

7 Schriften aus der Fakultät Wirtschaftsinformatik und Angewandte Informatik der Otto-Friedrich-Universität Bamberg

Gemeinschaftliche qualitätsgesicherte Erhebung und semantische Integration von raumbezogenen Daten

von Sebastian Matyas

UNIVERSITY OF BAMBERG PRESS

Schriften aus der Fakultät
Wirtschaftsinformatik und Angewandte Informatik
der Otto-Friedrich-Universität Bamberg

Schriften aus der Fakultät
Wirtschaftsinformatik und Angewandte Informatik
der Otto-Friedrich-Universität Bamberg

Band 7

University of Bamberg Press 2011

Gemeinschaftliche Qualitätsgesicherte Erhebung und Semantische Integration von Raumbezogenen Daten

von Sebastian Matyas



University of Bamberg Press 2011

Bibliographische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliographie; detaillierte bibliographische Informationen sind im Internet über http://dnb.ddb.de/ abrufbar

Diese Arbeit hat der Fakultät Wirtschaftsinformatik und angewandte Informatik der Otto-Friedrich-Universität als Dissertation vorgelegen
1. Gutachter: Prof. Dr. Christoph Schlieder
2. Gutachterin: Prof. Dr. Ute Schmid
Tag der mündlichen Prüfung: 15.02.2011

Dieses Werk ist als freie Onlineversion über den Hochschulschriften-Server (OPUS; http://www.opus-bayern.de/uni-bamberg/) der Universitätsbibliothek Bamberg erreichbar. Kopien und Ausdrucke dürfen nur zum privaten und sonstigen eigenen Gebrauch angefertigt werden.

Herstellung und Druck: Digital Print Group, Nürnberg
Umschlaggestaltung: Dezernat Kommunikation und Alumni

© University of Bamberg Press Bamberg 2011
http://www.uni-bamberg.de/ubp/

ISSN: 1867-7401
ISBN: 978-3-86309-002-9
eISBN: 978-3-86309-003-6
URN: urn:nbn:de:bvb:473-opus-3094

An dieser Stelle möchte ich zuallererst bei Prof. Schlieder bedanken. Ohne seine Unterstützung und vor allem seiner Toleranz gegenüber neuen Forschungsgebieten - seien sie auch noch so exotisch, wie das der ortsbezogenen Spiele - wäre diese Arbeit sicher nicht zu Stande gekommen.

Danke.

Eine Dissertation wird natürlich nicht im luftleeren Raum verfasst und deshalb gehen meine weiteren Danksagungen von ganzem Herzen an den tollen Lehrstuhl der Kulturinformatik und seine Mitarbeiter. Cornelia Pickel, für die Hilfe in allen Lebenslagen, Peter Kiefer, für die grandiose Zusammenarbeit bei allem was die Geogames betrifft (und allem darüber hinaus), sowie meinem Bruder, ohne den ich die vorliegende Arbeit sicher nicht vollendet hätte. Natürlich möchte ich Peter Wullinger, Klaus und Susanne nicht vergessen. Mein Dank gilt auch allen Hiwis, mit denen ich arbeiten durfte, vor allem an Steve Seypt und Manuel Beckmann.

Ein besonderer Dank geht an Mr. Yamada and Mr. Kunieda vom NEC C&C Innovation Research Laboratories, ohne deren Unterstützung die verteilte Nutzerstudie zwischen Bamberg und Fujisawa (Yokohama) nicht zu Stande gekommen wäre. Hiroko Mitarai und Maiko Kamata haben daran auch einen großen Anteil, meinen herzlichen Dank dafür an euch beide.

Die Zeit der Dissertation erstreckt sich natürlich nicht nur über die Arbeitszeit. Deshalb geht mein Dank auch an meine Eltern, die mich immer unterstützt haben. Ohne die teils notwendige Ablenkung durch die folgenden Personen, hätte ich die Zeit wohl nicht so gut überstanden: Carolina, Nora, Susanne, Nick, Nina, Britta und alle anderen vom Uni Bamberg Hockey-Team, die ich hier leider nicht alle namentlich nennen kann, sowie Christina. Vielen Dank für die schöne Zeit und moralische Unterstützung.

Zuletzt möchte ich noch einer Person im Speziellen danken, ohne die ich vor allem privat, niemals so lange durchgehalten hätte, der Ally [1].

Ich hoffe ich kann auch mal was von dir lesen.

[1] The rest is silence.

Inhaltsverzeichnis

1. Einleitung **1**

1.1. Problembeschreibung 2
1.1.1. Gemeinschaftliche qualitätsgesicherte Erhebung . . 2
1.1.2. Semantische Datenintegration 3
1.2. Hauptbeiträge 4
1.3. Gliederung 5

2. Gemeinschaftliche Datenerhebung **9**

2.1. Human Computation 13
2.2. Nutzergemeinschaften in der GIScience 18
2.2.1. Public Participation GIS (PPGIS) 18
2.2.2. Volunteered Geographic Information (VGI) 21
2.3. Ortsbezogene Spiele 27
2.3.1. Was ist ein ortsbezogenes Spiel? 28
2.3.2. Herausforderungen an das Design von ortsbezogenen Spielen 31
2.3.3. Ortsbezogene Spiele zur Datenerhebung 41
2.4. Semantische Datenintegration 43

3. Spielerische qualitätsgesicherte Erhebung raumbezogener Daten **55**

3.1. Klassifikation ortsbezogener Spiele zur Erhebung raumbezogener Daten 56
3.2. Strukturierung des Spielfeldes 59
3.3. Typen von sammelbaren Daten 62
3.4. Entwurfsmuster zur Motivation spielgetriebener Datenerhebung 64
3.4.1. *Log-Datei(en)* Konzept 65
3.4.2. *Rätsel/Aufgaben*-Entwurfsmuster 66
3.4.3. *Transporter*-Entwurfsmuster 67
3.4.4. *Erkundung*-Entwurfsmuster 68
3.4.5. *Produzent*-Entwurfsmuster 69

3.5. Entwurfsmuster zur Qualitätssicherung 70
3.5.1. *Peer Reviewing*-Entwurfsmuster 70
3.5.2. *Output-Input*-Entwurfsmuster 72
3.5.3. *Wisdom of the Crowd*-Entwurfsmuster 74
3.6. Weitere Eigenschaften 74
3.6.1. Wiederspielbarkeit 75
3.6.2. Asynchrones und langsames Spielen 76
3.6.3. Verteiltes Spielen 77
3.7. Design Framework 79
3.8. Anwendungsbeispiel: CityExplorer 80
3.8.1. Spielregeln 82
3.8.2. Zusatzregeln für ein verteiltes Spiel 86
3.8.3. Systemarchitektur 87
3.9. Zusammenfassung 93

4. Semantische Integration raumbezogener Daten 97
4.1. Semantischer Integrationsprozess 98
4.1.1. Semantisch annotierte raumbezogene Daten 101
4.1.2. Semantische Barrieren 105
4.2. Clustern von semantischen Sichtbereichen 106
4.2.1. Vorverarbeitug 106
4.2.2. Aggregation 108
4.3. Ausblick: Regelbasierte Datenaggregation 117
4.4. Zusammenfassung 120

5. Evaluation und Diskussion 121
5.1. CityExplorer-Studien 122
5.1.1. Nutzerstudien zum nicht-verteilten Spiel 122
5.1.2. Diskussion der Spielaspekte: nicht-verteiltes Spiel . 126
5.1.3. Nutzerstudie zum verteilten Spiel 133
5.1.4. Analyse der erhobenen raumbezogenen Datensätze . 134
5.1.5. Analyse der Geogames GPS-Tracks 139
5.2. Wisdom-of-the-Crowd Datenerhebung: GeoSnake 141
5.2.1. Hypothesen 144
5.2.2. Vorstudie 145
5.2.3. Hauptstudie 147
5.2.4. Variablen und Evaluierungsmodelle 149
5.2.5. Ergebnisse und Diskussion 150

5.3. Studie zum Integrationsprozess: SVAC-Algorithmus 153
5.3.1. Versuchsaufbau: Datengrundlage und Vergleichsalgorithmus 153
5.3.2. Ergebnisse und Diskussion 154
5.4. Ausblick: Simulationsumgebung 157

6. Zusammenfassung und Ausblick **159**
6.1. Ergebnisse 159
6.2. Ausblick 161
6.2.1. Spielerische qualitätsgesicherte Datenerhebung . . . 161
6.2.2. Semantischer Integrationsprozess 163

A. Anhang **165**
A.1. Fragebogen zu CityExplorer 165
A.1.1. Fragebogen 1 - Vor einem CityExplorer Spiel 165
A.1.2. Fragebogen 2 - Nach einem CityExplorer Spiel . . . 167
A.1.3. Ergebnis der Fragebogenauswertung zum verteilten Spiel (Japan) 172
A.2. Analyse der Raumbezogenen Datensammlung von CityExplorer 174
A.3. GPS Tracks Analyse 176
A.4. GeoSnake Materialien 179
A.4.1. Vorstudie: Allgemeine Spielanweisungen 179
A.4.2. Vorstudie: Objekte für GPS-Teil 181
A.4.3. Vorstudie: Objekte für Bildschirm-Teil 182
A.4.4. Hauptstudie: Allgemeine Spielanweisungen 184
A.4.5. Hauptstudie: Objekte für GPS-Teil 187
A.4.6. Hauptstudie: Objekte für Bildschirm-Teil 188
A.4.7. Fragebogen Vor- und Hauptstudie 190
A.4.8. Antworten Fragebogen Vorstudie 194
A.4.9. Antworten Fragebogen Hauptstudie 196
A.5. Material zum semantischen Integrationsprozess 198
A.5.1. Evaluierungsstudie zum SVAC-Algorithmus 198
A.6. GeoSnake - R Auswertung 206

Abbildungsverzeichnis

1.1. Prozentualer Anteil an aktiven Nutzern bei OpenStreetMaps 2

2.1. Beispiel für einen HIT auf der Mechanical Turk Webseite . . 11
2.2. Spielinterface des ESP Game 14
2.3. Spielprinzipien für GWAP-Spiele 14
2.4. Matrix zur Beteiligung und Klassifizierung der Öffentlichkeit nach [1] in PPGIS 19
2.5. Screenshot der *Ushahidi* Plattform 25
2.6. *Urban* und *Social Tapestries threads* 26
2.7. Nach [2], der Anteil an gespielten Miniaufgaben in Bezug auf den Spielort im Spiel *MobiMissions* 42
2.8. Nach [3], Iterationsschritte im *k-means* Algorithmus nach Ashbrook und Starner (2003) 46
2.9. Drei Möglichkeiten zur Bestimmung der Ähnlichkeit zweier Linienzüge 46
2.10. Neuer Anwendungsfall für Aggregationsansätze 1-D-dimensionaler raumbezogener Daten 54

3.1. Strukturierungsmöglichkeiten des Spielfeldes ortsbezogener Spiele: strukturiert - semi-strukturiert - unstrukturiert 60
3.2. Entscheidungspunkte bei der Spielfeldstrukturierung ortsbezogener Spiele 61
3.3. Kategorisierung der mit einem ortsbezogenen Spiel sammelbaren Daten 62
3.4. Schematische Darstellung des *Output-Input*-Entwurfsmusters 73
3.5. direkte und indirekt Interaktion zwischen Spielern in einem verteilten orstbezogenen Spiel (aus [4] entnommen) 78
3.6. Schematischer Aufbau des Design Framework zur Erstellung von ortsbezogenen Spielen zur Datenerhebung 79
3.7. Flowchart für den Designprozess eines ortsbezogenen Spieles zur Datenerhebung 81

3.8. Das Brettspiel Carcassonne 82
3.9. Die drei Spielphasen in CityExplorer 83
3.10. Web-basierter Wizard für das Peer-Reviewing Verfahren . . 85
3.11. Screendesign für ein verteiltes CityExplorer Spiel 87
3.12. Aufbau der Systemarchitektur von CityExplorer 88
3.13. Screenshots zu CityExplorer 89
3.14. Paketdiagram des Geogames J2ME Software Framework und die zugehörige CityExplorer Implementierung (entnommen aus Matyas (2007) [5]) 94
3.15. Screenshots der J2ME Implementierung und der DoJa (i-appli) Implementierung von CityExplorer 95

4.1. Ungelöstes Integrationsproblem im 1-D-dimensionalen Fall . 99
4.2. Ungelöstes Integrationsproblem im 0-D-dimensionalen Fall . 100
4.3. Zusätzliches Integrationsproblem im 0-D-dimensionalen Fall 101
4.4. Blickvektor als Unterscheidungsmerkmal von geographisch und semantisch ähnlichen POIs 102
4.5. Unzugänglichkeit eines Spielortes und rekonstruierte Sichtbereiche zweier Spielsteine 104
4.6. Datenstruktur eines CityExplorer Spielsteines 104
4.7. Sichtbereiche für die Spielsteine aus dem Restaurant-Szenario 105
4.8. Schritte in einem allgemeinem Integrationsprozess 107
4.9. Pufferbereiche nach dem Beschneiden der jeweiligen Sichtbereiche 109
4.10. Möglichkeiten der Clusterbildung aufgrund der geometrischen Nähe 110
4.11. Günstige Konstellationen für Vereinigung bzw. Schnitt als Methode zur Erstellung des Clustersichtbereiches s_c 112
4.12. 114
4.13. Simple View Area Clustering 115

5.1. Spieleinführung und Hardwareausgabe vor Beginn eines CityExplorer Spieles 123
5.2. Spielfeld für alle Studien zu CityExplorer in der Bamberger Innenstadt, sowie die von den Spielern gewählte Kategorien für diese Spiele 124
5.3. Endzustand des ersten CityExplorer Spieles in Bamberg . . . 125
5.4. Endzustand des zweiten CityExplorer Spieles in Bamberg . . 125

5.5. Spielverlauf des dritten CityExplorer Spieles in Bamberg und typisches Beispiel für das *Überflutungsproblem* bei ortsbezogenen Spielen zur Datenerhebung 131
5.6. Gesetzte Spielsteine in Spiel 3 insgesamt (blauer Balken) und von Spieler 2 (roter Balken) 132
5.7. Endzustand des verteilten Spieles zwischen Bamberg und Fujisawa und die im Spiel verwendeten Kategorien 134
5.8. OQM Verteilung für eine ausgewählte Stichprobe aus CityExplorer Spielsteinen 138
5.9. Vergleich von p_m und p_p für Spielsteine aus M und Google Maps POI aus Q 138
5.10. GPS Track erhoben während der Girls Day CityPoker [6] Spiele in der Innenstadt von Bamberg 140
5.11. Thematische Karten im Maßstab 1:25000 des Landesvermessungsamt Bayerns (TK25) für Bamberg und Coburg; in Rot mit den beiden Geogames erhobene GPS Tracks 141
5.12. Suchergebnis auf der Community-Webseite Locr (`http://de.locr.com/`) für raumbezogene Datensätze zum Stichwort „Altes Rathaus Bamberg“ 143
5.13. Beispiel für die Streuung der erhobenen GPS Koordinaten mit den Regelvarianten 1 (rote Markierung) und 2 (blaue Markierung) für die Orte „Geyerswörth“ und „St. Stephans Kirche“ in der Vorstudie zu GeoSnake 148
5.14. Beispiel für die Streuung der erhobenen GPS Koordinaten mit den Regelvarianten 1 (rote Markierung) und 2 (blaue Markierung), sowie der Markierungen (gelb), die im PC-Teil gesetzt wurden, für die Orte „Kleberstraße“ (oben), „Geyerswörth“ (unten links) und „Honer“ (unten rechts) 151
5.15. Simulationsumgebung für den Semantischen Integrationsprozess - Beispielhafte Einbindung des Aggregationsalgorithmus nach [7] 158

A.1. IQM des POV für die Datensammlung des ersten CityExplorer Spiels in Bamberg 174
A.2. IQM des APL für die Datensammlung des ersten CityExplorer Spiels in Bamberg 174
A.3. IQM des POV für die Datensammlung des zweiten CityExplorer Spiels in Bamberg 175

A.4. IQM des APL für die Datensammlung des zweiten CityExplorer Spiels in Bamberg 175
A.5. OQM Verteilung für die gesamte Stichprobe 176
A.6. Geographische Verteilung der Objekte in der Vorstudie zu GeoSnake im GPS-Teil, © Karte: Google Earth (`http://earth.google.com/intl/de/`). 181
A.7. Geographische Verteilung der Objekte in der Vorstudie zu GeoSnake im GPS-Teil, © Karte: Google Earth (`http://earth.google.com/intl/de/`). 182
A.8. Geographische Verteilung der Objekte in der Hauptstudie zu GeoSnake im GPS-Teil, © Karte: Google Earth (`http://earth.google.com/intl/de/`). 187
A.9. Geographische Verteilung der Objekte in der Hauptstudie zu GeoSnake im PC-Teil, © Karte: Google Earth (`http://earth.google.com/intl/de/`). 188

Tabellenverzeichnis

3.1. Klassifizierungsschema für ortsbezogene Spiele 59

5.1. Ergebnisse der Fragebogenauswertung - Likert Skala von 1 (Strongly disagree) bis 5 (Strongly agree) - bei n = 12 127
5.2. Mittelwert für die IQM Werte für Spiel 1 und 2 in Bamberg . 136
5.3. OQM Werte für die CityExplorer Datensammlung 137
5.4. Pseudo-zufällige Reihenfolge der Experimentteile und Regelvarianten (1S = Regelvariante 1 am Bildschirm, 1V und 2V = Regelvariante 1 und 2 mit Mobiltelefon) 146
5.5. Häufigkeit der besuchten Orte bezüglich der Regelvarianten; die Zahl in Klammern gibt die maximale Anzahl an abgegebenen GPS Koordinaten an und die Zahl vor der Klammer gibt die Anzahl an fehlerlosen GPS Koordinaten an. 147
5.6. Pseudo-zufällige Reihenfolge der Experimentteile und Regelvarianten (1S = Regelvariante 1 am Bildschirm, 1V und 2V = Regelvariante 1 und 2 mit Mobiltelefon, V1 und V2 = Spielfeld 1 bzw. 2) 149
5.7. Quelle: GeoSnake Spiel 1. Anmerkung: *** <0; ** <0.05; * <0.1 152
5.8. Quelle: GeoSnake Spiel 1. Anmerkung: *** <0; ** <0.05; * <0.1 152
5.9. Quelle: GeoSnake Spiel 2. Anmerkung: *** <0; ** <0.05; * <0.1 152
5.10. Quelle: GeoSnake Spiel 2. Anmerkung: *** <0; ** <0.05; * <0.1 153
5.11. Clustervergleich Ergebnisse 156

A.1. Ergebnisse der Fragebogenauswertung - Likert Skala von 1 (Strongly disagree) bis 5 (Strongly agree) - bei n = 3 173

A.2. Auflistung der gesammelten GPS Koordinaten in den 48 Geogames Partien - mit GeoTicTacToe und CityPoker - in Bamberg (Innenstadt und Feki), Coburg und Bremen; MUT = Mädchen und Technik [6]; EAI = Einführung in die Angewandte Informatik; WELT = Welterbetag [6]; BREMEN_1 = Verteiltes Spiel Bamberg vs. Bremen [4] 177
A.3. Auflistung der gesammelten GPS Koordinaten in den 48 Geogames Partien - mit GeoTicTacToe und CityPoker - in Bamberg (Innenstadt und Feki), Coburg und Bremen; MUT = Mädchen und Technik [6]; EAI = Einführung in die Angewandte Informatik; WELT = Welterbetag [6]; BREMEN_1 = Verteiltes Spiel Bamberg vs. Bremen [4] 178
A.4. Antowrten zum Fragebogen der Vorstudie 194
A.5. Antowrten zum Fragebogen der Vorstudie 195
A.6. Antowrten zum Fragebogen der Vorstudie 195
A.7. Antworten zum Fragebogen der Hauptstudie 196
A.8. Antworten zum Fragebogen der Hauptstudie 197
A.9. Antowrten zum Fragebogen der Hauptstudie 197
A.10. Verwendete Kategorien in der Evaluierungsstudie für den SVAC-Algorithmus und die in der CityExplorer gezählten Instanzen 198
A.11. Nach [8] berechnete Ähnlichkeiten zwischen den Kategorien 199
A.12. Nach [8] berechnete Ähnlichkeiten zwischen den Kategorien 200
A.13. Nach [8] berechnete Ähnlichkeiten zwischen den Kategorien 201
A.14. Nach [8] berechnete Ähnlichkeiten zwischen den Kategorien 202
A.15. Nach [8] berechnete Ähnlichkeiten zwischen den Kategorien 203
A.16. Nach [8] berechnete Ähnlichkeiten zwischen den Kategorien 204
A.17. Nach [8] berechnete Ähnlichkeiten zwischen den Kategorien 205

1. Einleitung

In den vergangenen Jahren ist die Verbreitung von mobilen Geräten mit integrierter Lokalisierungstechnologie (z.B. GPS) stark gestiegen. Jeder, der ein solches Gerät besitzt, kann zum Datenlieferanten werden und in einem Netzwerk von Freiwilligen die Rolle eines Datensensors übernehmen (voluntary sensor, siehe Goodchild 2007 [9]). Im Rahmen des *Web 2.0* (z.B. [10] oder [11]) sind eine Reihe von Anwendungen entstanden, die es einem Nutzer ermöglichen, selbst gesammelte raumbezogene Daten mit anderen Nutzern auszutauschen.

Die Internetportale OpenStreetMap[1] oder Wikimapia[2] sind dafür prominente Beispiele. Die dort von Freiwilligen, in den allermeisten Fällen keine geowissenschaftlichen Fachanwender, gelieferten raumbezogenen Daten beinhalten z.B. einzelne Orte (*Point of Interest*, *POI*) oder zurückgelegte Wege (*Tracks*), sowie Objekte, mit einer gewissen geographischen Ausdehnung wie Plätze, Wälder oder auch ein Gebäudekomplex. Zusätzlich kann der Anwender seine Daten mit semantischer Information anreichern, z.B. in Form einer Kategorisierung. Das Ziel bei beiden Portalen besteht in einer digitalen Abbildung der realen Welt.

Ein weiteres Beispiel - mit einem etwas weniger ambitionierten Ziel - ist das PARAMOUNT[3] Projekt (Loehnert et al., 2001) [12]. Hier wurde ein ortsbezogener Dienst für Wanderer und Bergsteiger aufgebaut. Die Anwender dieses Dienstes können nicht nur ihre Wandertouren hochladen, sondern beispielsweise auch persönliche Erfahrungen anderen mitteilen und/oder eine Bewertung zu jeder Tour abgeben. Bellavista et al. (2008) [13] behaupten sogar, dass für diese Art von Anwendungen ein von einer Nutzergemeinschaft getragener Web-2.0-Ansatz für den kommerziellen Erfolg unerlässlich sei.

[1] `http://www.openstreetmap.org/`
[2] `http://wikimapia.org/`
[3] `http://www.paramount-tours.com/`

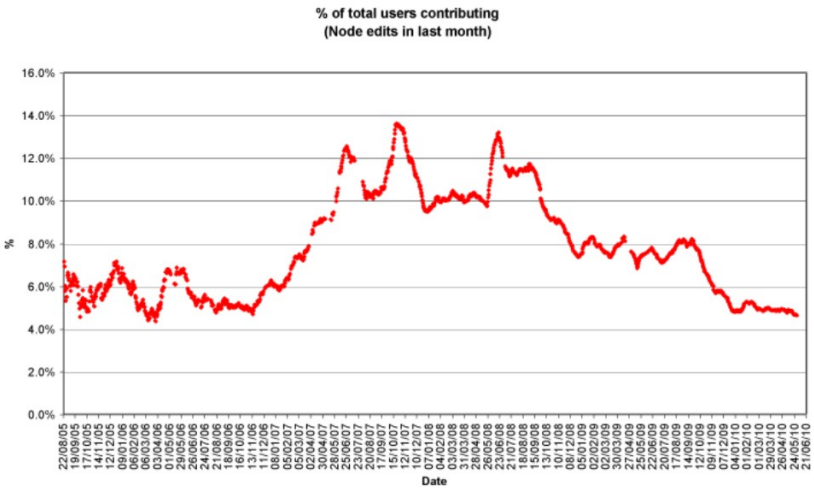


Abbildung 1.1.: Prozentualer Anteil der Nutzer, die Daten in OpenStreetMap editieren. Abbildung von `http://wiki.openstreetmap.org/wiki/Stats`

1.1. Problembeschreibung

In der vorliegenden Arbeit bearbeite ich zwei Problemstellungen, die im Kontext von raumbezogenen Daten, erhoben von Gemeinschaften aus fachlich (hier: geowissenschaftlich) nicht geschulten Nutzern, noch ungelöst sind:

1.1.1. Gemeinschaftliche qualitätsgesicherte Erhebung

Kuhn (2007) [14] stellte auf dem ersten Workshop für Volunteered Geographic Information (VGI) - ein Forschungsfeld, welches sich explizit mit Nutzergemeinschaften aus geowissenschaftlichen Laien beschäftigt - 2007 zahlreiche Herausforderungen für dieses Forschungsfeld vor. Unter anderem muss das Problem der Motivation gelöst werden, d.h. wie man geowissenschaftliche Laien dazu bewegen kann, freiwillig raumbezogene Daten zu sammeln und einer Nutzergemeinschaft zur Verfügung zu stellen. Idealerweise werden in so einer Nutzergemeinschaft alle Konsumenten auch zu Datenproduzenten. Aber das Beispiel OpenStreetMap zeigt, dass dies eben noch nicht der Fall ist. Die Plattform hat nur einen Anteil von maximal 14% Produzenten, sowohl auf Seiten der initialen Erhebung als auch auf der Seite der späteren

Nachbearbeitung der bereitgestellten Daten (siehe Abbildung 1.1).

Ohne die Möglichkeit einer Motivation durch eine entsprechende finanzielle Entlohnung (sog. *financial incentives* [15]), wie bei Fachanwendern/Experten üblich, müssen daher den potentiellen „produsers“ [16] oder „Prosumenten“ [17] neue Anreize geboten werden, damit sie überhaupt zu Datenproduzenten werden.

Damit ergibt sich folgende erste Forschungsfragestellung:

Wie können geowissenschaftliche Laien motiviert werden, freiwillig qualitätsgesicherte raumbezogene Daten zu erheben?

Der erste Beitrag meiner Dissertation liefert eine Antwort auf diese Frage. Er beinhaltet ein Framework zum Design von ortsbezogenen Spielen, die geowissenschaftliche Laien motivieren, raumbezogene Daten qualitätsgesichert in spielerischer Art und Weise zu sammeln. Um es mit den Worten von Peltola et al. (2006) [18] auszudrücken:

> game play is a natural motivator to participate in something that is not necessary or beneficial. [...] By controlling game events and perhaps game logics and rules, the agencies that ultimately use the gathered data, can steer players to do tasks supporting their needs.

1.1.2. Semantische Datenintegration

Obwohl schon qualitätssichernde Elemente in ein ortsbezogenes Spiel zur Erhebung von raumbezogenen Daten integriert werden können, ist die Güte der Daten stark abhängig vom eingesetzten Endgerät, der Lokalisierungstechnologie und dem fachgerechten Umgang mit dieser Technologie. So liegt die Genauigkeit von GPS als Lokalisierungstechnologie zwar im 5-Meter-Bereich, doch besitzen geowissenschaftliche Laien nicht die hochgenauen GPS-Geräte aus dem Vermessungswesen. Außerdem fehlt es ihnen oft an Wissen, um die Messergebnisse ihrer eigenen Hardware richtig zu interpretieren.

Im Rahmen meiner Arbeit möchte ich noch folgende zweite Forschungsfragestellung untersuchen:

Wie kann die Qualität von Sammlungen semantisch angereicherter raumbezogener Daten verbessert werden?

Eine wichtige Eigenschaft von ortsbezogenen Spielen zur Datenerhebung ist die wiederholte Spielbarkeit. Im Gegensatz zu anderen Domänen ist nämlich die Erstellung redundanter Daten im raumbezogenen Kontext sogar erwünscht. Das Zusammenführen von Daten mehrerer Nutzer kann man als Mehrfachmessung auffassen, die sich dazu nutzen lässt, die Qualität von Lokalisierung (Wo?) und Kategorisierung (Was?) zu verbessern. Zusätzlich können durch die Einführung von entsprechenden Spielregeln, raumbezogene Daten erhoben werden, deren Struktur neuartige Ansätze zur Integration erlauben. Der zweite Beitrag meiner Dissertation gibt eine Antwort auf die Fragestellung und besteht in einem Ansatz zur semantischen Integration der gesammelten raumbezogenen Daten.

1.2. Hauptbeiträge

Meine Dissertation behandelt den gesamten Weg der raumbezogenen Daten von der Erhebung in der realen Welt bis hin zur endgültigen Integration in eine digitale Datensammlung. Um diesen Prozess vollständig abzubilden, gehe ich folgendermaßen vor:

1. Auf Grundlage einer systematischen Betrachtung von mehr als 130 ortsbezogenen Spielen, stelle ich ein **Design Framework** zur Entwicklung von ortsbezogenen Spielen vor, welche Gemeinschaften von geowissenschaftlichen Laien dazu motivieren, raumbezogene Daten zu erheben. Inspiriert durch Paradigmen für nicht-ortsbezogene Spiele zur Datenerhebung von Luis von Ahn (2004/2008) - [19] and *Games with a Purpose* (GWAP) in [20] - und in Anlehnung an Arbeiten von Björk et al. (2004, 2007) [21] oder [22] bietet das Framework einem Spieledesigner verschiedene **Entwurfsmuster** für **Spielelemente** zur Erhebung raumbezogener Daten. Spezielles Augenmerk lege ich auch auf Entwurfsmuster, die eine **Qualitätssicherung** im Spiel selbst möglich machen.

2. Unter Verwendung des Design Frameworks habe ich einen **Prototyp**, das ortsbezogene Spiel *CityExplorer*[4], entworfen und implementiert. Der Prototyp wurde in mehreren Nutzerstudien verwendet, um die verschiedenen Aspekte des Spiels, bzw. die dort verwendeten Entwurfsmuster, zu evaluieren.

[4] `http://www.kinf.wiai.uni-bamberg.de/cityexplorer/`

3. Zusätzlich habe ich durch eine **emprirische Untersuchung** gezeigt, dass der *Wisdom of the Crowd*-Ansatz von Surowiecki (2004) [23], in Form eines entsprechenden Entwurfsmusters, zur Qualitätssicherung in ortsbezogenen Spielen erfolgreich eingesetzt werden kann.

4. Ich schlage einen **semantischen Integrationsprozess** vor. Auf Grundlage der speziellen Datenstruktur, der durch CityExplorer erhobenen raumbezogenen Daten, entwickle ich für den Aggregationsteil des Integrationsprozesses einen neuartigen Puffer-basierten Aggregationsalgorithmus, den ***Semantic View Area Clustering (SVAC)***. Des Weiteren führe ich den Begriff der ***semantischen Barriere*** ein, welche auf einem bisher noch nicht diskutierten Problemfall in der Aggregation von raumbezogenen Daten basiert. Der SVAC löst das Problem der *semantischen Barrieren*. Eine Evaluierung meines Ansatzes gegenüber einer Clusteranalyse nach Ashbrook und Starner (2003) [3] gibt erste Eindrücke über die Genauigkeit und Trefferquote von SVAC.

Für die weitere Evaluierung verschiedenster Ansätze des Aggregationsteils im semantischen Integrationsprozess habe ich prototypisch ein web-basierte Simulationsumgebung entwickelt. Mit dieser ist es möglich, auf die Datensammlung aus den *CityExplorer* Spielen beliebige Algorithmen anzuwenden. Außerdem präsentiere ich einen ersten Entwurf eines weiteren Aggregationsansatzes (*RaDAG, Rule-based Data Aggregation*), der auf einer regelbasierten beschreibungslogischen Sprache beruht.

1.3. Gliederung

In meiner Dissertation bin ich zuerst der Frage nachgegangen, wie man geowissenschaftliche Laien motivieren kann, freiwillig raumbezogene Daten qualitätsgesichert erheben zu lassen und sie in eine Datensammlung einzustellen. Danach beschäftigte ich mich mit einem semantischen Integrationsprozess, um die Datensammlung noch einmal qualitativ zu verbessern, insbesondere unter Ausnutzung von redundant erhobenen Datensätzen und des geographischen und semantischen Kontextes.

Dazu gebe ich in Kapitel 2 erst einmal einen Überblick auf die die vorliegende Arbeit betreffenden Forschungsgebiete. Nutzergemeinschaften und deren Einsatz zur Datenerhebung sind unter anderem in den Gebieten des *Human Computing*, des *Public Participation GIS* und der *Volunteered Geographic Information* Gegenstand der Forschung. In den letzteren zwei findet

eine spielerische Herangehensweise an die in Abschnitt 1.1 vorgestellte Problematik erst seit kurzem eine Erwähnung. Diese war im Bereich des *Human Computing* schon von Anfang an eine intensiv diskutierte Lösungsmöglichkeit. Die wichtigsten Ansätze und Überlegungen in diesen Gebieten werde ich im Hinblick auf eine spielerische Erhebung von raumbezogenen Daten dort diskutieren. Ebenfalls in diesem Kapitel stelle ich die bekanntesten Ansätze zur geographischen und semantischen Datenintegration vor und werde abschließend einen Anwendungsfall präsentieren, den diese bisher noch nicht untersucht haben.

Einer der Hauptbeiträge meiner Arbeit ist die Entwicklung eines Frameworks, welches es einem Designer ermöglicht, orstbezogene Spiele zur Datenerhebung zu erstellen. Dazu stelle ich zu Beginn von Kapitel 3 meine Klassifizierung der in der Forschung entworfenen und kommerziell erhältlichen ortsbezogenen Spielen vor, aus denen ich ein allgemeines Design Framework ableite. Dieses enthält zum einen Entwurfsmuster für entsprechende Spielelemente, die sowohl die Datenerhebung als auch die Qualitätssicherung betreffen. Zum Anderen sind auch die Strukturierung des Spielfeldes, welche Typen von Daten erhoben werden können, sowie weitere wichtige Eigenschaften ortsbezogener Spiele im Framework enthalten. Zum Abschluss von Kapitel 3 gehe ich detailliert auf das orstbezogene Spiel *CityExplorer* ein, das ich mit meinem Design Framework entworfen habe.

Durch den Entwurf entsprechender Spielregeln können mit CityExplorer raumbezogene Daten erhoben werden, die beliebige Komplexität aufweisen. Neben der formalen Definition dieser Komplexität, bzw. der daraus resultierenden Struktur, führe ich in Kapitel 4 noch den Begriff der *semantischen Barriere* ein. Mit Hilfe dieser speziellen Datenstruktur schlage ich einen neuartigen Algorithmus - den SVAC-Algorithmus - vor, der die *semantischen Barriere* in einer Sammlung von raumbezogenen Daten nutzt, um die Genauigkeit der Aggregation zu verbessern. Mit einem Ausblick auf einen weiteren Aggregationalsgorithmus - den RaDAG-Algorithmus - endet das Kapitel.

In Kapitel 5 diskutiere ich die Ergebnisse der verschiedenen Nutzerstudien, die ich mit *CityExplorer* durchgeführt habe und zeige, dass trotz der Einbindung von „Arbeitsaufgaben“, als welche das Erheben von raumbezogenen Daten angesehen werden kann, der Spielspaß und -fluss nicht beeinträchtigt wird. Das *Wisdom of the Crowd*-Entwurfsmuster, das in CityExplorer keine Anwendung fand, untersuche ich in einer zusätzlichen Nutzerstudie. Dazu verwende ich das ortsbezogene Spiel *GeoSnake*, welches ich für die Studie entworfen habe. Vor allem die Positionsgenauigkeit der raumbezogenen Datensätze steht hierbei im Fokus der Untersuchung und zeigt den positiven

Einfluss des Entwurfsmusters. Abschließend belegt die Auswertung der Ergebnisse des *SVAC*-Algorithmus, verglichen mit dem Clusteringalgorithmus nach Ashbrook und Starner (2003) [3], eine bessere Genauigkeit meines Ansatzes.

In Kapitel 6 fasse ich die Ergebnisse meiner Arbeit zusammen und gebe einen Ausblick auf weitere interessante Fragen aus dem Bereich der von mir bearbeiteten Forschungsfelder.

2. Gemeinschaftliche Datenerhebung

In meiner Arbeit verfolge ich den Ansatz, mit Hilfe von ortsbezogenen Spielen Gemeinschaften von geowissenschaftlichen Laien zu motivieren, Daten über ihre geographische Umwelt zu erheben. Dieses lokale Wissen ist für viele verschiedene Anwendungen, wie ortsbezogene Dienste[1], Navigationssysteme[2] oder ortsbezogene/pervasive Spiele[3] von großem Nutzen und wird dort im Moment hauptsächlich von Experten erhoben.

Die raumbezogenen Daten (rD), die in diesen Anwendungen verarbeitet und/oder angezeigt werden, werden üblicherweise als Dreitupel $rD \equiv_{\text{def}} \langle name, cat, footprint \rangle$ definiert - identisch zu Gazetteer-Einträgen z.B. nach Hill (2000/2006) [24] und [25] -, wobei:

$name$ Ein beliebiger Bezeichner für den Datensatz ist. Dieser muss nicht eindeutig sein, wie z.B. der Namen für einen Ort oder ein Restaurant. Hier kann auch ein Vektor verwendet werden, falls Alternativen für Namen mit aufgenommen werden sollen, die sich z.B. aus historischen Gründen gewandelt haben können.

cat Eine Ausprägung einer Kategorie/Klasse/Typenbezeichnung für den Datensatz, ausgewählt aus einem vorgegebenen Schema. In Ausnahmefällen kann ein Datensatz auch Ausprägungen verschiedener Kategorien aufweisen, wiederum als Vektor.

$footprint$ Ein Vektor aus Koordinaten, der entweder einen Punkt, eine Linie oder ein Polygon repräsentiert.

Die Idee der gemeinschaftlichen Datenerhebung kam zuerst außerhalb des Geographic Information Science (GIS) Forschungsfeldes im Bereich der Künstlichen Intelligenz auf. Eines der ersten größeren Initiativen meines Wissens

[1] `http://maps.google.de/`
[2] z.B.: `http://www.navigon.com/portal/sites.html`
[3] z.B.: `www.pervasive-gaming.org/`

war dabei die *Open Minds Initiative*[4], die im Jahr 2000 ins Leben gerufen wurde [26]. Nach eigenen Angaben ist das Ziel der Initiative ein „kollaboratives Framework zur Erstellung von „intelligenter“ Software, die das Internet als Medium verwenden“. Die Software, als auch die Daten, die die Software zu einem großen Teil erst „intelligent“ machen, ist dabei für alle offen. *Open Minds* unterscheidet sich aber in einem wichtigen Punkt von normalen Open Source Software Initiativen. Neben Experten, aus der entsprechenden Domäne, sowie Softwareentwicklern, sollen auch explizit normale Internetnutzer (*non-expert netizens* [26]) in die Projekte eingebunden werden. Die *netizens* sollen helfen, die für beispielsweise Handschrifterkennungssoftware nötigen Datengrundlagen, wie annotierte Textsammlungen (Corpora), zu erstellen [27].

Ein Beispiel, welches in [26] vorgestellt wird, ist das Programm *Animals*. Es ist als Frage-Antwort Spiel aufgebaut ist. Das Spiel hat folgende Regeln: Ein menschlicher Spieler überlegt sich zu Beginn einer Spielrunde ein beliebiges Tier. Dann bekommt er von einem KI-Gegner eine Reihe von maximal 20 Fragen gestellt, die er mit Ja oder Nein beantworten muss. Die Reihenfolge der Fragen entspricht einem Pfad in einem Binären Entscheidungsbaum. Wird ein Blattknoten durch so einen Frage-Pfad erreicht, wird dessen Inhalt als Vermutung für das ausgedachte Tier des menschlichen Spielers abgegeben. Ist die Vermutung richtig, endet das Spiel und eine neue Spielrunde kann gestartet werden. Ist die Vermutung falsch, wird der Spieler aufgefordert eine Frage zu formulieren, anhand der das Programm das fälschlich vorgeschlagene Tier von dem des Spielers unterscheiden kann. Diese Frage wird dann wiederum in den Entscheidungsbaum als neuer Ast eingebaut.

Geht man nach der Definition von Salen und Zimmermann [28] fehlt hier allerdings der Konflikt des Spielers mit dem Spielsystem (*conflict* [28]), um es zu einer wirklichen Spielerfahrung für den Spieler werden zu lassen. Denn eigentlich ist hier mehr das System mit dem Spieler in Konflikt als andersherum.

Dieser erste spielerische Ansatz wurde 2005 auf dem AAAI Spring Symposium zu „Knowledge Collection from Volunteer Contributors“ (KCVC05) weiter vertieft. Das dort von Ahn und Dabbish vorgestellte *ESP Game* [19] stellt eine konsequente Fortführung des Ansatzes der *Open Minds Initiative* dar. Durch weitere Arbeiten von Ahn und Kollegen wurde daraus ein eigenes Forschungsfeld: das Feld des *Human Computation* [29] [30].

Neben diesem Forschungsgebiet, das ich im folgenden Abschnitt im Detail

[4] `http://www.openmind.org/`

betrachten möchte (Abschnitt 2.1), soll auch das Gebiet des *Crowdsourcing* - welches eher in den Wirtschaftswissenschaften zu Hause ist - nicht unerwähnt bleiben.

Der Begriff selbst wurde zwar erst durch das Buch von Jeff How [31] bekannt, aber schon mit der Veröffentlichung der *Mechnical Turk*-Plattform[5] im November 2005 durch Amazon[6] wurde diese Art der Datenerhebung einer breiten Öffentlichkeit zugänglich gemacht.

Das Prinzip hinter *Mechanical Turk* ist relativ einfach. Nutzer der Plattform können sogenannte *HITs* (Human Intelligence Tasks) einstellen (dann als *Requester* bezeichnet), die mehr oder weniger komplexen Arbeiten entsprechen, die sie von anderen Nutzern (dann *Provider* genannt) erledigt haben möchten. Abbildung 2.1 zeigt beispielsweise einen HIT, bei dem überprüft werden soll, ob das gezeigte Bild zum Ausdrucken geeignet ist.

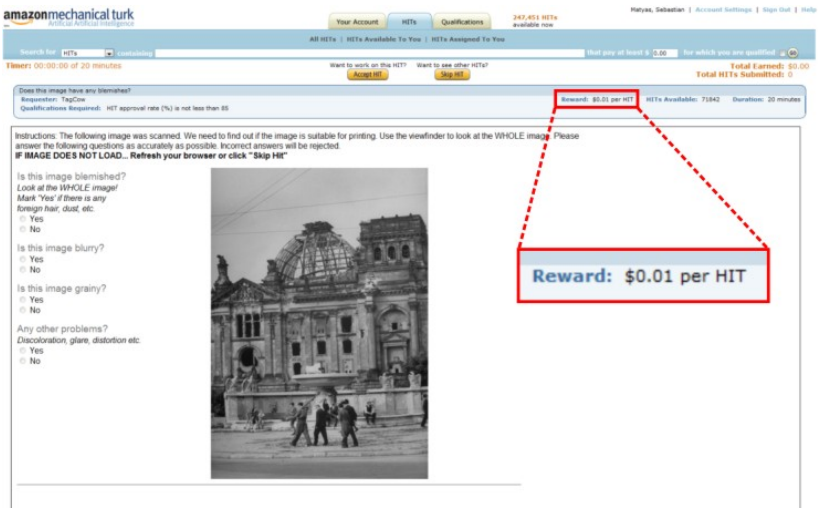


Abbildung 2.1.: Beispiel für einen HIT auf der Mechanical Turk Webseite, `https://www.mturk.com/mturk/welcome`

Anders als bei Projekten, wie zum Beispiel Wikipedia[7], die ganz auf eine freiwillige Teilnahme ihrer Nutzer am Datenerhebungs-/bearbeitungsprozess setzen, wird das Bearbeiten von HITs mit einer Vergütung finanzieller Art

[5] `https://www.mturk.com/mturk/welcome`
[6] `https://www.amazon.de/`
[7] `http://de.wikipedia.org/wiki/Wikipedia:Hauptseite`

belohnt. Im gezeigten Beispiel aus Abbildung 2.1 bekommt ein Nutzer für jedes bearbeitete Bild einen Geldbetrag von 0.01$ (*Financial Incentive* [15]) gutgeschrieben. *Requester* haben allerdings die Möglichkeit die Arbeit von *Providern* abzulehnen, falls die produzierten Daten nicht ihren Qualitätsansprüchen genügen. Es ist ebenso für *Requester* möglich einzustellen, dass *Provider* bestimmte Qualifikationen erfüllen müssen, damit sie einen HIT annehmen können und zusätzlich kann diese Voraussetzung durch einen Test überprüft werden.

Nichtsdestotrotz muss aber die Qualitätsüberprüfung von den *Requestern* selbst vorgenommen werden, auch wenn Studien gezeigt haben, dass die Qualität der Daten durchaus vergleichbar mit denen von Domänenexperten zu bewerten ist [32]. Mason und Watts (2009) [15] sind zusätzlich noch der Frage nachgegangen, ob ein höherer finanzieller Gegenwert mit einer höheren Qualität, der von den *Providern* abgelieferten Daten, einhergeht. Zwar konnten die Autoren einen Zusammenhang zwischen Bezahlung und Quantität feststellen, aber nicht in Bezug auf die Qualität. D.h. auch bei einer höheren Bezahlung war die abgelieferte Arbeit nicht signifikant qualitativ besser, als bei geringerer oder gar keiner Bezahlung. Mason und Watts führen diesen Effekt auf den „Anchoring“ Effekt nach [33] zurück.

Zu denselben Ergebnissen kam man auch Mrozinski et al. (2008) in [34]. Hier wurde sogar festgestellt, dass eine höhere Bezahlung mit einer Verringerung der Qualität einhergeht. Sie führen dieses Ergebnis auf die Zusammensetzung der Gemeinschaft der *Provider* bei *Mechanical Turk* zurück. Ab einem gewissen *Incentive*-Schwellwert, so die Argumentation in [34], werden vor allem solche *Provider* angezogen, die einen HIT primär aus Geldgründen bearbeiten und dabei natürlich so schnell wie möglich vorgehen, um ihren Ertrag zu maximieren. Darunter leidet aber die Qualität. Bei einem niedrigeren *Incentive*-Wert finden sich dagegen vor allem solche Nutzer, die weniger des Geldes wegen die Plattform nutzen, sondern eher nach interessanten Aufgaben suchen.

Auch wenn das *Crowdsourcing* eine Rolle im Bereich des Mobile Computing spielt, siehe [35] und [36], bleibt die Frage der Qualitätssicherung doch auch hier entweder unbeantwortet, bzw. Aufgabe der Institution, die die Aufgaben stellt.

2.1. Human Computation

Das Feld des Human Computation beinhaltet neben Ansätzen zur spielerischen Datenerhebung natürlich auch nicht-spielbezogene. Diese beschäftigen sich z.B. mit der Fragestellung, wie man menschliche Internetnutzer von Bots unterscheiden kann [37]. Aber im Kontext meiner Arbeit möchte ich mich auf eine Betrachtung des Themenbereiches der spielbezogenen Ansätze beschränken.

Wie oben schon erwähnt, markiert das von Ahn und Dabbish (2005) entwickelte Spiel *ESP Game* [19], welches auch als *Google Image Labeler*[8] bekannt ist, sozusagen den Beginn der systematischen Betrachtung der spielerischen Datenerhebung. Das Ziel des Spieles ist es, sinnvolle Beschreibungen (Tags), die nur aus einem Wort bestehen, für digitale Bilder zu generieren. Die Bilder stammen von Webseiten aus dem gesamten Internet. Das Spielprinzip gestaltet sich wie folgt - das zugehörige Spielinterface wird in Abbildung 2.2 dargestellt:

- Zwei Spieler werden anonym und zufällig miteinander gepaart.
- Beide bekommen dasselbe Bild angezeigt und müssen dieses mit entsprechenden Begriffen (Tags) beschreiben.
- Sobald ein Spieler einen Tag eingibt, der schon vom anderen Spieler eingegeben wurde, bekommen beide Punkte gutgeschrieben und das nächste Bild wird angezeigt. Die Punktzahl richtet sich danach, wie schnell die Spieler einen gemeinsamen Tag finden.
- Die auf der linken Seite in Abbildung 2.2 angezeigten Taboo-Wörter dürfen für das Bild nicht mehr verwendet werden.
- Nach einer gewissen Anzahl an erfolgreich verschlagworteten Bildern bekommen die Spieler, je nach Geschwindigkeit, Bonuspunkte gutgeschrieben.

Ausgehend von diesem ersten Spiel haben Ahn und Kollegen zahlreiche andere Spiele zur Datenerhebung entwickelt, die sie selbst als *Games with a purpose* (*GWAP*) [20] bezeichnen. In Ahn und Dabbisch (2008) [20] werden diese Spiele von den Autoren in drei Kategorien eingeteilt: *Output-agreement games Inversion-problem games Input-agreement games*. Diese sind in Abbildung 2.3 von links nach rechts dargestellt.

[8] `http://images.google.com/imagelabeler/`

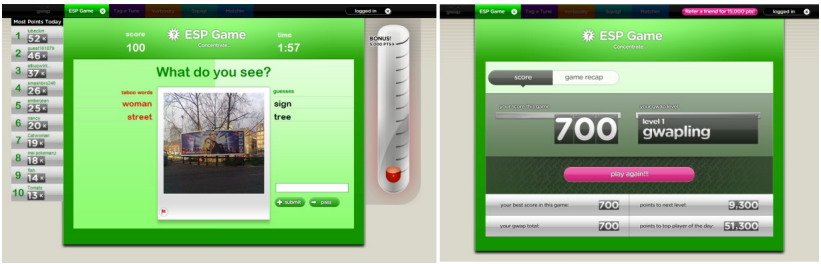


Abbildung 2.2.: Screenshot des Spielinterface des ESP Game von der GWAP Webseite (`http://www.gwap.com/gwap/`

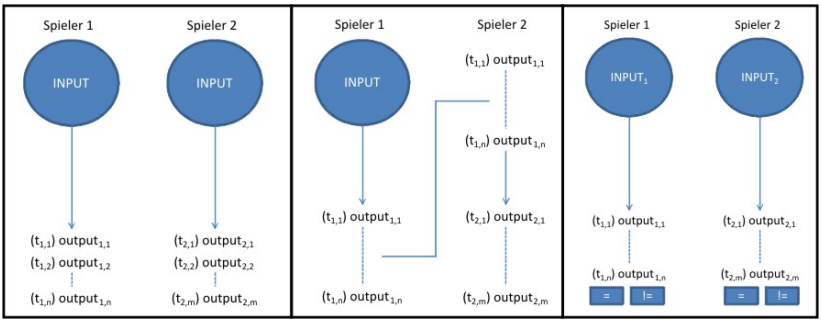


Abbildung 2.3.: Aus [20], drei Spielprinzipien für *Games with a Purpose (GWAP)*

Bei *Output-agreement games* geht es darum, dass zwei Spieler auf Grundlage desselben vorgegebenen Inputs denselben Ouput im Spiel produzieren. Der Output muss nicht zum selben Zeitpunkt erstellt werden, aber innerhalb einer vorgegebenen Zeitspanne. Das Spielsystem hat kein Wissen über den Input, sondern vergleicht nur den Output miteinander. Das *ESP Game* [19] ist ein prominentes Beispiel für dieses Spielprinzip.

Inversion-problem games sind Spiele, bei dem die Spieler unterschiedliche Rollen einnehmen - aus Gründen der Fairness abwechselnd. Es gibt einen „Beschreiber“ (*describer* [20]) und einen „Rater“ (*guesser* [20]). Der Beschreiber bekommt einen vorgegebenen Input und muss dem Rater durch das Bereitstellen eines entsprechenden Outputs klar machen, welchen Input er bekommen hat, ohne ihn explizit zu benennen. Das Spielsystem besitzt Wissen über den Input, möchte aber Wissen in einer anderen Form erhalten. Beispiele

sind die Spiele *Verbosity* [38], *Peekaboom* [39] und *Phetch* [40].

In *Input-agreement games* bekommen die Spieler einen Input angezeigt und müssen durch das Produzieren von aussagekräftigen Outputs entscheiden - meistens über textuelle Beschreibungen -, ob sie denselben Input bekommen haben. Das Spielsystem weiß in diesem Spielprinzip nur, ob der Input identisch ist oder nicht. Das Spiel *TagATune* [41] [42] folgt beispielsweise diesem Spielprinzip.

Ein zusätzliches Spielprinzip für GWAP-ähnliche Spiele wird in [43] angesprochen und zwar das Prinzip des *output-optimization*. Die Autoren führen das Spiel *The Restaurant Game* [44] als Beispiel auf. Die Spieler übernehmen in diesem Szenario entweder die Rolle von Gästen oder Bedienungen in einem Restaurant. Die Spieler sollen sich nun entsprechend ihrer Rolle verhalten. Das Verhalten - vor allem die Konversationen der Spieler - wird dabei aufgezeichnet und dient als Datengrundlage für autonome soziale Agenten.

Yuen et al. [43] verallgemeinern dieses Spielprinzip folgendermaßen: Die Spieler in einem *output-optimization* Spiel bekommen denselben Input und produzieren ihren individuellen Ouput, wobei der Output der jeweils anderen Spieler einem Spieler als Hinweis zur Verfügung steht. Anders als bei den anderen Spielprinzipien leitet sich hieraus aber noch keine direkte Gewinnbedingung ab, die als Motivation für die Spieler dienen könnte. D.h. es findet keinerlei Bewertung der einzelnen Outputs statt.

Neben diesen Spielprinzipien stellen Ahn und Dabbish [20] noch weitere Spielelemente vor, um speziell die Qualität des Outputs von Spielen zur Datenerhebung zu gewährleisten. Zu aller erst sollten Spieler immer zufällig für ein Spiel gepaart oder gruppiert werden (*random matching*). Damit soll verhindert werden, dass sie sich vorher absprechen und so das Ergebnis verfälschen. Des Weiteren sollen die Spieler mit bereits von anderen Spielern erhobenen Daten getestet werden (*player testing*), um ihre Glaubwürdigkeit zu bestimmen. Weiterhin sollten die erhobenen Daten dadurch abgesichert werden, dass ein bestimmter Output erst dann außerhalb des Spielkontextes übernommen wird, wenn er von mehreren Spielern unabhängig voneinander generiert wurde (*repetition*). Ein letztes Mittel ist die Angabe von Tabu-Outputs, d.h. von Daten, die schon in vorherigen Spielrunden erhoben wurden, und von den Spielern nicht noch einmal erhoben werden sollen.

In der Literatur finden sich zahlreiche Beispiele, in denen die genannten Spielprinzipien auf andere Problemstellungen angewendet werden. Hier ist vor allem das Prinzip des *Output-agreement game* stark vertreten, was wahrscheinlich auf die große Bekanntheit des *ESP-Game* zurückzuführen ist.

Einige Beispiele aus dem Forschungsfeld des Semantic Web will ich kurz vorstellen:

Die von Siorpaes und Hepp (2008) in [45] und [46] vorgestellte *OntoGame* Serie bildet laut den Autoren den gesamten Semantic Web Lebenszyklus ab, d.h. von der Erstellung von Ontologien (*OntoPronto*) über die semantische Annotation der Daten (*OntoTube* und *OntoBay*) bis hin zum Abgleich zweier oder mehrerer unterschiedlicher Ontologien (*SpotTheLink*). Zum Beispiel müssen zwei Spieler in *OntoPronto* zuerst anhand eines zufällig ausgewählten Wikipedia Artikels (wobei nur der erste Absatz angezeigt wird) entscheiden, ob dieser eine Klasse von Objekten oder ein Individuum beschreibt. Stimmt ihre Wahl überein, bekommen sie Punkte gutgeschrieben. In einem zweiten Schritt muss dieser Artikel dann in den spezifischsten Ast der PROTON Ontologie[9] von den Spielern eingeordnet werden. Bei gleicher Wahl gibt es wieder Punkte.

Krause und Aras (2009) gehen in ihrer Arbeit [47], der dem Bereich des *Social Tagging* zugeordnet werden kann, ähnlich vor. Ziel ist es, Webseiten (und Absätze innerhalb dieser) mit entsprechenden (semantischen) Stichwörtern (Tags) zu versehen. Das Spiel hat ebenfalls zwei Phasen und wird mit zwei Spielern gespielt. In der ersten übernimmt ein Spieler die Rolle des „Startspielers“, der einen zufälligen Abschnitt aus einer Webseite angezeigt bekommt. Diesen muss er dann mit einem einzigen Tag versehen.

In der zweiten Phase bekommen beide Spieler mehrere verschiedene - aber beide Spieler dieselben - Abschnitte aus der im ersten Schritt verwendeten Webseite angezeigt. Nun müssen sie entscheiden, welche der gezeigten Abschnitte zu dem vom „Startspieler“ vorgeschlagenen Tag gehören. Für jede Übereinstimmung bekommen beide Spieler Punkte gutgeschrieben. Nach dem Ende der Runde wird der andere Spieler dann „Startspieler“ und das Spiel beginnt mit einer anderen Webseite von vorne. Aus den Ergebnissen des Spiels werden sog. *Folksonomies* (von Thomas Vander Wal[10] eingeführter Begriff) zu Webseiten erstellt.

Im Spiel *Phrase Detectives* [48] von Poesion (2009) ist die Aufgabe der Spieler anaphorische Verbindung in vorgegebenen Texten zu annotieren, bzw. zu überprüfen. Dazu wird das Spiel in zwei Teile aufgeteilt. Im ersten Teil (*Annotation Mode* [48]) annotieren die Spieler die Wörter, die jeweils zu einer anaphorische Verbindung gehören. Im nachfolgenden zweiten Teil (*Validation Mode* [48]) bekommen die Spieler die Annotationen anderer Spieler

[9] `http://proton.semanticweb.org/`
[10] `http://vanderwal.net/folksonomy.html`

angezeigt. Diese können entweder angenommen oder abgelehnt werden, wobei es für beide Entscheidungen Punkte gibt. Wird eine Annotation allerdings angenommen, bekommt auch derjenige Spieler Punkte gutgeschrieben, der die Annotation ursprünglich erstellt hatte.

Ma et al. (2009) haben in ihrer Arbeit das Spiel *Page Hunt* in [49] vorgestellt, welches zum Ziel hat, Suchmaschinen zu verbessern. In *Page Hunt*, das alleine gespielt wird, bekommt der Spieler eine zufällig ausgewählte Webseite angezeigt, allerdings ohne deren URL. Er muss nun geeignete Suchanfragen formulieren, damit die Webseite innerhalb der ersten fünf Treffer in einer vom System bereitgestellten Suchmaske auftaucht. Schafft er das, bekommt er vom System Punkte gutgeschrieben. Neben dieser Einzelspielervariante stellen die Autoren noch eine kompetitive Variante (*PageRace*) vor, die ebenfalls nach dem *Output-agreement game* Spielprinzip [20] abläuft, und eine kollaborative Zweispielervariante (*PageMatch*). Letztere folgt dem *Input-agreement games* Spielprinzip [20].

Ein letztes Beispiel, welches zwar nicht direkt in den Bereich des *Semantic Web* fällt, aber zeigt, dass diese Form der spielerischen Datenerhebung durchaus mit traditionellen Formen der Verbesserung von algorithmischen Ansätzen mithalten kann, ist das *Video Store Clerk* Spiel[11]. Der *Cinematch*-Algorithmus wird nach [50] von Netflix[12] eingesetzt, um Nutzern entsprechend ihrer vorangegangenen Vorlieben, neue Filme vorzuschlagen. 2007 rief Netflix in diesem Kontext einen Wettbewerb ins Leben, bei dem das Ziel war, den *Cinematch*-Algorithmus um 10% zu verbessern. Neben zahlreichen Ansätzen, die z.B. auf statistischen Modellen beruhen, gab es einen Ansatz der mit der Idee des Human Computation arbeitete.

Jay Sandhaus und Andrew Bergmann (Team *JUST A GUY IN A GARAGE*) entwickelten dazu das *Video Store Clerk* Spiel. Im Spiel übernimmt der Spieler die Rolle eines Angestellten einer Videothek und muss nun in einer Spielrunde für zehn virtuelle Kunden, auf Grundlage von deren in der Vergangenheit ausgeliehenen und bewerteten Filmen, jeweils einen weiteren Film für diese Kunden bewerten. Als Grundlage nutzen Sandhaus und Bergmann, die Originaldaten aus der Netflix Datenbank. Bewertet nun der menschliche Spieler den Film genauso wie im System hinterlegt, bekommt er Punkte und kann bei insgesamt sieben richtigen Bewertungen in die nächste Runde voranschreiten. Durch diesen Spiel-basierten Ansatz konnte der *Cinematch*-Algorithmus um 8.667% verbessert werden, was einem 17. Platz in der

[11] `http://www.videostoreclerk.com/`
[12] `http://www.netflix.com/`

Endabrechnung entspricht[13]. Dies zeigt deutlich das Potenzial eines solchen „simplen“ Vorgehens.

In Kapitel 3 und 4 werde ich unter anderem zeigen, dass man mit einem Human Computation Ansatz nicht nur Rohdaten sammeln kann, um einen schon bestehende Algorithmen zu verbessern, sondern auch, dass durch die Erhebung entsprechend strukturierter Daten, neue Algorithmen möglich werden.

2.2. Nutzergemeinschaften in der GIScience

Geographische Informationssysteme (GIS) waren lange Zeit ein Werkzeug ausschließlich für Experten [51]. Auch die Daten, die mit GIS analysiert und bearbeitet werden, wurden nur teilweise einer breiteren Öffentlichkeit, bzw. geowissenschaftlichen Laien, zugänglich gemacht. Dies änderte sich mit dem Aufkommen von zwei neuen Teilgebieten in der GIScince Forschungslandschaft: *Public Participation GIS* (PPGIS) und etwas später *Volunteered Geographic Information* (VGI).

2.2.1. Public Participation GIS (PPGIS)

In der GIScience Forschung wurde schon früh über die Frage der Beteiligung (*participation*) der Öffentlichkeit (*Public*) in GIS-Systemen oder Prozessen nachgedacht, laut Tulloch (2008) [52] schon bei Chrisman in [53]) aus dem Jahre 1987. Dennoch wurde das Thema erst durch eine Reihe von Tagungen und Workshops Anfang bis Mitte der 1990er Jahre wirklich prominent. Der Begriff PPGIS selbst wurde nach [54] auf zwei Meetings des *National Center for Geographic Information and Analysis* (NCGIA) erstmals eingeführt. Das Ziel dieser Tagungen war es, der Frage nach der nächsten Generation des GIS, oder auch GIS/2, nachzugehen. Dabei sollte das GIS/2 den technischen Fortschritt in diesem Bereich im sozialen und politischen Kontext verankern. Zuerst eine eher nordamerikanische Angelegenheit, wurde mit einem Workshop zum Thema *Participation and Access*, welcher von der *European Science Foundation* (ESF) und der *National Science Foundation* (NSF) - amerikanisch - gemeinsam gesponsert wurde, das Thema auch in Europa breiter aufgestellt. In den folgenden Jahren hat sich so ein eigenständiges Forschungsfeld innerhalb der GIScience gebildet.

[13] `http://www.netflixprize.com//leaderboard`

Die Arbeiten von Dunn (2007) [55], Sieber (2006) [54] und Elwood (2006) [56] geben hierzu einen detaillierten Überblick. Des Weiteren bietet die eigens für diesen Themenkomplex eingerichtete Webseite[14] einen aktuellen Überblick über das Feld. An dieser Stelle möchte ich besonders die Aspekte des PPGIS ansprechen, die für meine vorliegende Arbeit von Interesse sind, bzw. meine Arbeit in den Kontext des PPGIS eingliedern.

Beteiligung und Öffentlichkeit sind zwei der zentralen Punkte bei der Betrachtung von PPGIS Anwendungen. Aber wie genau gestaltet sich die Beteiligung der Öffentlichkeit an PPGIS-Anwendungen? Und wer genau ist überhaupt mit „der Öffentlichkeit“ gemeint?

Die Beteiligung an Entscheidungsprozessen wurde von Arnstein in [57] schon im Jahre 1969 als hierarchisches Konstrukt, als sog. *eight rungs of a ladder* [57], betrachtet. Die Idee hinter dieser Sichtweise ist, dass je nachdem auf welcher Stufe der Leiter sich die Beteiligten befinden, um so mehr Macht haben sie an der Entscheidungsfindung. Schlossberg und Shuford (2005) [1] haben dieses Prinzip, zusammen mit anderen Ansätzen auf diesem Gebiet, mit einer Klassifizierung der Öffentlichkeit gekreuzt. Abbildung 2.4 zeigt die nach [1] resultierende Matrix.

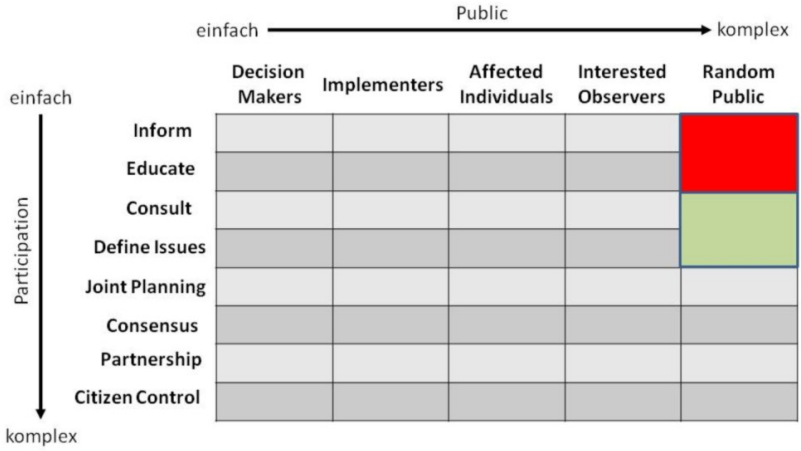


Abbildung 2.4.: Nach [1], mögliche Formen der Beteiligung (*Participation*) und eine Klassifizierung der Öffentlichkeit (*Public*)

Die Aufteilung der Öffentlichkeit zeigt auch gleichzeitig ein zentrales Ele-

[14] http://www.ppgis.net/

ment des PPGIS Ansatzes: Die Verschiebung der Machtverhältnisse von wenigen Experten zur breiten Masse der Nicht-Experten, welche ja auch hauptsächlich von den Entscheidungen der Experten betroffen sind. Dieser soziale Wandel und die Bemächtigung der zuvor Machtlosen durch dafür eingesetzte GIS Ansätze und Technologien wird dabei durchaus kritisch gesehen. Einerseits wird dadurch denjenigen Macht gegeben, die sie vielleicht gar nicht bekommen sollen oder wollen [58]. Andererseits werden gleichzeitig Personen ausgeschlossen, die nicht über die entsprechenden technischen Möglichkeiten bzw. Fähigkeiten verfügen das Machtinstrument PPGIS einzusetzen [54]. Diesen sozialen Aspekt werde ich aber in meiner Arbeit nicht weiter ausführen. Der interessierte Leser sei auf [56] oder [58] verwiesen. Der Vorteil meines spiel-basierten Vorgehens ist es, dass innerhalb eines Spieles die Realität und damit der vielleicht vorhandene politische Kontext, in dem das Spiel statt findet, ausgeblendet werden kann.

Der in dieser Arbeit vorgestellte Ansatz gliedert sich, in der in Abbildung 2.4 gezeigten Matrix, in die rot eingefärbten Matrixfelder ein. Ortsbezogene Spiele machen es meiner Meinung nach möglich, eine breite Öffentlichkeit zu erreichen, wobei deren Zusammensetzung meistens zufällig ist. Allerdings reduziert erstens der spielerische Charakter und zweitens die heterogene Zusammensetzung der einbezogenen Öffentlichkeit die Möglichkeiten der Beteiligung dieser am Spiel. Wie ich in Abschnitt 2.3 zeige, werden ortsbezogene Spiele schon im Zusammenhang raumbezogener Daten eingesetzt, um eine zufällige Öffentlichkeit (*random public* [1]) zu informieren (*inform* [1]) oder ihr etwas beizubringen (*educate* [1]). Potential sehe ich für meinen Ansatz auch in den grün eingefärbten Feldern in Abbildung 2.4.

Neben dem PPGIS beschäftigt sich noch ein anderes Feld innerhalb der GIScience ebenfalls mit Auswirkungen der Einbindung der Öffentlichkeit in einen GIS-gestützten Entscheidungsprozess: Das *Volunteered Geographic Information (VGI)*. Vor allem mit der Bereitstellung zahlreicher GIS Web 2.0 Anwendungen, wie Google Maps[15] und Google Earth[16] wurden hier zahlreiche Veröffentlichungen publiziert. Bisher ist allerdings noch nicht klar, inwieweit PPGIS und VGI in ihren Forschungsfragen und -lösungen überlappen. Klar ist nur, dass sie nicht identisch sind. Tulloch (2008) macht in [52] einen ersten Versuch die beiden Felder gegeneinander abzugrenzen und stellt fest, dass sich VGI mehr mit Anwendungen und der darin zur Verfügung gestellten raumbezogenen Information auseinandersetzt. PPGIS dagegen legt mehr

[15] `http://maps.google.de/`
[16] `http://earth.google.com/intl/de/`

Augenmerk auf den Prozess und das Endresultat, d.h. den Auswirkungen des Entscheidungsprozesses auf die Nutzer [52]. Elwood (2008) kommt zu einem ähnlichen Schluss [59], in dem sie festhält, dass:

> The existing literature on VGI suggests that an essential mechanism through which this phenomenon will impact society is the data themselves - their content and use; the technological structures used to obtain, store, and share them; and the situated knowledge claims and politics they may be used to advance. *[59], Seite 181*

Aus diesem Grund sehe ich meine Arbeit mehr im Feld des VGI, welches ich im nachfolgenden kurz vorstelle.

2.2.2. Volunteered Geographic Information (VGI)

Auf dem ersten Workshop zum Forschungsfeld des VGI[17] im Jahr 2007 wurde zuallererst einmal versucht, dieses noch junge Feld gegen die traditionellen Themengebiete (wie oben erwähnt auch zum PPGIS) der GIScience abzugrenzen, bzw. zu diskutieren, wie diese durch VGI betroffen werden.

Gleichzeitig wurde ein erster Versuch unternommen, der VGI eine eigene Identität zu geben. Wie der Name schon vorgibt, sind zwei Themenkomplexe von besonderem Interesse in VGI Forschung: die freiwilligen Teilnehmer (*Volunteered*) und die Daten, die sie bereit stellen (*Geographic Information*). Obwohl sich diese Begrifflichkeit in den letzten Jahren gefestigt hat, finden sich nach Elwood (2208) [60] auch unter anderen Namen Ansätze, die sich damit beschäftigen, wie man die neue Zugänglichkeit der Web-GIS Anwendungen nutzen kann, bzw. welchen Einfluss sie auf das Verständnis der Öffentlichkeit zu geographischen Fragestellungen haben. Der bekannteste ist dabei meiner Meinung nach der Begriff der *neogeography*, der nach Sui (2008) [61] von Turner (2006) in [62] eingeführt wurde. *Neogeography* macht laut [61] eine Geographie ohne Geographen, sprich Experten, möglich; durch die Einbindung von geowissenschaftlichen Laien.

Im Kontext meiner Arbeit ist das Bild des „menschlichen Sensors“ (*human sensor*) nach Goodchild (2007), in [9], [63] und [64] am Zutreffendsten als Bild für die Idee des VGI geeignet. Durch die breite und erschwingliche Verfügbarkeit von mobilen Geräten mit integrierter Lokalisierungstechnologie, wird es für Jedermann immer einfacher, Informationen über seine

[17] `http://www.ncgia.ucsb.edu/projects/vgi/`

persönliche oder öffentliche geographische Umwelt aufzunehmen und mit anderen zu teilen. Somit wird die Öffentlichkeit zu einem großen Sensornetzwerk, in dem jeder Einzelne die Möglichkeit besitzt, ein Knoten in diesem Netzwerk zu sein.

Wie bei der Betrachtung der PPGIS Literatur im letzten Abschnitt, will ich den interessierten Leser an dieser Stelle neben den Arbeiten von Goodchild vor allem auf die Arbeiten von Elwood - [60] und [59] -, Kuhn (2007) [14] und Craglia et al. (2008) [65] verweisen, falls man sich einen Überblick über das Feld des VGI verschaffen möchte.

Nachfolgend gehe ich auf zwei konkrete Fragestellungen im Detail ein und präsentiere einige prominente Beispiele für VGI, die im Zusammenhang mit meiner Arbeit stehen.

Genau wie die Frage nach dem unterschiedlichen Verständnis von Öffentlichkeit im PPGIS, stellt sich beim VGI die Frage, wer denn hier freiwillig als Sensor auftritt? D.h. wie kann eine mögliche Aufteilung der „Random Public" aus [1] aus Abbildung 2.4 aussehen? Coleman et al. (2209) gibt in [16] darauf eine Antwort. Die Autoren teilen dazu die *produser* [16] in fünf Klassen ein: *neophyte*, *interested amateur*, *expert amateur*, *expert professional* und *expert authority*. Diese unterscheiden sich vor allem auf Grundlage ihres theoretischen und praktischen Wissens in der Materie GIS. Die Einteilung hilft die unterschiedliche Motivation der Teilnehmer in VGI Anwendungen zu identifizieren (siehe ebenfalls [16]). Trotzdem bleibt für mich die Frage bestehen, wie man aus dieser heterogenen Öffentlichkeit Rückschlüsse auf die Qualität der von ihr bereitgestellten Daten ziehen kann.

Eine weitere Fragestellung des VGI ist: Wie man die Qualität der Daten beurteilen kann, die von einer so breit gefächerten Öffentlichkeit stammen?

Flanagin und Metzger (2008) schlagen in [66] vor, erst einmal zwischen Glaubwürdigkeit (*credibility*) und Genauigkeit (*Genauigkeit*) von VGI-basierten Daten zu unterscheiden. VGI besteht neben objektiven Daten, wie der Standort eines Restaurants, auch zu einem großen Teil aus subjektiven Daten, deren Qualität schwer messbar ist, z.B. ein Meinungsbild zu einem Bauprojekt in einer bestimmten geographischen Nachbarschaft. Daraus resultieren nach Ansicht der Autoren zwei Arten des Qualitätsbegriffes: einmal die *credibility-as-accuracy* und die *credibility-as-perception*. Letzteres hängt stark mit der Vertrauenswürdigkeit *believability* der Daten zusammen, d.h. ob die Nutzer in die Qualität der Daten Vertrauen haben oder nicht. Damit wird das subjektive Empfinden der Nutzer abgebildet. Ob die Daten auch objektiv gut sind, drückt der erste Begriff aus.

Den Ansatz, den ich in dieser Arbeit verfolge, versucht beide Sichtwei-

sen mit dem Begriff der Qualität zu verbinden. Indem die Qualitätssicherung in ein ortsbezogenes Spiel eingebunden wird, gibt man den Spielern erstens die Möglichkeit ihr Vertrauen in die Qualität, der von den anderen Spielern erhobenen Daten zum Ausdruck zu bringen. Gleichzeitig bietet sich je nach verwendetem Entwurfsmuster die Möglichkeit entsprechende objektive Qualitätsmaße zu berechnen. Dazu aber in Kapitel 3 Abschnitt 3.5 mehr.

Als Beispiel um die Qualität von nutzergenerierten Daten zu berechnen, möchte ich die Arbeit von Bashir et al. (2007/2008) ([67] und [68]) kurz vorstellen. Der Ansatz berechnet die Qualität von, wie sie sie nennen, *Collaboratively Contributed Geographic Information (CCGI)* auf Grundlage eines „trust and reputation model" ([68], Seite 1). Das Modell basiert auf der Annahme, dass raumbezogene Information, wie zum Beispiel Nachrichtenmeldungen, vertrauenswürdiger sind, wenn sie von einer Quelle stammen, oder bewertet werden, die sich geographisch nahe an der Stelle befindet, die zu der Information gehört. Zusätzlich spielt auch das Ansehen der Quelle innerhalb einer Nutzergemeinschaft eine Rolle. Der Wert der Qualität speist sich daher aus zwei Netzwerken, einmal einem sozialen Netzwerk, welches die Beziehung zwischen verschiedenen Akteuren abbildet, und einem Nachrichtennetzwerk, welches abbildet, welcher Akteur, welche Nachrichten erstellt oder bewertet hat. Für beides, den Akteur und für die Information, müssen Ortsangaben vorliegen.

Die Formel zur Berechnung der Qualität/Trust nach [68] für eine beliebige Nachricht t_{m_k} ist dann folgende:

$$t_{m_k} = \sum_{i=1,g=1}^{k} \frac{t_{n_g * r_{(n,m)}}}{log(c_i)} wobei c_i > 1 \tag{2.1}$$

$r_{(n,m)}$ ist hier die Bewertung einer Information m durch einen beliebigen Akteur n. Der Logarithmus $log(c_i)$ dient zur Normierung der geographischen Entfernung c_i zwischen der von einem Akteur bereitgestellten Information und dessen Standort.

Das Ansehen t_{n_g} eines Akteurs im sozialen Netzwerk berechnet sich wie folgt:

$$t_{n_g} = \frac{\sum_{n_i \in adj(n_g), i=1}^{N} (t_{n_i,n_g})(k_{n_i,n_g})}{N} \tag{2.2}$$

N ist die Menge aller Akteure, wobei sich t_{n_g} aus den Vertrauenswerten berechnet, die ihm/ihr von mit ihm/ihr verbundenen Akteuren im Netzwerk gegeben werden. t_{n_i,n_g} ist die Bewertung, die Akteur i gegenüber Akteur g

abgibt und k_{n_i,n_g} ist die soziale Stärke der Beziehung zwischen den beiden Akteuren.

Bashir et al. erweitern in ihrer Arbeit somit ein Vertrauensmaß mit einem Maß für geographische Nähe. Problematisch bleibt aber, wie die Vertrauenswerte t_{n_i,n_g} zwischen den Akteuren im sozialen Netzwerk erhoben werden. Wie motiviert man Nutzer eines solchen sozialen Netzwerkes, anzugeben, inwieweit er/sie einem Freund vertraut und ist dieser Wert nicht abhängig vom Kontext? Manchen Freunden vertraue ich bei der Frage, wo das beste Restaurant ist mehr als bei der Frage nach dem besten Arzt. Diese Fragen bleiben bisher unbeantwortet. Deswegen habe ich in meinem Ansatz die Qualitätssicherung direkt in den spielerischen Datenerhebungsprozess eingebunden (siehe Kapitel 3) und arbeite nicht mit einem auf Vertrauenswerten beruhenden Ansatz, obwohl solch ein Ansatz großes Potential birgt.

Ein bekanntes und typisches Beispiel für eine Anwendung im Bereich VGI ist die *Ushahidi* (Swahili für „Zeugnis“ („testimony“)) Plattform[18], welche auch schon in der medialen Berichterstattung ein breites Echo gefunden hat[19]. In 2008 als einfacher (Google Map) Mashup [69] gestartet, um Nutzern die Möglichkeit zu geben, Vorfälle von Gewalt in Kenya zu melden[20], wird es nun als eine Open-Source-Plattform angeboten. Wie Abbildung 2.5 zeigt, besteht solch eine Webseite aus einer Kartendarstellung, die mit entsprechenden thematischen Overlays versehen werden kann, und detaillierten Beschreibungen der kartographisch dargestellten Information. Bei dem abgebildeten Beispiel aus Kenya wurden alle gemeldeten Vorfälle vor der Veröffentlichung manuell überprüft.

Prinzipiell sind die meisten VGI Anwendungen ähnlich aufgebaut. Auf Grundlage von frei zur Verfügung stehenden Web-GIS Plattformen/Frameworks, wie Google Maps oder OpenStreetMaps [70], werden entsprechende frei zugängliche Mashups erstellt. Dazu wird die Kartengrundlage mit verschiedensten semantischen Informationen mit sog. *Overlays* überlagert. Die Überprüfung der eingestellten Information wird entweder von Experten übernommen, wie im Beispiel von *Ushahidi*, oder den Nutzern selbst überlassen, wie am Beispiel von OpenStreetMaps[21] geschehen.

Neben diesen reinen Web-basierten VGI Anwendungen, finden sich in der Literatur Beispiele von mobilen VGI Anwendungen, auch wenn sie zur Zeit ihrer Veröffentlichung noch nicht in dieses Forschungsfeld gezählt wurden.

[18] http://www.ushahidi.com/
[19] http://www.nytimes.com/2009/11/09/business/09link.html?_r=1
[20] http://legacy.ushahidi.com/
[21] http://www.openstreetmap.de/

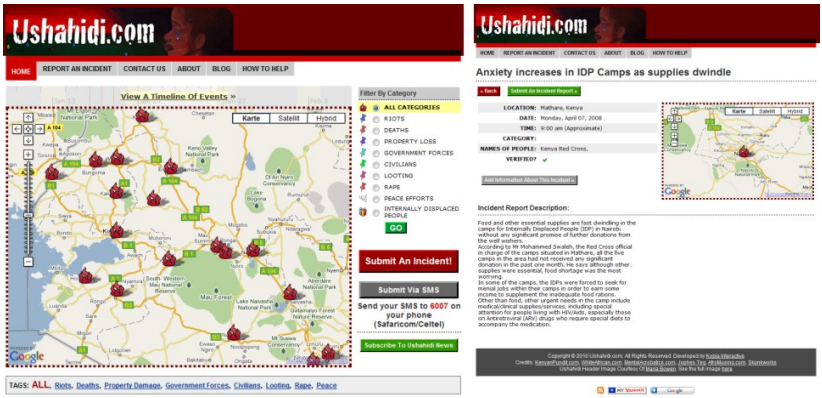


Abbildung 2.5.: Nutzung der *Ushahidi* Plattform zur Meldung von Gewalttaten in Kenya 2008, von `http://legacy.ushahidi.com/`

Eines der ersten Beispiele ist die *GeoNotes* Anwendung, die von Espinoza et al. (2001) in [71] vorgestellt wurde. *GeoNotes* erlaubt dem Nutzer mit Hilfe von mobilen Geräten, wie PDAs, mit einer geographischen Position versehene virtuelle Notizen (*notes*) in der realen Welt zu hinterlassen. Diese Notizen können z.B. ein Hinweis auf ein gutes Restaurant in der Nähe der Notiz sein oder auch aus einer Nachrichten für Freunde, die vielleicht später an der Stelle vorbei kommen, bestehen.

Das Designstudio *Proboscis*[22] geht in seiner Arbeit [72] noch eine Schritt weiter. Es lässt die Nutzer nicht nur simple Notizen zu einzelnen Orten erstellen, sondern versucht das soziale Verständnis, bzw. Wissen, zu Orten zu bekommen, die in einer Stadt von den Nutzern als interessant empfunden werden. Die Projekte *Urban Tapestries* und *Social Tapestries*[23] ließen die Teilnehmer unter anderem ihr lokales Wissen in sogenannten *threads* darstellen. Diese *threads* sind ein thematischer Pfad aus verbundenen Plätze in der persönlichen geographischen Umgebung. Die Pfade verbinden Plätze aus drei unterschiedlichen Kategorien: Politisch, sozial oder historisch. Abbildung 2.6 gibt ein Beispiel für *threads* aus den drei genannten Bereichen.

Die Arbeit [73] ist ein aktuelles Beispiel aus diesem Bereich des VGI. Bamford et al. (2007) stellen ein Anwendung vor, die es Reisenden ermöglicht,

[22] `http://proboscis.org.uk/`
[23] `http://urbantapestries.net/`

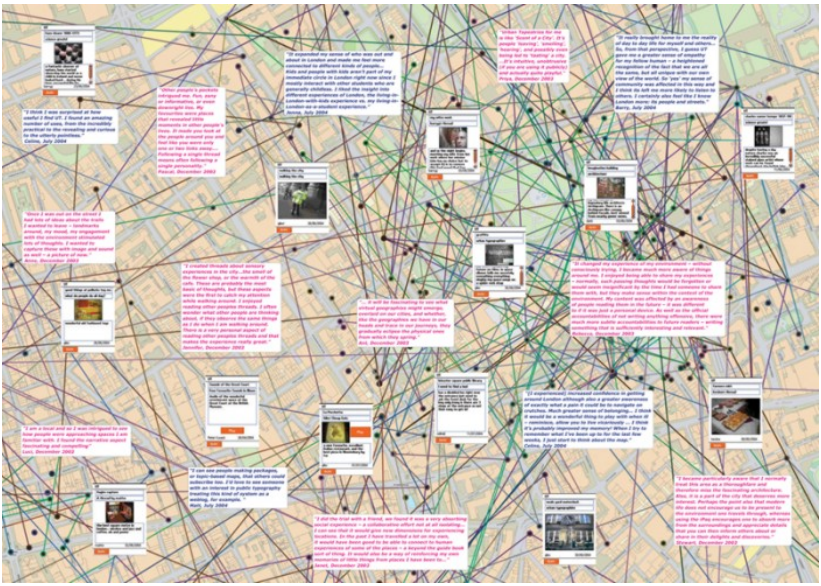


Abbildung 2.6.: Visualisierung der von den Teilnehmern in den *Urban* und *Social Tapestries* Projekten angelegten *political/historical/-social threads* - entnommen von `http://proboscis.org.uk/projects/urban-tapestries/`

mit einem mobilen Gerät, hier ein Mobiltelefon, aufgenommene Fotos vor Ort auf einen Blog hochzuladen und mit entsprechenden textuellen Bemerkungen zu versehen. Diese und zahlreiche ähnliche Arbeiten, siehe Yuen et al. (2009) [43], zeigen die große Beliebtheit solcher Foto-basierten Anwendungen.

Ein Lösungsansatz wurde allerdings von der GIScience Gemeinschaft übersehen, der auf dem AAAI Spring Symposium *KCVC05* 2005 prominent vertreten war: Die Nutzung von Spielen als natürlichem Motivator zur Einbindung der Öffentlichkeit in den Datenerhebungsprozess und die Qualitätssicherung. Vor allem ortsbezogene Spiele, die das Spielgeschehen vom Bildschirm in die reale Welt verlagern, sind hierfür ideal geeignet. Durch die Nutzung von Lokalisierungstechnologien wie GPS, sind auch die Voraussetzungen gegeben, durch entsprechende Spielelemente, spezifisch strukturierte raumbezogene Daten erheben zu lassen. Dazu gebe ich im nächsten Abschnitt einen Überblick über die aktuelle Forschungslandschaft im Bereich der orstbezogenen Spiele.

2.3. Ortsbezogene Spiele

Peltola und Karsten (2006) geben im folgenden Satz, den sie in ihrer Arbeit äußern, ein weit verbreitetes Vorurteil wieder:

> We authors, over the years, have felt many pangs of conscience as too much time is spent on games. (p. 6 [18])

Sie schwächen ihre Aussage aber ab, indem sie festhalten, dass

> playing in general is a necessary part of the experience of meaningful life. (p. 6 [18])

Mit diesem Satz fassen sie das Einführungskapitel des Buches „Homo Ludens“ von Huizinga (1971, ursprünglich aus dem Jahre 1939) zusammen: Das Spielen ist ein notwendiger Bestandteil des menschlichen Lebens. Huizinga identifizierte einige formale Eigenschaften, die Spiele mit magischen Ritualen, welche innerhalb eines „magischen Zirkels“ (*magic circle*) durchgeführt werden, gemeinsam haben.

Huizinga argumentiert, dass Spiele definiert werden durch den Ort, an dem das Spiel stattfindet, die Zeit, die ein solches Spiel dauert, und die Regeln, die während des Spieles gelten. Um am Spielgeschehen teilnehmen zu können, stimmen die Teilnehmer freiwillig dieser Änderung der Regeln ihres alltäglichen Lebens zu. Alle Menschen, die dies nicht tun, befinden sich automatisch außerhalb des Spieles. Auch die Spieler, die gegen Regeln verstoßen, sind somit nicht mehr Teil des sozialen Kreises des Spieles. Des Weiteren stellt Huizinga fest, dass keine materiellen Interessen mit dem Spiel verbunden sind und kein Profit, bzw. ein wie auch immer gearteter materieller Wert, daraus resultiert.

Ähnlich sieht es Caillois [74] (ursprünglich aus dem Jahre 1958) in seiner Definition des Spieles. Ein Spiel sei eine unproduktive menschliche Tätigkeit. Er unterscheidet dabei noch genauer zwischen dem Spiel an sich (*game = ludus*) und dem Schauspiel (*(free) play = paidia*). Im Deutschen ist der Begriff Spiel nicht ganz eindeutig, da beides unter dem Begriff „Spiel“ zusammengefasst werden kann. Beide Tätigkeiten erschaffen aber nach Caillois ebenfalls nichts von Wert für außerhalb des Spieles oder des Schauspiels.

Natürlich gelten die oben genannten Eigenschaften eines Spieles auch für digitale Spiele. Der Spieler bewegt sich während des Spieles in einem räumlich begrenzten Raum, indem er vor dem Bildschirm innerhalb seines Zimmers verweilt. Durch den Bildschirm ist es auch klar abgegrenzt von der

realen Welt Er spielt in zeitlich abgegrenzten Spielrunden und ist in seiner Interaktivität durch die Regeln begrenzt, die im Spiel gelten.

Die Abgegrenztheit von der realen Welt wird in dem Bild des „magischen Zirkels“ festgehalten. Dieser wurde 2004 durch das Buch „Rules of Play“ von Salen und Zimmerman (2004) [28] bekannt wurde (abgeleitet von dem Vergleich Huizinga des Spieles mit magischen Ritualen). Die Spieler betreten den „magischen Zirkel“ am Anfang des Spieles und verlassen ihn erst wieder, wenn das ganze Spiel oder eine Spielrunde vorbei sind. Durch den Simulationsaspekt, der digitalen Spielen nach Aarseth(2001) [75] inne wohnt, ist auch die Unproduktivität ein fester Bestandteil dieser Spiele. Anhand der fest implementierten Regeln kann eben nichts neues erschaffen werden. Auch nichts was auch von Wert für außerhalb des Spieles sein könnte. Der Zirkel erstreckt sich dementsprechend über eine räumliche, zeitliche, soziale Dimension. Wie lassen sich aber nun ortsbezogene Spiele von traditionellen digitalen Spielen abgrenzen?

2.3.1. Was ist ein ortsbezogenes Spiel?

In der Literatur finden sich zahlreiche Definitionen für ortsbezogene Spiele und wie sie sich von traditionellen, d.h. hier von digitalen Spielen unterscheiden. Außerdem muss festgehalten werden, dass selbst eine einheitliche Bezeichnung für diese neuartige Spielform in der Literatur noch nicht fest eingeführt ist. Es finden sich zahlreiche Begriffe, die teilweise parallel, synonym oder auch konträr verwendet werden.

Am häufigsten werden ortsbezogene Spiele unter dem Begriff des *Pervasive Game* [76] eingeordnet. Hierzu hat vor allem das von der EU geförderte Projekt IPerG (Integrated Project on Pervasive Gaming, FP6 - 004457) beigetragen, welches von 2004 bis 2008 durchgeführt wurde - siehe auch `http://www.pervasive-gaming.org` und [77].

Die dort verwendete Definition eines *Pervasive Game* beruht ebenfalls auf dem Konstrukt des „Magischen Zirkels“ nach Huizinga, bzw. Salen und Zimmerman. Nach Montola und Kollegen (2005/2009)- siehe [78] und [77] - sind den pervasiven Spielen alle Spiele zuzuordnen, die die drei oben genannten Dimensionen des „Magischen Zirkels“ erweitern, bzw. aufweichen: Die räumliche, die zeitliche und die soziale.

- **Räumlich**: Ein pervasives Spiel findet nicht nur „im Rechner“ statt, also innerhalb einer virtuellen Welt, sondern auch in der realen Welt. Der Spieler selbst wird zur Spielfigur. Er bewegt sein virtuelles Eben-

bild (Avatar) in der virtuellen Welt nicht mehr mit Hilfe eines Interfaces wie z.B. ein Spielecontroller, sondern er selbst wird zum Interface. D.h. seine Aktionen in der realen Welt haben einen direkten Einfluss auf die virtuelle Welt des Spieles. Die virtuelle Welt überlagert sozusagen die reale (z.B. in [79]). Damit verschwimmt die Grenze zwischen beiden, da sich die Spieler ja gleichzeitig in beiden befinden und auch mit beiden interagieren.

- **Zeitlich**: Durch die räumliche Ausdehnung des Spieles in die reale Welt, können Spiele in der zeitlichen Dimension nicht mehr exakt vom normalen Leben abgegrenzt werden. Pervasive Spiele können praktisch permanent laufen und Aktionen im Spiel jederzeit statt finden. Eines der ersten Spiele, die diese Aufweichung des „magischen Zirkels“ umsetzte, war *Killer* [80]. In diesem Spiel versuchen sich die Spieler gegenseitig mit „Spielzeugwaffen“ umzubringen. Dabei kennt jeder Spieler zwar sein eigenes Opfer, nicht aber wer sonst noch am Spiel teilnimmt, bzw. wer auf ihn als Killer angesetzt ist. Das Spiel selbst läuft 24 Stunden am Tag, solange bis nur noch ein Killer übrig ist. Dadurch ist man permanent der Gefahr ausgesetzt, von wildfremden Mitspielern „getötet“ zu werden. Die Trennung zwischen Spiel und Nicht-Spiel, sprich der Realität, wird somit außer Kraft gesetzt.

- **Sozial**: Durch die räumliche und zeitliche Aufweichung des „magischen Zirkels“, ist die Trennung zwischen Spielern und Nicht-Spielern ebenfalls nicht mehr klar zu erkennen. Wird in der realen Welt gespielt, ist bei manchen Spielkonzepten, wie in Steve Jackson „Killer“, nicht von vornherein ersichtlich, wer am Spiel teilnimmt und wer nicht. Auch können Nicht-Spieler von Spielern unfreiwillig in das Spielgeschehen miteinbezogen werden [81] [82]. Beispielsweise werden in dem Spiel *Uncle Roy all Around You* [83] Schauspieler eingesetzt, um die Spieler im Spiel voranzubringen. Doch kennen die Spieler nur das grobe Aussehen von diesen, d.h. die Spieler können auch unabsichtlich normale Passanten für Schauspieler halten und versuchen, diese in das Spielgeschehen zu integrieren .

Die materielle Komponente, die bei Huizinga oder auch Caillois angesprochen wird, findet bei Montola et al. keinerlei Erwähnung.

2. Gemeinschaftliche Datenerhebung

Nach Nieuwdorp (2007) [84] finden sich folgende Begriffe für das angesprochene Konzept der ortsbezogenen Spiele ebenfalls in der Literatur - wobei dies nur die aus meiner Sicht wichtigsten sind, in [84] sind noch weitere aufgelistet:

- Ubiquitous games [85]
- Augmented reality games [86]
- Mixed reality games [87]
- Mobile games [88]
- Alternate reality games [89]
- Crossmedia games [90]
- Location-based/aware game [91]
- Wireless Computer Game [92]

Alle diese Bezeichnungen werden nicht nur von unterschiedlichen Autoren benutzt, sondern einzelne Autoren verwenden auch unterschiedliche Begriffe für ein und dieselbe Arbeit. So ordnet Björk und Kollegen im Jahr 2002 ihre Spiele unter dem Begriff „Ubiquitious Gaming“ [93] ein, nur um dieselben Spiele 2007 als Datengrundlage zu verwenden, um für „Pervasiv Games“ Design-Entwurfsmuster herauszufinden. Nieuwdorp [84] stellt noch weitere Beispiele vor, wobei die Wahl des Begriffes, mehr vom Kontext der angestrebten Veröffentlichung abhängt, als von einer kohärenten Definition im Forschungsgebiet.

Aus diesem Grund gehe ich in meiner Arbeit von der oben genannten Definition von pervasive Spiele aus - und verweise hier nur auf weitere Arbeiten zu diesem Themenkomplex. Eine alternative Definition wird beispielsweise von Walther (2005) [94] vorgeschlagen und Brown (2007) [95] setzt sich eher kritisch mit dieser Begriffsvielfalt auseinander.

Zwei Einschränkungen sollen die Defintion von pervasiven Spielen noch etwas konkretisieren und sicherstellen, dass nur ortsbezogene Spiele darunter fallen. Diese habe ich in Kiefer et al. (2006) [96] vorgeschlagen. Und zwar besitzen danach ortsbezogene Spiele zwei Einschränkungen, die sie von pervasiven Spielen, wie sie in [77] definiert werden, abgrenzen, eine technologische und eine spielerische:

1. Ortsbezogene Spiele nutzen im Spiel eine oder mehrere Lokalisierungstechnologien, seien es globale wie GPS oder relative wie Bluetooth. Damit sind beispielsweise Spiele wie *Killer* keine ortsbezogenen Spiele, da sie ohne jegliche technologische Unterstützung auskommen.

2. Die physische Bewegung der Spieler im geographischen Raum wird als ein zentrales Element im Spiel eingesetzt, d.h. Spielaktionen können nur an bestimmten realen Orten im realen Spielgebiet ausgeführt werden und auch nur, wenn sich die Spieler physisch dort befinden. Dies deckt sich mit der räumlichen Erweiterung des „magischen Zirkels" von pervasiven Spielen. Damit fallen auch pervasives Spielzeug (*Smart Toys* [76]) und augmentierte Brettspiele (*Augmented Tabletop Games* [76]) nicht in die Kategorie der ortsbezogenen Spiele.

Die materielle Dimension des „magischen Zirkels" wurde in der bisherigen Diskussion zu ortsbezogenen Spielen nur am Rande angesprochen, soll aber mit dieser Arbeit in den Mittelpunkt gerückt werden. In Kapitel 3 werde ich auf diesen besonderen Aspekt genauer eingehen und stelle dort meinen Prozess zur Entwicklung von ortsbezogenen Spielen zur gemeinschaftlichen Datenerhebung vor.

Zuvor will ich aber ausgehend vom Stand der Forschung, noch auf die Herausforderungen zu sprechen kommen, die ein Entwickler von „traditionellen" ortsbezogenen Spielen zu beachten hat. Diese Herausforderungen spielen natürlich auch bei der Entwicklung von ortsbezogenen Spielen zur Datenerhebung eine Rolle.

2.3.2. Herausforderungen an das Design von ortsbezogenen Spielen

In der Arbeit von Benford et al. (2005) [97] wurden fünf Herausforderungen bei der Entwicklung von pervasiven und damit auch bei ortsbezogenen Spielen präsentiert: Umgang mit Ungenauigkeit (*dealing with uncertainty* [97]), hybride Architekturen (*hybrid architectures* [97]), gekoppelte Welten (*hefting domains* [97]), Spielvorbereitung (*configuration* [97]) und Spieldurchführung (*orchestration* [97]).

In meiner eigenen Forschung konnte ich noch drei weitere Herausforderungen identifizieren: Verteiltes Spielen (*geographic separation of players* [98]), die Integration von nicht-spielbezogenen Aufgaben (*integration of non-gaming tasks* [98]) und die Entwicklung von intelligenten ortsbezogenen KI-Gegnern [99].

Umgang mit Ungenauigkeit

Alle Lokalisierungstechnologien - eine Übersicht findet sich z.B. in der Arbeit von Rashid et al. (2006) [100] -, die in einem ortsbezogenen Spiel eingesetzt werden können, sind mit einer gewissen Ungenauigkeit behaftet. Diese Ungenauigkeit wird durch zwei Umstände verursacht: Der lückenhaften Abdeckung des Spielfeldes mit der gewählten Lokalisierungstechnologie. Vielleicht sind nur eine begrenzte Anzahl an WLAN-Einwahlknoten vorhanden. Und/oder dem Fehler in der Genauigkeit und Präzision, den die jeweilige Positionsberechnung mitbringt. Z.B. ist die Empfangsgüte des GPS Signals unter anderem von der Tageszeit abhängig und auch von der jeweiligen geographischen Umgebung - d.h. wird das Spiel eher in einer städtischen Umgebung oder einer ländlichen gespielt [101].

Benford et al. (2006) schlagen in [102] für diese Herausforderung fünf Lösungsstrategien vor, wie im Spielkontext mit dieser Ungenauigkeit umgegangen werden kann: Man kann sie beseitigen (*remove*), verstecken (*hide*), verwalten (*manage*), anzeigen (*reveal*) oder ausnutzen (*exploit*). Drozd et al. (2006) [103] gehen sogar noch einen Schritt weiter und schlagen vor, diese Strategien für alle möglichen *seams* [103], die bei pervasiven Anwendungen auftreten, anzuwenden. Allgemein gesprochen, ist ein *seam* [104] ein beliebiger Fehler in einem Computersystem, der eine fehlerfreie Interaktion zwischen Nutzer und System verhindert. Im Zusammenhang mit orstbezogenen Spiele, beschränke ich mich im Folgenden aber auf die Betrachtung von Lösungsstrategien für *seams*, die durch die dort verwendete Lokalisierungstechnologie verursacht werden [102].

Wenn man den Gedanken von Mark Weiser (1999) [105] folgt, hätte ein System ohne bemerkbare *seams* das Potential, für den Nutzer unsichtbar zu sein. Auf diese Weise könnte sich der Nutzer dann voll und ganz auf seine Aufgabe konzentrieren, die er mit Hilfe des System durchführen möchte. Im Kontext von Spielen wäre eine solche „Aufgabe“, die völlige Immersion [106] [107] in das Spielgeschehen, bzw. den *game flow* [108].

Um *seams* vollständig zu beseitigen, hat ein Entwickler von ortsbezoge-

nen Spielen die Möglichkeit immer auf die neuste Technologiegeneration zu setzen, z.B. bei GPS immer den neuesten Empfänger mit seinem Spiel zu unterstützen. Des Weiteren muss er für den Fall von schlechtem Signalempfang entsprechende Algorithmen einsetzen, die die entstehenden Fehler aus der Positionierung heraus rechnen. Technisch ist es aber zurzeit schwierig, *seams*, vor allem bei der Lokalisierungstechnologie GPS, vollständig zu beseitigen.

Um die am wahrscheinlichsten zu beobachteten Fehler vor dem Spieler zu verbergen, d.h. *seams* zu verstecken, muss die Benutzeroberfläche vom Spielentwickler entsprechend entworfen werden. Eine Idee ist zum Beispiel, dem Spieler seine Position nicht auf einem detaillierten Satellitenfoto anzuzeigen, sondern auf einer Detail ärmeren thematischen Karte. Da eine thematische Karte schon eine Abstraktion von der realen Welt darstellt, werden nur sehr grobe Fehler bemerkt. Man kann auch ganz auf eine Karte verzichten und die Spieler nur über das Erreichen eines relevanten Spielortes benachrichtigen - ohne dessen exakte Position oder Form visuell darzustellen. Im ortsbezogenen Spiel *Typhoon* [109] beispielsweise identifizieren GSM-Zellen-IDs spielrelevante Orte, ohne dass dessen exakte Ausdehnung dem Spieler angezeigt wird. Die Spieler wissen nur, ob sie sich gerade innerhalb der jeweiligen Zelle befinden oder nicht. Dies verbirgt die Tatsache, dass die Ausdehnung einer GSM-Zelle zeitlich schwankt und nicht immer gleich ist.

Bei der Verwaltung von *seams* geht es darum, eine Alternativstrategie zu haben, falls solch ein Fehler während des Spieles auftritt. Wenn beispielsweise für einige Regionen im Spielfeld kein GPS Signal zur Verfügung steht, könnte das Spiel den Spielern erlauben ihre Position selbst anzugeben - wie in [110] und [111] aufgezeigt. Eine etwas teurere Lösung im Hinblick auf die Organisation eines ortsbezogenen Spieles wäre es, menschliche Helfer einzusetzen, die das Spiel in Echtzeit überwachen und bei Bedarf eingreifen. In etwas abgewandelter Form wird diese Strategie auch in [112] eingesetzt. Bei LARP (Live Action Role Playing) Spielen überwachen mehrere Spielleiter das Spielgeschehen und leiten die Spieler durch das gesamte Spiel.

Ein Spielentwickler kann sich auch dafür entscheiden, alle vorkommenden *seams* den Spielern anzuzeigen. Eine Beispiel findet sich in [102]. In *Can You See Me Now?* wird die Position des Spielers nicht exakt (punktförmig) dargestellt, sondern kreisförmig, wobei sich der Radius des Kreises je nach Empfangsbedingungen vergrößert oder verkleinert. Solch eine Strategie wird auch bei der mobilen Variante von Google Maps angewendet[24]. Dadurch be-

[24] `http://www.google.de/mobile/maps/`

kommen die Spieler ein Gefühl dafür, wo und wie sich die Genauigkeit der eingesetzten Lokalisierungstechnologie auf ihr Spiel auswirkt. Dementsprechend können sie dann Strategien entwickeln, um diesem Phänomen entgegen zu wirken.

Die letzte Strategie ist es, auftretende *seams* auszunutzen, um daraus eine eigene Spielerfahrung zu entwickeln, d.h. diese in das Spielgeschehen als Spielelement zu integrieren. Diese Idee wurde erstmals von Chalmers et al. (2004/2005) [113] [114] in ihrem Spiel *Treasure/Bill game* umgesetzt. In diesem werden zufällig virtuelle Spielobjekte in einem Spielfeld verteilt. Dabei wird aber vorher nicht überprüft, ob die eingesetzte Lokalisierungstechnologie, hier über WLAN-Lokalisierung realisiert, an diesen Orten in der realen Welt es den Spielern auch möglich macht, diese aufzusammeln. Das WLAN wird aber nicht nur zur Positionierung der Spieler verwendet, sondern dient auch dazu, aufgesammelte Spielobjekte zum Spielserver zu schicken. Die Empfangsgüte des WLAN wird nun dadurch zum Spielelement, dass die Spieler belohnt werden, wenn sie Orte im Spielfeld finden, wo die Qualität der verwendeten Kommunikationstechnologie im Spiel (WLAN) gut ist. Werden Spielobjekte von Orten mit gutem WLAN Empfang hochgeladen, bringen sie dem Spieler mehr Punkte, als von Orten mit weniger Empfang. Solch ein Ansatz zur Entwicklung von Anwendungen/Spielen ist in der Literatur als *Seamful design* bekannt - siehe hierzu die Arbeiten von Barkhuus et al. (2005) [115], Broll et al. (2006) [116] oder Broll und Benford (2005) [117].

Hybride Architekturen

Die zugrunde liegende Architektur von ortsbezogenen Spielen kann, wie bei anderen Anwendungen auch, entweder als Client-Server oder Peer-to-Peer Architektur, oder als Kombination von beiden realisiert werden. Obwohl diese Herausforderung eher die technische Seite von ortsbezogenen Spielen betrifft, auf die ich hier nicht im Detail eingehen möchte (siehe hierzu die Arbeit von Greenhalgh et al. (2007) [118]), hat die gewählte Realisierung doch auch Auswirkungen auf das Spieledesign.

Verbunden mit der Frage, in welcher Architektur das ortsbezogene Spiel umgesetzt werden soll, ist die Frage, ob das Spiel einem synchronen oder einem asynchronen Spielprinzip folgen soll oder vielleicht sogar beiden, wie Saarenplaa et al. (2009) [119] vorschlagen. Dabei spreche ich von synchroner Spielweise, wenn der Spielablauf eines ortsbezogenen Spieles folgende zwei

Einschränkungen erfüllt:

1. Die Spieler nehmen simultan am Spiel teil, d.h. Interaktionen zwischen den Spielern finden nur statt, wenn die Spieler gleichzeitig am Spielgeschehen teilnehmen.

2. Jede Spielaktion wird in Echtzeit an alle betroffenen Spieler kommuniziert.

Damit besitzen alle Event-basierten ortsbezogenen Spiele, nach der Definition von Paelke et al. (2008) [120], wie beispielsweise *Can You See Me Now?* [102], *Human Pacman* [121], *PAC-LAN* [122], *Uncle Roy All Around You* [123] oder *CityPoker* [99], eine synchrone Spielweise. Ebenso beinhalten auch Spiele mit einer offenen Spielwelt (*open-world games*), in der die Spieler in das Spiel zu jedem Zeitpunkt aus- und wieder einsteigen können - z.B. bei Botfighters [79] [124] [125], Alien Revolt [91], or Songs of North [126] - eine synchrone Spielweise.

Das asynchrone Spielprinzip ist vor allem in letzter Zeit mehr in den Fokus gerückt - siehe z.B.die Arbeiten von Flintham und Kollegen (2007) [127]. Die grundlegende Idee ist es hierbei, Spielern die Möglichkeit zu geben, das Spiel zeitlich getrennt zu spielen - meistens mit Hilfe eines Servers als Zwischenschicht. Über diese Zwischenschicht werden dann die von den Spielern getroffenen Spielaktionen indirekt an die betroffenen Spieler vermittelt, sobald sie dem Spiel beitreten. Die indirekte Interaktion zwischen den Spielern kann entweder ganz simpel realisiert sein, z.B. über eine Server-basierte Rangliste, wie im Spiel *Feeding Yoshi* [128], oder komplex, z.B. über eine komplette Community Webseite, wie im Spiel *Gopher* [129] oder *MobiMission* [2].

Gekoppelte Welten

In einem ortsbezogenen Spiel wird die reale Welt mit einer virtuellen überlagert. Die zentrale Frage, die sich dadurch stellt, ist: Welche Spielelemente werden virtuell und welche werden in Form eines Realwelt-Objektes oder sogar beidem repräsentiert.

Meines Wissens wurde dies durchaus interessante Fragestellung bisher noch nicht systematisch untersucht. Einzig Cheok et al. (2004) haben in ihrer Evaluierung des Spieles *Human Pacman* [121] berichtet, dass sich die Verwendung von Realwelt-Objekten positiv auf die Spielerfahrung der Teilnehmer

ausgewirkt hat. Von den 23 befragten Spielern gaben 78.3% an, dass „graspable interaction offered by the collection of real objects enhances the game (p. 80).“ Im Spiel wurden die „Kraftpillen“, die es Pacman erlaubt, die ihn verfolgenden Geister zu fressen, mit einer kleinen Truhe repräsentiert. In dieser befand sich ein Bluetooth-fähiges Gerät, welches die Spieler nutzen konnten, um den Effekt der „Kraftpillen“ zu aktivieren.

In der Tat verwenden nur wenige ortsbezogene Spiele Realwelt-Objekte in ihrem Spielablauf. Ein Beispiel ist das Spiel *CityPoker* [99]. Hier werden reale Pokerkarten verwendet und im Spielfeld versteckt. Diese müssen von den Spielern mit Hilfe von mobilen Geräten und GPS Lokalisierung innerhalb von fünf vorher festgelegten Spielfeldregionen gefunden werden.

Ein anderes Beispiel ist *Insectopia* [130], in dem jedes Bluetooth-fähiges Gerät als Spielobjekt verwendet wird. Die Aufgabe der Spieler ist es, mehr Insekten zu sammeln als ihre Mitspieler. Insekten sammeln die Spieler durch das Verbinden mit beliebigen Bluetooth-fähigen Geräten, die sie in ihrer alltäglichen Umgebung finden,.

Spielvorbereitung und -durchführung

Die Spielvorbereitung beinhaltet nicht nur das Problem, wie man am Besten ein orstbezogenes Spiel für eine geographische Umgebung konfiguriert, sonder auch wie man das Spiel am Besten auf eine andere Umgebung überträgt. D.h. welche Orte wähle ich als spielrelevante Orte aus, an denen die Spieler Spielaktionen durchführen können, bzw. wie übertrage ich diese geographische Anordnung von Spielorten auf eine andere geographische Umgebung.

Bei der Durchführung geht es um die Überwachung des Spielgeschehens in Echtzeit, möglichst aber ohne Eingriff in den eigentlichen Spielverlauf.

Zur Spielvorbereitung werden in der Literatur zahlreiche Softwarewerkzeuge vorgeschlagen. Dies beschäftigt sich hauptsächlich mit der Fragestellung, die bei der Vorbereitung eines ortsbezogenen Spieles, und auch bei der Entwicklung, eine entscheidende Rolle spielt: Wo im Spielfeld habe ich für meine eingesetzte Lokalisierungs- und Kommunikationstechnologie ausreichend Empfang? Davon hängt dann entscheidend ab, wo die spielrelevanten Orte im Spielfeld gesetzt werden können.

Das *ColourMaps* [131] Werkzeug erlaubt es Designern von ortsbezogenen Spielen digitale Repräsentationen des Spielfeldes, die aus Satellitenbildern

oder thematische Karten bestehen, mit farbigen Overlays zu versehen. Der Farbwert der Pixel kann dabei je nach Art des Overlays unterschiedliche Information repräsentieren, z.B. GPS Empfangsbedingungen oder auch Spielinhalte.

Oppermann et al. (2006) [109] argumentieren, dass auf Grund der besonderen Struktur von ortsbezogenen Anwendungen, zu denen auch Spiele zählen, beim Design dieser Anwendungen drei Schichten an Information wichtig sind: Der *physical environment layer*, der eine Repräsentation der geographischen Umgebung darstellt. Der *infrastructure layer*, als Repräsentant der im Spielfeld zur Verfügung stehenden Infrastruktur bezüglich der Lokalisierungs- und Kommunikationstechnologie, und dem *content layer*, der die Anwendungsinhalte repräsentiert.

In [132] werden die zwei oben genannten Ansätze in einen Designprozess zusammengefasst, unter dem Stichwort „artist-led design“. Die Idee ist, geowissenschaftlich nicht versierten Designern, wie z.B. Künstlern, geeignete Werkzeuge an die Hand zu geben, damit Fragen der Spielvorbereitung von diesen einfach beantwortet werden können. Der vorgestellte Prozess gibt aber nur Hilfestellung bei den technischen Aspekten der Spielentwicklung und nicht beim Design der im Spiel eingesetzten Spielelemente.

Ein weiteres Werkzeug, welches zur Spielvorbereitung verwendet wird, ist das Geogame Tool [133]. Mit diesem kann die z.B.in CityPoker [99] oder GeoTicTacToe [134] notwendige Synchronisationszeit (*syncTime* [134]) je nach Konfiguration des Spielfeldes berechnet werden. Die Synchronisationszeit ist nur für ortsbezogene Spiele wichtig, die der Definition in [134] genügen und damit als *Geogames* zu klassifizieren sind. Somit ist das Werkzeug als spielspezifische Lösung zu bewerten.

Für die Spieldurchführung finden sich in der Literatur meines Wissens nur spielspezifische Lösungen. Prominente Beispiele sind Werkzeuge für die Spiele *Epidemic Menace* [90] und *Savannah* [135]. In *Epidemic Menace* wird eine Kombination von Schauspielern und einer computer-gestützen Überwachung aller Spielzustände verwendet, um das Spiel durchzuführen.

Das für *Savannah* entwickelte computer-gestützte System wird dort nicht nur zum Monitoring eingesetzt, um z.B. die Bewegung der Spieler live darzustellen, sondern auch als Spielelement. Zwischen zwei Spielrunden wird das System genutzt, um den Spielern zusätzliche Lerninhalte zu vermitteln. Dieses zusätzliche Wissen kann von den Spielern dann in nachfolgenden Spielrunden eingesetzt werden. Wie man sieht, existieren zwar Softwarewerkzeuge, um Spiele in Echtzeit zu überwachen [87], bisher wurde aber noch kein allgemeines Framework vorgeschlagen.

Verteiltes Spielen

Um ein ortsbezogenes Spiel z.B. in meiner Heimatstadt zu spielen, brauche ich für die allermeisten Spiele weitere menschliche Mitspieler. Aber was mache ich, wenn ich gerade in meinem Ort keine oder nicht genügend Mitspieler finde? Kann ich dann gegen Freunde aus einer anderen Stadt spielen? In dieser Herausforderung geht es darum, dass Spieler aus unterschiedlichen Städten miteinander spielen können, auch wenn sie sich nicht in demselben Spielfeld aufhalten. D.h. die geographisch verteilte Spielfelder in den jeweiligen Städten müssen logisch so aufeinander abgebildet werden, dass es von den Spielern als ein einziges angesehen wird.

Ich werde in Kapitel 3 im Abschnitt 3.6.3 noch einmal genauer auf diese Herausforderung eingehen. Kiefer et al. (2007) machen in [4] klar, dass der kritische Aspekt eines verteilten Spieles der ist, festzulegen, wie Spieler in den aufeinander abgebildeten Spielfeldern Spielaktionen ausführen können. Dazu unterscheiden sie zwischen Spielfeldern in denen Spielaktionen überall durchgeführt werden können und solchen wo dies nur an ausgewählten Orten möglich ist. Weiter identifizieren sie drei Quellen warum Spielfelder als unterschiedlich empfunden werden und welche die Abbildung aufeinander schwierig macht: Das räumliche Ausmaß (*spatial scale*), die statischen (*static nature*) und die dynamischen Merkmale (textitdynamic conditions) des Spielfeldes. Falls diese Unterscheidungsmerkmale nicht in irgendeiner Art und Weise im Spieldesign behandelt werden, wird das resultierende ortsbezogene Spiel mit Sicherheit als unfair von den teilnehmenden Spielern empfunden werden.

Beispiel für verteilte Spiele, die aber nur einen begrenzten Ortsbezug und minimale Bewegung der Spieler aufweisen, finden sich bei Müller und Kollegen (2004/2006) [136] oder [137]. In *Airhockey* und anderen Spielen der Autoren werden z.B. zwei Tische eines Airhockeyspieles aufeinander abgebildet. Die Bewegung der Spieler wird nur indirekt über die Bewegung des Pucks auf das Spiel übertragen.

Entwicklung von intelligenten ortsbezogenen KI-Gegnern

Für den Bereich der Künstlichen Intelligenz in traditionellen Videospielen, findet sich eine große Anzahl an Ansätzen, die in der Praxis angewendet werden - die Bücherreihe [138] gibt einen guten Überblick dazu. Meines Wissens

existiert aber bisher nur eine einzige Arbeit, die sich mit der Entwicklung von KI-Gegnern für ortsbezogene Spiele beschäftigt. In der Datenbank für ortsbezogene Spiele, die ich für meine Klassifizierung ortsbezogener Spiele (siehe Kapitel 3) erstellt habe, findet sich ebenfalls nur ein Spiel, welches KI-Gegner enthält: *ARQuake* [86]. Dieses Spiel ist eine direkte Umsetzung des Computerspielklassikers *Quake* und besitzt dieselben KI-Routinen wie das Original. Alle anderen mir bekannten ortsbezogenen Spiele, sind solche, die mit einem oder mehreren menschlichen Spielern, gegen oder miteinander, gespielt werden müssen. Jakl (2006) präsentiert in [139] das Einzelspieler Abenteuerspiel *Journey*.

Aus diesem Grund schlagen Kiefer et al. (2005) [99] vor, KI-Gegner zu entwerfen, die exakt dem Verhalten von anderen menschlichen Spielern entsprechen. Diese virtuellen „Spielerkopien“ könnten dann auch dazu eingesetzt werden, Spieler asynchron gegen andere Spieler antreten zu lassen oder diese als Probegegner vor einem Spiel gegen das menschliche Gegenstück einzusetzen. Dadurch könnte auch die von den Autoren vermutete geringere Motivation, gegen KI-Gegner in einem ortsbezogenen Spiel anzutreten, erhöht werden. Da die dort geäußerten Bemerkungen aber nur theoretischer Natur sind und auch von neueren Studien des Spieles *Alien Revolt* von Silva (2008) [91] in Frage gestellt werden, wären detailliertere Studien zu dieser Herausforderung erstrebenswert.

Integration von nicht-spielbezogenen Aufgaben

Jedes ortsbezogene Spiel muss mit einer Restriktion in seinem Design auskommen. Seien es Restriktionen, die der geographischen Umgebung geschuldet sind, oder es gibtbspezielle Designziele, die eingehalten werden müssen. Ein solches Designziel kann die Einbindung von Spielelementen zur Erhebung von raumbezogenen Daten darstellen.

Ein Beispiel für das Beachten von geographischen Restriktionen sind die Arbeiten von Kiefer et al. (2007) [140] und Chen und Benford et al. (2007) [141]. In beiden Arbeiten werden ortsbezogene Spiele entworfen, die speziell dafür gedacht sind, dass sie auf einem linearen Weg von einem festen Ausgangspunkt bis zu einem ebenfalls fixen Endpunkt gespielt werden. Die Spieler könnten beispielsweise auf einer Fahrrad- oder Wandertour unterwegs sein. Auf der sind sie nicht gewillt, den vorgegebenen Weg zu verlassen. Dadurch, dass sich die Spieler nur in eine Richtung bewegen, muss das Spielfeld und das Spiel selbst entsprechend entworfen sein.

Kiefer et al. (2007) [140] schlagen vor, mit Hilfe von raumfüllenden Kurven [142] die in ortsbezogenen Spielen verwendeten zweidimensionalen Spielfelder zu linearisieren. Zusätzlich zur linearen Anordnung der spielrelevanten Orte, muss das lineare Spielfeld auf den Weg so abgebildet werden, dass es möglich ist, sich so auf dem Spielfeld zu bewegen, dass man die spielrelevanten Orte trotzdem in einer beliebigen Reihenfolge abarbeiten kann, ohne sich vor und zurück zu bewegen. Um dies möglich zu machen, bilden die Autoren das lineare Spielfeld einfach mehrmals hintereinander auf dem Weg ab. Alle Spielfelder repräsentieren logisch ein und dasselbe Spielfeld, d.h. führe ich eine Spielaktion an einem Ort im ersten Spielfeld aus, ist diese Spielaktion im darauf folgenden Spielfeld immer noch aktiv. Somit kann ich auf dem ersten Spielfeld auch einen spielrelevanten Ort überspringen und erst im nächsten Spielfeld zu einem späteren Zeitpunkt besuchen. Mit dem in [140] vorgestellten Simulationsframework kann zusätzlich noch berechnet werden, wie oft man das Spielfeld für eine gegebene Wegstrecke hintereinander anordnen sollte, damit ein spannendes - nach der Definition der Geogames aus [134] - Spiel entsteht.

Neben diesem allgemeinen Ansatz, der nicht nur auf das von den Autoren vorgestellte Spiel *Alak* [143] beschränkt ist, finden sich in der Literatur noch weitere Ansätze wie in [144] oder auch [145]. Bichard und Kollegen (2006/2006) beschäftigen sich mit ortsbezogenen Spielen, die auf dem Rücksitz eines Autos gespielt werden. Unter anderem untersuchen sie den Themenkomplex des Storytelling für solche Spiele [144]. Da die Bewegungsfreiheit der Spieler aber durch das gewählte Fortbewegungsmittel eingeschränkt ist, eignen sich diese Art der Spiele eher nicht, um damit raumbezogene Daten sammeln zu lassen.

Ein Designziel, welches entscheidenden Einfluss auf Spielelemente eines ortsbezogenen Spieles haben kann, ist die Einbindung von Lerninhalten. Das in der Literatur vor allem als *Serious Gaming* [146] bekannte Forschungsgebiet - bei dem es darum geht Lernen mit Spielen zu motivieren -, bekommt mit dem Ortsbezug eine ganz neue Dimension hinzu. Beim mobilen Lernen („Mobile Learning“ [147]) steht aber nicht das Lernen unterwegs im Vordergrund [148] [149], sondern vielmehr, dass die Lerninhalte direkt vor Ort erfahren werden. Kiefer et al. (2006) [6] argumentieren, dass der Umstand, dass sich die Lernenden zum Zeitpunkt des Lernens innerhalb der geographischen Umgebung aufhalten, die mit den Lerninhalten direkt verknüpft ist, erhöht das Lernpotential (*spatial/temporal grounding* [6] [150]) ohne den Spielspaß zu verringern.

Eine einfacher in das Spieldesign einzuarbeitende Zielvorgabe ist die Inte-

gration von touristisch interessanten Inhalten oder auch Orten. Ein prominentes Beispiel findet sich in den Arbeiten von Walz und Kollegen (2006/2007) [151] und [152]. Das Spiel *REXplorer* [152] wurde speziell für die Stadt Regensburg entwickelt, um Touristen die Stadt spielerisch erkunden zu lassen.

Natürlich finden sich in der Literatur noch zahlreiche weitere solcher speziellen Designziele, wie beispielsweise die Verbindung von ortsbezogenen Spielen mit der persönlichen Gesundheitsvorsorge [153]. Ich will im nächsten Abschnitt ein Designziel detaillierter ansprechen, das im Zusammenhang mit meiner Arbeit von großem Interesse ist: die Vorgabe mit ortsbezogenen Spielen raumbezogene Daten zu erheben.

2.3.3. Ortsbezogene Spiele zur Datenerhebung

Seit längerem ist man sich vor allem in der Forschung zu pervasiven Spielen bewusst, dass

> „game play is a natural motivator to participate in something that is not necessary or beneficial. [...] By controlling game events and perhaps game logics and rules, the agencies that ultimately use the gathered data, can steer players to do tasks supporting their needs“. (p. 7 [18])

Konzeptionelle Ansätze, wie dieser Anspruch in die Praxis umzusetzen wäre, finden sich in einigen Arbeiten zu diesem Thema. Capra et al. (2005) waren eine der ersten, die sich in [154] mit der Erhebung von raumbezogenen Daten beschäftigten. Das *Hitchers Framework* [103] beruht auf der einfachen Idee, dass Spieler virtuelle Objekte irgendwo in der realen Welt aufsammeln und an anderer Stelle wieder ablegen können. Wie von den Autoren angemerkt, ist dies eine Weiterentwicklung des Spielkonzeptes des *Treasure/Bill* Spieles von Chalmers und Kollegen (2004/2005) - siehe hierzu [114] und [113]. In diesem Spiel wurden auch schon zusätzliche Daten, Informationen zur eingesetzten Lokalisierungs- und Kommunikationstechnologie, während des Spieles aufgezeichnet. Im Gegensatz zum *Hitchers Framework* werden dort aber die virtuellen Objekte nur in einem begrenzten Spielfeld verteilt und können dieses auch nicht verlassen, was das Sammeln von raumbezogenen Daten auf diesen geographischen Raum beschränkt.

Das Framework wurde in den folgenden Jahren zur Entwicklung von zwei Spielen eingesetzt, die interessant für meine Arbeit sind: *MobiMission* [2] und das *Gopher* [129] Spiel. Beide folgen demselben Spielprinzip: Die Spieler

produzieren selbst die Spielinhalte. Später schlugen Wolff und Grüter (2008) den Begriff *gamers as producers* [155]) für solche Spiele vor. In beiden Spielen geht es darum, dass die Spieler Miniaufgaben erstellen - einmal „Missions“ und einmal „Gopher“ genannt -, die die anderen Spieler dann spielen können. In den von den Spielern erstellten Miniaufgaben geht es meistens darum Bilderrätsel zu lösen oder virtuelle Objekte von einem Ort zu einem anderen zu transportieren. Das liegt vor allem daran, dass den Spielern nur die Möglichkeit geboten wird, die Aufgabe anhand eines Photos und einiger Zeilen Text zu beschreiben. Die Missionen werden ortsbezogen verteilt und aufgesammelt. D.h. in beiden Spielen werden sie mit einer GSM-Funknetzzelle - mit der entsprechenden Zell-ID - assoziiert und können dann auch nur dort aufgesammelt werden. Die Aufgaben selber müssen aber nicht zwangsweise orstbezogen durchgeführt werden, sprich da wo man sie aufgesammelt hat.

Wahrscheinlich auch deshalb haben beide Arbeiten gezeigt, dass ein sehr großer Teil der erstellten Miniaufgaben an sich keinen ortsbezug im Sinne eines ortsbezogenen Spieles aufwies. Die Spieler mussten also gerade nicht physisch ihre Position verändern, um die Aufgabe zu lösen. Auch wurde der Ortsbezug beim Aufsammeln der Minaufgaben dadurch abgeschwächt, dass, wenn in einer Zelle keine Miniaufagben mehr vorhanden waren, man auch Miniaufagben von benachbarten Zellen aufsammeln konnte. Dies wurde nötig, da besonders zu Beginn des Spieles nur wenige Missionen für die Spieler zur Verfügung standen und langes Suchen vermieden werden sollte.

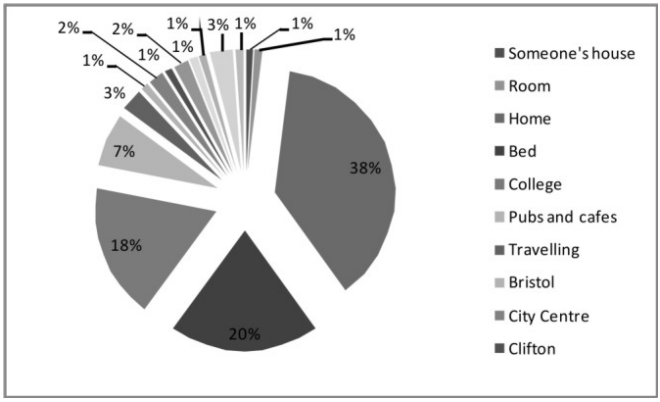


Abbildung 2.7.: Nach [2], der Anteil an gespielten Miniaufgaben in Bezug auf den Spielort im Spiel *MobiMissions*

Abbildung 2.7 zeigt, dass sich im Spiel *MobiMissions* von [2] - für [129] liegt eine solche Auswertung leider nicht vor - 75% der Miniaufgaben auf genau drei Spielorte verteilen. Somit ist natürlich die Intention, mit diesen Spielen raumbezogene Daten für beispielsweise eine Stadt erheben zu wollen, schwierig. Wenn man ihnen die Wahl lässt, scheinen die Spieler eher geneigt zu sein, das Spiel vom Bett aus zu spielen, als draußen.

Ein Schritt in die richtige Richtung ist meiner Meinung nach das Spiel *Feeding Yoshi* welches in [128] vorgestellt wurde. Die Aufgabe der Spieler ist es, in ihrer alltäglichen geographischen Umgebung offene und geschlossene WLAN-Knoten zu finden, um entsprechende Spielaktionen durchzuführen. Auf diese Weise sind die Spieler „gezwungen“ sich in der realen Welt zu bewegen und darüber Daten zu erheben.

Aufbauend auf den hier erwähnten Ansätzen in den Forschungsgebieten des PPGIS 2.2.1, der VGI (Abschnitt 2.2.2), des Human Computation (Abschnitt 2.1) und der ortsbezogenen Spiele (Abschnitt 2.3), stelle ich im nächsten Kapitel meinen Ansatz zum Design solcher Spiele vor. Die dort präsentierten Ansätze und Ergebnisse finden auch teilweise schon in Arbeiten anderer Autoren Erwähung, siehe beispielsweise [156], [157] oder [155], auf die ich in den folgenden Kapiteln an geeigneter Stelle eingehen werde.

2.4. Semantische Datenintegration

Der Fokus der vorliegenden Arbeit liegt in der qualitätsgesicherten Erhebung von raumbezogenen Daten. Die Daten werden aber nicht nur ein einziges Mal erhoben, sondern sind in der damit aufgebauten Datensammlung im Normalfall mehrmals vorhanden. Durch die Ungenauigkeit in den eingesetzten Lokalisierungstechnologien, beispielsweise GPS, sind diese Daten aber nicht vollständig identisch, sondern nur bis zu einem gewissen Grad. Um zu bestimmen, ob eine Aggregation der redundanten Datensätze zu einem neuen Datensatz sinnvoll ist, muss zuerst die Ähnlichkeit zwischen den Datensätze berechnet werden. Ansätze zur Berechnung entsprechender Maße, finden sich in der Literatur zahlreich. An dieser Stelle kann ich daher nur auf die im Kontext der vorliegenden Arbeit relevanten eingehen.

Ausgehend von der Forschung durch Schwering (2008) [158] - die sich an [159] orientiert - gibt es zwei maßgebliche Ansichten, wie man den Begriff der Ähnlichkeit betrachten kann. Einmal kann man anhand der *Gemeinsamkeiten und Unterschieden* von zwei Konzepten oder Ausprägungen dieser Konzepte - d.h. die einzelnen Datensätze selbst, nachfolgend auch Objek-

te genannt - die Ähnlichkeit zwischen diesen festlegen. Zum Anderen kann die *semantische Distanz* zwischen Konzepten oder Objekten als Maß für die Ähnlichkeit angelegt werden. Alle Konzepte/Objekte müssen dazu im selben Framework mit derselben Metrik repräsentiert werden. Die Distanz zischen den Konzepten/Objekten, die in einem solchen Framework verortet werden, kann auf einer räumlichen Entfernung oder einer Entfernung in einem Graph beruhen. Oder die Ähnlichkeit wird durch die Anzahl, bzw. Komplexität, an Transformationen berechnet, die notwendig sind, um das eine Konzept/Objekt in das andere zu überführen.

Für raumbezogene Daten rD, wie ich sie zu Beginn dieses Kapitels vorgestellt habe, findet sich vor allem die Ansicht der Ähnlichkeit als Distanzmaß in vielen Arbeiten in der Literatur wieder. Einer der bekanntesten Ansätze ist das *geometrische Modell*, welches nach [158] zuerst in der Psychologie verwendet wurde (nach [158] z.B. in [160]). In diesen Modellen werden Konzepte, bzw. Objekte, in einem mehrdimensionalen Raum modelliert und die räumliche Entfernung als Ähnlichkeitsmaß verwendet.

Das am meisten in der Literatur verwendete Entfernungsmaß ist die Minkowski-Distanz:

$$d_{ij} = \left[\sum_{k=1}^{n} |x_{ik} - x_{jk}|^r \right]^{1/r} \tag{2.3}$$

Wobei n die Anzahl der Dimensionen, x_{ik} der Wert in Dimension k für ein Objekt i und x_{jk} der Werte in Dimension k für ein Objekt j wiedergibt. Diese generische Formel kann für $r = 1$ Entfernungen als City-Block-Distanz ausrechnen und für $r = 2$ als euklidische Distanz. *Geometrische Modelle* beschränken sich häufig auf die Betrachtung von Ähnlichkeit zwischen Objekten. Eine der wenigen Ausnahmen ist das Modell der *Conceptual Spaces* nach Gärdenfors (2000/2004) (siehe [161] oder [162]) oder dessen Erweiterung nach Schwering (2005) [163], welches auch auf Konzepte angewendet werden kann. In der vorliegenden Arbeit beschränke ich mich allerdings auf die Betrachtung der Objektähnlichkeit in Bezug auf die Definition rD zu Anfang des Kapitels.

Im Bereich der semantischen Integration ist ein Gebiet, in dem schon sehr lange auf ein solches räumliches Entfernungsmaß zurückgegriffen wird, das der Clusteranalyse - siehe für einen Überblick die Arbeit von Eckes und Rossbach (1980) [164]. Ziel einer jeden Clusteranalyse ist es, eine Datensammlung so in einzelne Cluster zu unterteilen, dass die Objekte eines Cluster ähnlicher zueinander sind, als zu Objekten außerhalb ihres Clusters. Verfahren

der Clusteranalyse lassen sich nach [164] grob in partitionierende (iterativ-global und iterativ-partiell) und hierarchische (agglomerativ und divisiv) Verfahren einteilen. Erste Ansätze gehen von einer vorgegebenen Anzahl von Clustern aus und versuchen dann durch Neuzuweisung von Objekten aus den initialen Clustern zu anderen Clustern das Gesamtergebnis zu verbessern. Dieser Schritt wird solange wiederholt, bis sich eine globale oder lokale Zielfunktion nicht mehr weiter verbessern lässt.

Unter den iterative-partiellen Verfahren findet sich beispielsweise der *k-means* Algorithmus, siehe z.B. Smith et al. (2007) [164], der speziell im Kontext von raumbezogenen Daten Anwendung findet [165]. Der *k-means* Algorithmus findet sich in vielen Varianten in der Literatur wieder. Als Beispiel will ich kurz, die von Ashbrook und Starner (2003) in [3] vorgeschlagene Variante, welche ich auch im Kapitel 5 im Abschnitt 5.3.1 verwende, präsentieren.

Ashbrook und Starner´s Variante des *k-means* Algorithmus, vorgestellt in [3], berechnet aus aufgezeichneten GPS-Tracks interessante Orte, die die Ersteller der Tracks besucht haben. Dazu gruppiert er die in den GPS-Tracks enthaltenen GPS-Punkte nach folgendem Verfahren zu entsprechenden Clustern:

Um einen zufällig ausgewählten GPS-Punkt wird ein Kreis mit einem vorgegebenen Radius gelegt. Alle GPS-Punkte innerhalb des Radius gehören vorerst einmal zu diesem Cluster und werden markiert. Dann wird der Clustermittelpunkt - der Schwerpunkt in diesem Fall - ausgehend von diesen GPS-Punkten neu berechnet (siehe die ersten beiden Schritte in Abbildung 2.8). Weicht der neue Mittelpunkt vom alten ab, verschiebt sich das Cluster und der Mittelpunkt wird neu berechnet. Sobald sich das Cluster nicht mehr verschiebt (siehe letzter Schritt in Abbildung 2.8), werden alle Punkte, die in diesem Cluster liegen, im nächsten Schritt des Algorithmus nicht mehr weiter betrachtet und der Algorithmus beginnt mit einem neuen GPS-Punkt von vorne.

Ein ähnliches Vorgehen findet sich auch in der Arbeit von Kang et al. (2005) [166], wobei hier noch die Zeitspanne zwischen den zwei entferntesten GPS-Punkten in einem Cluster betrachtet wird, um Ausreißer, die vom GPS Fehler verursacht wurden, abzufangen.

Verfahren zur Berechnung der Ähnlichkeit von eindimensionalen Merkmalen, wie Linienzüge, beruhen weitestgehend auf drei Ansätzen, die in Abbildung 2.9 dargestellt sind. Eine Möglichkeit ist, die einzelnen Knoten individuell als Vergleichskriterium heranzuziehen [167]. D.h. sind zwei Knoten nah genug voneinander entfernt, werden sie zusammengeführt. Ein andere

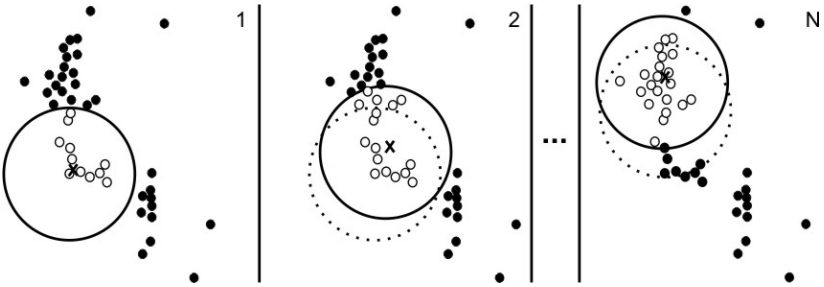


Abbildung 2.8.: Nach [3], Iterationsschritte im *k-means* Algorithmus nach Ashbrook und Starner (2003)

Möglichkeit besteht darin, dass man die Linienzüge in ihrer Gesamtheit betrachtet, d.h. alle zugehörigen Knotenpunkte. Aufgrund eines globalen Ähnlichkeitsmaßes wird entscheiden, ob die Linienzüge aggregiert werden oder nicht [7]. Als weitere Möglichkeit wird der Linienzug um einen Puffer erweitert und damit dieser als flächenhaftes Merkmal repräsentiert, d.h. die einzelnen Knoten spielen für das Ähnlichkeitsmaß keine Rolle mehr - nur bei der Entscheidung wo und wie der Puffer am Linienzug entlang läuft [168]. Zusätzlich spielen bei einigen Ansätzen andere Eigenschaften der Linienzüge eine Rolle, wie z.B. der Winkel an Überschneidungspunkten von zwei Linienzügen [169] oder die Anzahl der ausgehenden Kanten [167].

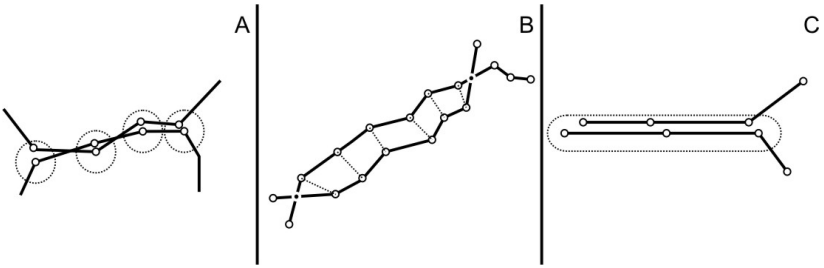


Abbildung 2.9.: Drei Möglichkeiten zur Bestimmung der Ähnlichkeit zweier Linienzüge

Möglichkeit A in Abbildung 2.9 wird beispielsweise von Volz (2006) [167] eingesetzt, um Straßendaten im GDF (Geographic Data Files) Format und aus ATKIS (Amtliches Topographisch-Kartographisches Informationssystem) zu-

sammenzuführen. Laut [167] werden ATKIS Daten von der Arbeitsgemeinschaft der Vermessungsverwaltungen der Länder der Bundesrepublik Deutschland (AdV) erhoben[25] und GDF Daten von zwei privaten Firmen, NAVTEQ[26] und TeleAtlas[27]. Volz versucht möglichst viele $1:1$ Beziehungen zwischen Knotenpunkten herzustellen, indem er ähnliche Knotenpaare im ATKIS und GDF Datensatz sucht. Ein Knotenpaar wird als ähnlich angesehen, wenn die euklidische Entfernung zwischen ihnen gering genug ist und sich der Winkel der ausgehenden Kanten nicht mehr als 5 Grad unterschiedet. Die Kanten zwischen den ähnlichen Knoten werden in einem anschließenden Schritt aufeinander abgebildet.

Möglichkeit B findet z.B. in der Arbeit von Morris et al. (2004) [7] Anwendung. Morris et al. aggregieren einzelne GPS-Tracks, die von Mountainbikefahrern während ihrer Touren aufgezeichnet wurden, zu einem einzigen Gesamtnetz. Der Ähnlichkeitswert, der dabei zur Anwendung kommt, basiert auf der Hausdorff-Entfernung $H(A, B)$ die nach den folgenden zwei Formeln berechnet wird:

$$h(A, B) = max_{a \in A} \{min_{b \in B} \{d(a, b)\}\} \tag{2.4}$$

$$H(A, B) = max \{h(A, B), h(B, A)\} \tag{2.5}$$

$h(A, B)$ berechnet die gerichtete maximale Entfernung von einer Punktmenge A zu einer Punktmenge B. $d(a, b)$ gibt die Entfernung zwischen zwei Punkten a und b. Diese kann durch jedes beliebige Entfernungsmaß repräsentiert werden, meistens wird aber die euklidische Distanz nach der Minkowski-Formel verwendet. $H(A, B)$ ist die allgemeine Formel der Hausdorff-Entfernung für beide Richtungen.

Walter und Fritsch (1999) [169] sowie Sayda (2005) [168] bestimmen ähnliche Linienzüge mit Hilfe eines Puffers, wie in Möglichkeit C dargestellt. Dabei wird der Linienzug um eine bestimmte Fläche vergrößert und dann überprüft, welcher andere Linienzug diese Fläche schneidet. Walter und Fritsch lassen den Puffer jeweils bis zu einem neuen Knoten eines Linienzuges wachsen, wenn dieser nicht mehr als 45% vom letzten Knoten abweicht. Diese Regel tritt aber nur in Kraft, falls ein Knoten mehr als zwei mit ihm verbundene Kanten besitzt. Dies schränkt den Suchraum des Algorithmus ein.

[25] http://www.atkis.de/
[26] http://www.navteq.com/
[27] http://www.teleatlas.com/

Sayda hat hingegen keine Einschränkungen in Bezug auf die Pufferoperation, verwendet aber das euklidische Distanzmaß in Form von Normalen zwischen den zwei Linienzügen, um zu entscheiden, welche Position der aggregierte Linienzug haben soll. Je nach Gewichtung der Linienzüge, wird im einfachsten Fall der Mittelwert der jeweils betroffenen Knoten der zwei Linienzüge als neue Position verwendet.

Das geometrische Modell, auf dem all die oben genannten Ansätze beruhen, bringt Eigenschaften mit, die in Bezug auf ein Ähnlichkeitsmaß umstritten sind:

1. *Minimalität*: Dieses Axiom sagt aus, dass wenn die Entfernung zwischen zwei Konzepten/Objekten null beträgt, diese identisch sind.

2. *Symmetrie*: Die Entfernung und damit die Ähnlichkeit zwischen zwei Konzepten/Objekten ist immer gleich, egal von welchem Konzept/Objekt ich diese betrachte.

3. *Dreiecksungleichung*: Nimmt man die Entfernung zwischen zwei Konzepten/Objekten ist diese immer kleiner oder gleich, als wenn ich die Entfernung zwischen diesen Konzepten/Objekten über ein drittes Konzept/Objekt messe.

Alle diese metrischen Eigenschaften wurden mit entsprechenden Experimenten in Frage gestellt, vor allem von Tversky in den späten 1970er Jahren [170] (siehe auch [158]). Um die Defizite des *geometrischen Modells* zu beheben, haben Tversky und Kollegen [171] ein auf der Mengentheorie beruhendes Ähnlichkeitsmaß vorgeschlagen. Das Maß $s(a, b)$ berechnet die Ähnlichkeit zwischen zwei Konzepten a und b als eine Funktion ihrer gemeinsamen ($A \cap B$) und nicht-gemeinsamen ($A - B, B - A$) Merkmale.

$$s(a, b) = F(A \cap B, A - B, B - A) \tag{2.6}$$

Tversky (1977) schlägt in seiner Arbeit [170] zwei konkrete Ausprägungen dieses allgemeinen Modells vor: das *contrast model* (Gleichung 2.7) und das *ratio model* (Gleichung 2.8), wobei letzteres die Normalisierung des *contrast model* darstellt:

$$S(a, b) = \theta * f(A \cap B) + \alpha * f(A - B) + \beta * f(B - A) \tag{2.7}$$

$$S(a, b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha * f(A - B) + \beta * f(B - A)} \tag{2.8}$$

Beide Modelle folgen dem Prinzip, dass Ähnlichkeit gerichtet und asymmetrisch ist. Ein Beispiel für die Verwendung des *ratio model* in Bezug auf raumbezogene Konzepte findet sich in der Arbeit von Rodriguez und Egenhofer (2003/2004) [172] [173]. Mit dem *Matching-Distance Similarity Measure* (MDSM) (Gleichung 2.9) berechnen Rodriguez und Egenhofer die Ähnlichkeit zwischen Konzepten, die wie das Beispiel aus 2.1 aufgebaut sind. Die den Konzepten zu Grunde liegende Ontologie haben die Autoren aus Daten von WordNet (siehe [174] und [175]) und SDST (Spatial Data Transfer Standard[28] eigenhändig aufgebaut, wobei aber nur raumbezogene Konzepte mit einbezogen wurden.

```
entity_class{
 name: {stadium, ball, arena}
 description:
        large often unroofed structure in
        which atheletic events are held
 is_a: {construction*}
 part_of: {}
 whole_of: {athletic_field*}
 parts:
        {{athletic_field, sports_field, playing_field
            },
        {dressing_room},{foundation},
        {midfield},{spectator_stands, stands},
        {ticket_office, box_office, ticket_booth}}
 function:
        {{play, compete},{play, practice},
        {recreate, play}}
 attributes:
        {{architectural_property},
        {covered/uncovered},{name},
        {lighted/unlighted},{owner_type},
        {sports_type},{usertype}}
}
```

Listing 2.1: Beispiel für ein in der MDSM Ontologie definiertes Konzept *Stadium*

[28] http://mcmcweb.er.usgs.gov/sdts/standard.html

Das Ähnlichkietsmaß ist nach [172] eine gewichtete Summe aus den Ähnlichkeitswerten für die drei Konzepteigenschaften *parts*, *functions* und *attributes*, siehe Gleichung 2.9.

$$S(c_1, c_2) = \omega_p * S_p(c_1, c_2) + \omega_f * S_f(c_1, c_2) + \omega_a * S_a(c_1, c_2) \quad (2.9)$$

Für die einzelnen Eigenschaften verwenden sie die an Tverskys *ratio model* angelehnte Formel nach Gleichung 2.10.

$$S_t(c_1, c_2) = \frac{|C_1 \cap C_2|}{|C_1 \cap C_2| + \alpha(c_1, c_2) * |C_1 \backslash C_2| + (1 - \alpha(c_1, c_2)) * |C_2 \backslash C_1|} \quad (2.10)$$

Die Funktion $\alpha(c_1, c_2)$ definiert die Entfernung zwischen den Konzepten c_1 und c_2 und dem direkten Oberkonzept, welches beide subsumiert. Die Berechnung wird nach Gleichung 2.11 durchgeführt:

$$\alpha(c_1, c_2) = \begin{cases} \frac{d(c_1, lub)}{d(c_1, c_2)} & d(c_1, lub) \leq d(c_2, lub) \\ 1 - \frac{d(c_1, lub)}{d(c_1, c_2)} & d(c_1, lub) > d(c_2, lub) \end{cases},$$
$$d(c_1, c_2) = d(c_1, lub) + d(c_2, lub) \quad (2.11)$$

Für das Maß $\alpha(c_1, c_2)$ muss eine hierarchisch aufgebaute Ontologie vorhanden sein, die in Form eines Graphen dargestellt werden kann. Die Knoten im Graph entsprechen den Konzepten. Die Verbindungen zwischen den Konzepten können dort entweder durch *part-of* oder *is-a* Beziehungen dargestellt werden.

Eine graphenbasierte Repräsentation wird auch in Ansätzen aus dem Bereich der *Netzwerkmodelle* favorisiert, die dann auf Grundlage der Weglänge von einem Konzept zum anderen deren Ähnlichkeit bestimmen. Einer der bekanntesten Vertreter dieser Art der Berechnung der Ähnlichkeit ist das Maß nach Resnik (1995) [8]. Dieses Modell verwendet ebenfalls WordNet (siehe [174] und [175]) als hierarchische Wissensbasis. Ich werde in Kapitel 4 Abschnitt 4.2.2 den Ansatz genauer erläutern und verweise an dieser Stelle nur auf die Arbeit von D´Ulizia et al. (2007) [176], die Resnik´s Ansatz auf geographische Konzepthierarchien angewendet haben.

Bernstein und Kollegen (2004/2005) stellen in [177] und [178] fünf verschiedene Prinzipien vor, wie man ein Ähnlichkeitsmaß für Ontologien berechnen kann. Dabei wählen sie eine ähnliche Einteilung wie Schwering (2008)

in [158]. Eine entsprechende Softwareunterstützung findet sich beispielsweise in [179] in Form des Programms *SimPack*.

- Ein erstes Maß, welches auf der Entfernung im Graphen beruht, der die Ontologie repräsentiert (*Ontology Distance*).

- Ein zweites Maß, dass auf dem Prinzip des Informationsgehaltes nach Resnik [8] beruht (*Information-theoretic entropy measures*.

- Ein drittes Maß, welches Konzepte in den Vektorraum abbildet (*Vector Space Approaches*) und dann die Entfernung zwischen den einzelnen Vektoren berechnet.

- Maß vier beruht auf der Levenstein Entfernung zwischen zwei Konzeptnamen (*Edit Distance*)

- Das letzte Maß stammt aus der Information Retrieval Literatur, welches zwei Dokumente mit Hilfe eines gewichteten Histogramm´s, das die Häufigkeit der Wörter des jeweiligen Dokumentes enthält, auf Ähnlichkeit vergleicht.

Ein verwandtes Forschungsfeld beschäftigt sich mit dem Problem des Schema-Matching, bei dem zwei Datenbankschams aufeinander abgebildet werden. Dazu müssen mit Hilfe entsprechender Ähnlichkeitsmaße ähnliche Konzepte in beiden Schemas gefunden werden, um sie ineinander zu überführen. Die Überblicksartikel von Rahm und Kollegen bieten hier einen guten Einstieg in die Thematik, siehe [180] und [181].

Ansätze die auf dem *Aligment Modell* oder dem *Transformational Modell* beruhen, sind meiner Meinung nach nicht sinnvoll auf raumbezogene Datensätze in der von mir präsentierten Form anwendbar. Desweswegen möchte ich hier nur auf die repräsentativen Arbeiten einmal von Markman und Gentner (1993) [182] und zum anderen von Goldstone (1994) [183], sowie Imai (1997) [184] und Hahn et al. (2003) [185] verweisen.

Neben Modellen und Ansätzen, die sich entweder auf den geometrischen oder auf den semantischen Aspekt der raumbezogenen Datensätze rD konzentrieren, d.h. entweder auf die $footprint$ oder auf die $name$ und cat Information, finden sich noch Ähnlichkeitsmaße, die beides kombinieren.

Zwei repräsentative Beispiele will ich an dieser Stelle kurz ausführen. Jones et al. (2001) schlagen in ihrer Arbeit [186] eine Linearkombination der

folgenden zwei Maße vor: ein Maß, welches die räumliche Nähe widerspiegelt (*TSD(q,c)*) und eines, welches die semantische Nähe zwischen zwei Objekten angibt (*TD(a,b)*). Der Anwendungskontext ist eine Anfrage q an ein Informationssystem, welches Fundorte von Gegenständen von archäologischen Ausgrabungen enthält, und Kandidaten c für die Antwort. $TSD(q,c)$, siehe Gleichung 2.12, ist allerdings selbst schon ein kombiniertes Maß, welches ein räumliches (die euklidische Distanz $ED(q,c)$) und ein semantisches Entfernungsmaß (siehe Gleichung 2.13) kombiniert.

$$TSD(q,c) = w_e * ED(q,c) + w_h * HD(q,c) \tag{2.12}$$

w_e und w_h sind beliebig zu setzende Gewichte für die beiden Maße $ED(q,c)$ und $HD(q,c)$. $HD(q,c)$ wird auf Grundlage einer hierarchischen Strukturierung von Ortsangaben berechnet, nach folgender Gleichung 2.13:

$$HD(q,c) = \sum_{x \in \{q.PartOf - c.PartOf\}} \frac{\alpha}{L_x} + \sum_{x \in \{q.PartOf - c.PartOf\}} \frac{\beta}{L_y} + \sum_{z \in \{q,c\}} \frac{\gamma}{L_z} \tag{2.13}$$

L_x, L_y und L_z geben die Hierarchieebene der einzelnen Orte in einer Hierarchiestruktur an. x ist die Menge der unterscheidbaren Orte, die sich direkt oberhalb des Anfrageterms q befinden, aber nicht eines möglichen Kandidaten c. y ist die Menge der unterscheidbaren Orte, die sich direkt oberhalb eines möglichen Kandidaten c befinden, aber nicht des Anfragebegriffs q. Die Orte z sind der Anfrage- und Kandidatenbegriff selbst. $q.PartOf$ und $c.PartOf$ beziehen sich auf die transitive Hülle der Orte oberhalb von q und c in der *part-of* Hierarchie. α, β und γ bieten die Möglichkeit z.B. die von Tversky [170] angesprochene Asymmetrie abzubilden.

$$TD(a,b) = \left(\frac{C_{a,x_1}}{L_{x_1}} + \frac{C_{x_1,x_2}}{L_{x_2}} + \frac{C_{x_2,x_n}}{L_{x_n}} + ... + \frac{C_{x_n,b}}{L_b} \right) \tag{2.14}$$

Nach [186] wird die thematische Distanz zwischen nicht-räumlichen Anfragebegriffen auf Grundlage des kürzesten Weges nach [187] berechnet. Dafür wird die Konzepthierarchie des AAT (Art & Architecture Thesaurus)[29]

[29] `http://www.getty.edu/research/tools/vocabulary/aat/`

verwendet. L_i ist die Ebene des Begriffes i, während $C_{j,k}$ Gewichte darstellen, die zur Verknüpfung zwischen den Begriffen i und j angewendet werden können.

Ebenfalls eine Linearkombination schlägt Hastings (2008) in seiner Arbeit [188] vor. Besonders interessant macht dessen Ansatz, dass er direkt mit Gazetteer Datensätzen der von mir vorgestellten Form rD arbeitet. Er präsentiert für jedes Element von rD ein Ähnlickeitsmaß, welche dann entsprechend gewichtet einfach addiert werden, siehe Gleichung 2.15:

$$s_{rD} = ws * Gs + wt * Gt + wn * Gn \tag{2.15}$$

Gs berechnet den Grad der geometrischen Überlagerung (Funktion $Area$) von zwei geometrischen Merkmalen F_1 und F_1:

$$Gs = Area(F_1 \cap F_2 / 0.5 * (Area(F_1) + Area(F_1)) \tag{2.16}$$

Gt berechnet die thematische Ähnlichkeit zweier Kategorien $T1$ und $T2$ in einer vorgegebenen Taxonomie T^*. Die Funktion $Upstep$ zählt dabei die Schritte in der Taxonomie, die nötig sind, bis von beiden Kategorien eine gemeinsame Oberkategorie erreicht wird.

$$Gt = 2^{-(Upstep(T1,T2)+Upstep(T2,T1))}, \quad gegeben\ eine\ Taxonomie\ T^* \tag{2.17}$$

Gn liefert einen Ähnlichkeitswert für die Namen der zwei Datensätze. Hierzu werden die beiden Zeichenketten zuerst in einzelne Wörter (*Tokens*) aufgeteilt und diese dann miteinander verglichen:

$$Gn = (Match(Tokens(S_1), S_2) + (Match(Tokens(S_2), S_1))) / Count(Token(S_1 + S_2)) \tag{2.18}$$

Häufiger werden allerdings die semantischen Information „nur“ als Filter angewendet, um die Ähnlichkeit zwischen semantisch identischen Datensätzen zu berechnen, z.B. in den Arbeiten von Gösseln und Sester (2003 /2004) in [189] und [190]. Ansonsten behandeln die vorgestellten Ansätze, die Maße auf Grundlage der geometrischen Information berechnen, raumbezogene Daten derselben semantischen Ausprägung, wie z.B. Straßendaten oder GPS-Tracks von Fahrradtouristen.

Ich werde in Kapitel 4 noch genauer darauf eingehen. An dieser Stelle möchte ich darauf hinweisen, dass alle hier vorgestellten Ansätze einen Anwendungsfall, wie er in Abbildung 2.10 abgebildet ist, bisher nicht betrachten. Diese Lücke will ich mit meiner Arbeit schließen.

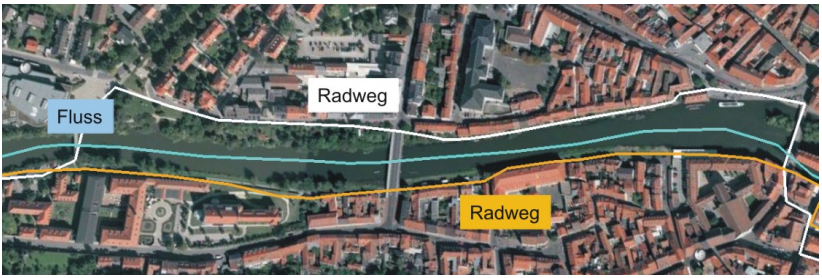


Abbildung 2.10.: Neuer Anwendungsfall für Aggregationsansätze 1-D-dimensionaler raumbezogener Daten, © Karte: Google Earth (`http://earth.google.com/intl/de/`).

Das Problem liegt darin, dass alle Aggregationsansätze, bzw. die dort verwendeten Ähnlichkeitsmaße, den Kontext in dem sich zwei zu aggregierende raumbezogene Merkmale befinden, nicht weiter betrachten. Das führt aber im gezeigten Anwendungsfall zu einer fälschlichen Zusammenführung der zwei abgebildeten GPS-Tracks. In dem gezeigten Beispiel würden die zwei raumbezogenen Datensätze aus der Kategorie *Radwege* zusammengeführt werden - da sie geometrisch sowie semantisch ähnlich genug, bzw. identisch sind. Das Resultat der Aggregation würde im Fluss liegen, was jedoch offensichtlich nicht korrekt sein kann. Der (semantisch) unähnliche Datensatz des Flusses wird von den Ansätzen, die ich in diesem Kapitel vorgestellt habe, bisher nicht betrachtet. Wie ich in Kapitel 4 zeigen werde, ist dieser entscheidend, um den Anwendungsfall korrekt zu behandeln.

3. Spielerische qualitätsgesicherte Erhebung raumbezogener Daten

Im folgenden Kapitel stelle ich meinen Lösungsvorschlag für die erste Fragestellung der vorliegenden Arbeit vor:

Wie können geowissenschaftliche Laien motiviert werden, freiwillig qualitätsgesicherte raumbezogene Daten zu erheben?

Wie in Kapitel 1 Abschnitt 1.1.1 angesprochen, schlage ich vor, ortsbezogene Spiele als Instrument einzusetzen, um das Motivationsproblem der geowissenschaftlichen Laien zu lösen. Dazu muss das Spielgeschehen so gestaltet sein, dass als *Nebenprodukt* (von Bell et al. 2009 als „by-product" [156] bezeichnet) zum Spielgeschehen Daten unterschiedlichen Typen erhoben werden. Gleichzeitig soll es möglich sein, die Qualitätssicherung der erhobenen Daten innerhalb des Spiels zu realisieren. Dabei darf der Spielspaß, bzw. der Spielfluss („game flow" [108]), nicht verloren gehen.

In diesem Kapitel werde ich dazu die folgenden Teilaspekte der obigen Fragestellung betrachten:

1. Zuerst werden in Abschnitt 3.1 durch eine Klassifikation existierender ortsbezogener Spiele die Bestandteile für ein Design Framework identifiziert. Das Framework kann für das Design beliebiger ortsbezogener Spiele zur qualitätsgesicherten Erhebung von raumbezogenen Daten eingesetzt werden.

2. Die gefundenen Elemente werden in den darauf folgenden Abschnitten (3.2 bis 3.5) näher erläutert und auch miteinander in Beziehung gesetzt.

3. Abschnitt 3.6 geht auf optionale Eigenschaften ein, die ein solches ortsbezogenes Spiel aufweisen kann, um sowohl die Quantität der sammelbaren Daten, als auch den Spielspaß, zu erhöhen.

4. Die vorgestellten Bestandteile des Design Frameworks werden in Abschnitt 3.7 in einen kompletten Design Prozess zusammengefasst.

5. Im letzten Teil des Kapitels (Abschnitt 3.8) wird als Anwendungsbeispiel des Design Frameworks das ortsbezogene Spiel CityExplorer eingeführt. Dabei wird auch auf die Besonderheiten bei der technischen Umsetzung des Spieles eingegangen.

3.1. Klassifikation ortsbezogener Spiele zur Erhebung raumbezogener Daten

In der Literatur finden sich zahlreiche Ansätze zur Klassifikation von ortsbezogenen Spielen. Diese werden hauptsächlich dazu genutzt, eine Definition für diese neue Art von Spielen einzuführen, siehe z.B. [78] und [77], [94], [84] oder [95]. Die Definition wird dann dazu verwendet, bereits existierende Spiele, bzw. Spielkonzepte, gegeneinander abzugrenzen. Ein Beispiel hierfür ist die Arbeit von Magerkurth et al. (2005) [76]. Laut Magerkurth et al. sind ortsbezogene Spiele eine Unterart von Pervasiven Spielen („Location-Aware Games"), werden aber nicht weiter unterteilt. Auch für den Bereich von Spielen zur Datenerhebung findet sich ein Übersichtsartikel [43] in der Literatur. Der zur Erhebung von raumbezogenen Daten wichtige Ortsbezug, spielt aber in der von Yen et al. (2009) [43] vorgestellten Klassifizierung keine Rolle, obwohl unter dem Begriff des „geospatial tagging" einige ortsbezogene Spiele genannt werden.

Einen etwas anderen Ansatz verfolgen Kiefer et al. (2006) in ihrer Arbeit [96]. Die Autoren entwickelten ihr Klassifikationsschema mit dem Ziel das Design von neuen Spielen, bzw. Spielkonzepten, aus bekannten zu ermöglichen. Der "Design durch Variation"-Ansatz (*design by variaton* [96]) erlaubt es, Spiele in den Dimensionen "Einbettung der Spielumgebung" - *game environmental embedding* -, "Spielkonzepte" - *game concepts* - und "Räumlich und Zeitliche Einteilung" - *Spatial and temporal dimension* - zu variieren. Durch die Veränderung eines Spieles in diesen drei Dimensionen entsteht ein ganz neues Spielerlebnis.

Zum Beispiel wird das Geogame GeoTicTacToe in drei oben genannten Dimensionen mindestens einmal verändert und gezeigt, wie sich dadurch das Spiel grundlegend ändert. Kiefer et al. [96] geben aber nur einen kurzen Hinweis darauf, wie man mit ihrer Methode vollkommen neue Spiele entwerfen kann. Des Weiteren wird der Aspekt des Einsatzes zur Datenerhebung nicht

weiter betrachtet und folglich finden sich auch keine Vorschläge zum Design solcher Spiele.

Um die Bestandteile für ein Design Framework für ortsbezogene Spiele zur Erhebung von raumbezogenen Daten zu identifizieren, habe ich 132 Spiele untersucht. Die Spiele selbst sind dabei sowohl Forschungsprototypen als auch kommerziell erhältliche Titel. Für eine detaillierte Beschreibung der einzelnen Spiele verweise ich interessierte Leser auf die Webseite `http://www.kinf.wiai.uni-bamberg.de/lbgdb/`. Die *location-based games database* enthält alle von mir gesammelten Spiele.

Neben den allgemeinen Kriterien - Name des Spiels, einer kurzen Beschreibung, einer Quellenangabe, (falls vorhanden) Veröffentlichungen und der Institution/Firma, die das Spiel entworfen hat - enthält das Klassifikationsschema die in Tabelle 3.1 aufgelisteten Kriterien. Die aus Kiefer et al. (2006) [96] entnommenen Ausprägungen *Chase*, *Puzzle*, *Strategy* und *Item Hunt* der Design Dimension „Spielkonzepte“ (*Game Conceptual dimension* [96]) habe ich auf einzelne Kriterien verteilt, um Mehrfachausprägungen für ein Kriterium zu vermeiden.

Kriterium	Beschreibung	Ausprägungen	Referenz
Players	Mögliche Spieleranzahl	1-255 (=MM)	-
Chase	Hauptspielelement (Fangen)	nein (0) oder ja (1)	[96]
Puzzle	Hauptspielelement (Rätsel lösen)	nein (0) oder ja (1)	[96]
Strategy	Hauptspielelement (Strategische Entscheidungen)	nein (0) oder ja (1)	[96]
Item Hunt	Hauptspielelement (Orte/Artefakte suchen)	nein (0) oder ja (1)	[96]
Technology	Programmiersprache	J2ME, .NET, DoJa, Android, Flash, Python, HTML, BREW, C++, Native	-

Hardware	Mobiler Endgerätetyp	PDA/Pocket PC, Mobile Phone, GPS Device, Custom	-
Localization	Lokalisierungstechnologie	GPS, WiFi, Bluetooth, Self-Positioning, none	[5]
Defined	Spielfeldstrukturierung	unstrukturiert (0), strukturiert (1) oder semi-strukturiert (2)	[5]
Monitored	Manuelle Spielleitung erforderlich	nein (0) oder ja (1)	-
Internet	Internetverbindung erforderlich	nein (0) oder ja (1)	-
Typ	Klasse des ortsbezogenen Spiels	location-based game (LBG), Mixed reality location-based games (MR), Augmented-reality location-based games (AR)	[96]
Spatiotemp	Raum- und zeitliche Einteilung des Spielgeschehens	sdtd, sdtc, sctc, sctd	[96]
Collect	Typus von sammelbaren Daten	LCN, GE, NGI oder GD	[5]

Status	Forschungs-prototyp oder kommerzielles Spiel	f oder k	-

Tabelle 3.1.: Kategorien und Ausprägungen des Klassifizierungsschema für ortsbezogene Spiele

Ausgangspunkt meiner Untersuchungen waren dabei die folgenden vier Fragestellungen:

1. Wie ist das Spielfeld strukturiert? (Im Schema unter dem Kriterium *Defined* erfasst)
2. Welche Typen von Daten werden gesammelt? (Im Schema unter dem Kriterium *collect* erfasst)
3. Werden zur Datenerhebung spezielle Spielmechanismen eingesetzt?
4. Wie wird die Qualität der gesammelten Daten gesichert?

Die Antworten auf diese Fragestellungen sollen die Definition eines Design Frameworks zur Erstellung von ortsbezogenen Spiele zur qualitätsgesicherten gemeinschaftlichen Erhebung von raumbezogenen Daten ermöglichen. Bevor wir uns aber das Framework in seiner Gesamtheit in Abschnitt 3.7 anschauen, werde ich im folgenden die einzelnen Bestandteile bzw. Fragestellungen näher beleuchten.

3.2. Strukturierung des Spielfeldes

Die Analyse der klassifizierten Spiele ergibt drei mögliche Strukturierungsarten für Spielfelder von ortsbezogenen Spielen. Diese sind dabei unabhängig davon, welche Merkmale das Spielfeld enthält - wie Orte, Wege oder Regionen - und welches Ausmaß es hat. Wie Abbildung 3.1 zeigt, kann das Spielfeld entweder strukturiert (*Defined* = 1), semi-strukturiert (*Defined* = 2) oder unstrukturiert (*Defined* = 0) sein.

Der Grad der Strukturiertheit wird dadurch festgelegt, wie frei die Spieler darin sind, im Spielfeld Spielaktionen durchzuführen. Diese Freiheit der Aktionsauswahl definiert sich dadurch, ob das Spielfeld vom Game Designer



Abbildung 3.1.: Strukturierungsmöglichkeiten des Spielfeldes ortsbezogener Spiele: strukturiert - semi-strukturiert - unstrukturiert

im Vorhinein festgelegte spielrelevante Merkmale - siehe Abbildung 3.2 - enthält, an denen Aktionen ausgeführt werden können, oder ob die Spieler diese im Laufe des Spiels selbst wählen können.

Sind Aktionen allein an vorher festgelegten Merkmalen möglich, spreche ich von einem strukturiertem Spielfeld. Zum Beispiel ist das Setzen eines Symbols - X oder O - in GeoTicTacToe [134] nur an neun vorher festgelegten Orten möglich.

Ein weiteres Beispiel ist das ortsbezogene Spiel *On the Streets* [88]. Zwei oder mehrere gegnerischer Teams (*Gangs*) versuchen im Spielverlauf so viel Territorium wie möglich zu erobern. Das Spielfeld besteht aus festgelegten rechteckigen Regionen (*fields*), welche eingenommen werden können. Alle Spielaktionen können nur innerhalb dieser Regionen ausgelöst werden. Also man muss nur in derselben Region sein, um einen gegnerischen Spieler anzugreifen, und nicht am selben Ort wie dieser.

Bei einem semi-strukturiertem Spielfeld sind ebenfalls Merkmale vorgegeben, auf denen man bestimmte Aktionen auslösen kann. Darüber hinaus sind die Spieler entweder vollständig frei wo sie weitere Aktionen ausführen oder zumindest werden die Merkmale nicht von der Spielsoftware vorgegeben sondern von der realen geographischen Umgebung.

Das Spiel *Treasure* [115] beispielsweise weist in seinem Spielfeld zufällig verteilte Münzen auf, die der Spieler einsammeln kann. Die Orte, wo die Münzen platziert werden, gibt die Spielsoftware vor. Zusätzlich können die Spieler aber noch Aktionen an beliebigen Orten im Spielfeld durchführen. So ist es z.B. möglich anderen Spielern Münzen zu stehlen und zwar immer dann, wenn ich mich auf eine bestimmte Entfernung diesem nähere. Des Weiteren müssen die Spieler die gesammelten Münzen per WLAN auf den

Spielserver hochladen. Die Punktzahl, die sie dafür bekommen, wird mit der jeweiligen Signalstärke des nächsten WLAN-Knotens modifiziert. So sind hochgeladene Münzen mehr wert, je besser der Empfang an dem Ort ist, wo sie hochgeladen werden. Obwohl die Orte der WLAN-Knoten dem Spieldesigner bekannt sind, ist die WLAN-Abdeckung, d.h. die Regionen in denen die Spielaktion „Münzen hochladen“ möglich ist, vor einem Spiel unbekannt und muss von den Spielern während des Spieles erkundet werden.

Ist das Spielfeld dagegen unstrukturiert, gibt es überhaupt keine vorbestimmten Merkmale und die Spieler sind entweder vollständig frei, wo sie Aktionen ausführen, oder die spielrelevanten Merkmale werden ausschließlich von der realen geographischen Umgebung bestimmt. Dies reduziert dann auch entscheidend den Aufwand, der bei den anderen Spielfelstrukturierungen zur Vorbereitung und Durchführung betrieben werden muss, siehe hierzu Oppermann et al. (2008) [132].

Feeding Yoshis [128] Spielfeldstrukturierung ist hierfür ein Beispiel. Die Aufgabe der Spieler ist es, virtuelle Kreaturen (*Yoshis*) mit Früchten zu füttern, um so einen möglichst hohen Punktestand zu erreichen. Früchte und Yoshis finden sich an jeweils geschlossenen und offenen WLAN-Knoten. Die WLAN-Knoten wurden aber nicht vorher vom Spieldesigner in der Spielumgebung verteilt. Die Spiel scannt einfach die reale Spielumgebung nach vorhandenen WLAN-Knoten ab und präsentiert diese dem Spieler dann entweder als Früchte (geschlossen) oder als Yoshi (offen).

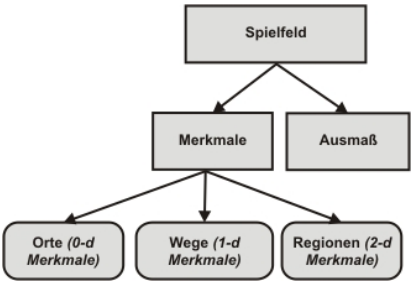


Abbildung 3.2.: Entscheidungspunkte bei der Spielfeldstrukturierung ortsbezogener Spiele

Der Spieldesigner muss nicht nur das wie und wo von Spielaktionen festlegen, sondern auch von welcher Art die geographischen Merkmale sind, bei denen diese Aktionen ausgelöst werden können. Abbildung 3.2 zeigt hier die verschiedenen Möglichkeiten für den Spieldesigner auf. Grundsätzlich lassen

sich drei Arten von geographischen Merkmalen unterscheiden: 0-dimensionale (Orte), 1-dimensionale (Wege) und 2-dimensionale (Regionen). Ob die Merkmale dabei vordefiniert sind (strukturiert) oder während des Spiels von den Spielern „gefunden“ werden (unstrukturiert), wird dann in einem zweiten Schritt entschieden. Zusätzlich kann noch das Ausmaß des gesamten Spielfeldes definiert werden. Die Spielfeldgrenze kann entweder fest vorgegeben sein oder sich aus der Verteilung der spielrelevanten Merkmalen erschließen.

3.3. Typen von sammelbaren Daten

Abhängig von der Strukturierung des Spielfeldes, aber unabhängig von der Entscheidung, welche geographischen Merkmale des Spielfeldes relevant sind, ergibt sich eine Kategorisierung der sammelbaren Daten durch ein ortsbezogenen Spiel, wie in Abbildung 3.3 gezeigt. Die Abhängigkeit kann allerdings auch umgekehrt werden, indem man erst den/die Typus/Typen an sammelbaren Daten festlegt, die man sammeln möchte und daraufhin die Strukturierung des Spielfeldes. Die Kategorisierung der sammelbaren Daten lehnt sich an meine Arbeit in Matyas (2007) [5] an.

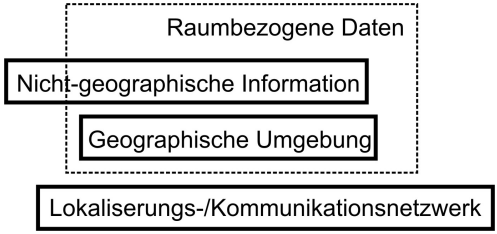


Abbildung 3.3.: Kategorisierung der mit einem ortsbezogenen Spiel sammelbaren Daten

Ausgehend von meiner Analyse der Datenbank ortsbezogener Spiele, kann nachfolgende Unterscheidung an sammelbaren Daten getroffen werden. Die Einteilung ergibt sich aus der Überlegung, wieviel Semantik die unbearbeiteten Daten mitbringen.

Auf der untersten Ebene ist es möglich, Daten über das verwendete **Lokalisierungs-/Kommunikationsnetzwerk** (*Collect* = LCN) zu erheben. Wenn wir von der Definition von Kiefer at al. (2006) [96] ausgehen, welche besagt, dass in allen Arten von orstbezogenen Spielen eine Form von Lokalisierungstechnologie benutzt wird, ist es möglich, mit allen ortsbezogenen Spielen

diesen Typus von Daten zu sammeln. Mit Ausnahme von wenigen ortsbezogenen Spielen für Einzelspieler, wie z.B. *Journey I und II* aus der Arbeit von Jakl (2006) [139], wird immer auch eine Kommunikationstechnologie verwendet, um beispielsweise den globalen Spielzustand zu synchronisieren oder zumindest am Ende eines Spiels den persönlichen Highscore auf einen Server hochzuladen.

Die einzige Voraussetzung hierfür ist eine reine technische. Die Spielesoftware muss es möglich machen, die empfangenen Daten aus den verwendeten Netzwerken (für Lokalisierung und Kommunikation) der Spieler in diskreten Zeitabschnitten aufzuzeichnen und persistent zu speichern, z.B. über Log-Dateien. Beispielsweise nutzt das Spiel *Can You See Me Now! (CYSMN)* von Benford et al. (2006) [102] die Lokalisierungstechnologie GPS. Die gesammelten Daten zu den Empfangsbedingungen im GPS Netzwerk werden hier nach einem Spiel visualisiert, um die Güte des Empfangs im Spielfeldes für spätere *CYSMN* Spiele zu bestimmen. Nativ enthalten die Daten keine Semantik, sondern diese werden erst bei der Auswertung durch entsprechende Anwendungen hinzugefügt bzw. durch einen externen Standard festgelegt.

Auf der nächsthöheren Ebene erheben die Spieler Daten über die **geographische Umgebung** (*Collect* = GE), in der das Spiel stattfindet. Da die Bewegung der Spieler ein Hauptelement in einem ortsbezogenen Spiel ist [96], erlauben z.B. die aufgezeichneten Bewegungsspuren Wege in der geographischen Umgebung zu identifizieren. Theoretisch ist damit die Erhebung von Daten zu jeder der Merkmalsklassen aus Abbildung 3.2 möglich. D.h. die Spieler zeichnen während des Spiels entweder Orts-, Wege- oder Regionendaten oder eine beliebige Kombination davon auf. Dabei wird die Merkmalsklasse zu den Lokalisierungsdaten, wie GPS Koordinaten, von den Spielern oder der Anwendung schon bei der Aufnahme explizit angegeben bzw. vorgegeben. Nehmen wir das Spiel *GeoTicTacToe* aus der Arbeit von Schlieder et al. (2006) als Beispiel [134]. Die Spielanwendung zeichnet die Bewegung der Spieler als Folge von GPS Koordinaten auf. Da die Spielregeln vorgeben, dass die Spieler nur zu Fuß unterwegs sein dürfen, können die so aufgezeichneten Koordinatenfolgen später automatisch als Wege interpretiert werden.

Auf der obersten Ebene werden **nicht-geographische Informationen** (*Collect* = NGI) erhoben, d.h. Daten, die entweder nur indirekt einen Bezug zur geographischen Umgebung des Spielfeldes haben oder die direkt Daten der geographischen Umgebung näher beschreiben.

Beispiele für diesen Typus von Daten sind eine Klassifizierung der Merkmale aus Abbildung 3.2 - z.B. ob ein Ort ein Hotel, Restaurant oder Museum ist - oder weiterführende Informationen wie Hotelpreise, Speisekarten von Restaurants oder Öffnungszeiten von Museen.

Bringt man nun die Daten zur geographischen Umgebung und die nichtgeographischen Information zusammen, bzw. lässt sie durch ein entsprechendes Spielelement gemeinsam erheben, bekommt man von den Spielern freiwillig die Daten, die aktuelle ortsbezogene Dienste meist teuer anbieten: Lokales Wissen oder informatisch gesehen **raumbezogene Daten** (*Collect* = GD). Sei es z.B. der Standort des neuen indischen Restaurants mitsamt der aktuellen Speisekarte, Öffnungszeiten und Telefonnummer zur Reservierung eines Tisches oder die neuen Zimmerpreise eines mit GPS Koordinaten verorteten Hotels.

Im nachfolgenden Abschnitt präsentiere ich verschiedene Entwurfsmuster zum Design von Spielelementen, die Spieler eines ortsbezogenen Spieles dazu bringen, Daten der oben genannten Typen auf unterschiedlich strukturierten Spielfeldern zu erheben. Ich werde vor allem darauf eingehen, welches Entwurfsmuster am Besten mit welcher Strukturierung des Spielfeldes gewählt wird und welchen Typus von Daten man am Besten so erheben kann.

3.4. Entwurfsmuster zur Motivation spielgetriebener Datenerhebung

Meine Analyse der Spielesammlung ergab insgesamt fünf Entwurfsmuster, die es einem Designer ermöglichen, Spieler zur Datenerhebung zu motivieren. Zur Beschreibung meiner Entwurfsmuster lehne ich mich grob an das Schema von Björk and Holopainen (2004) [21] an, welches Sie verwendeten, um Entwurfsmuster für Videospiele darzustellen. Wo immer möglich verweise ich auf das entsprechende Entwurfsmuster n nicht-ortsbezogenen Spielen auf (*Relations* in [21]). Wie am Ende des letzten Abschnitts erwähnt, betrachte ich zusätzlich ich noch zwei Punkte, die in [21] keine Erwähnung finden: Mit welcher Spielfeldstrukturierung das Entwurfsmuster am Besten umgesetzt werden kann (*Spielfeldwahl*) und welcher Typus von Daten damit am Sinnvollsten gesammelt werden kann (*Datensammlung*).

Die Schwierigkeit bei der Umsetzung der im nachfolgenden vorgestellten Entwurfsmuster ist die Validierung der Aktionen der Spieler. Da die Idee der hier vorgestellten Entwurfsmuster darin besteht, die in Abschnitt 3.3 besprochenen Typen von Daten durch die Spieler erheben zu lassen, sollten

möglichst wenige Daten vom Designer im Vorhinein erhoben werden. Zum Beispiel, ob die nicht-geographischen Informationen zu einem geographischen Merkmal im Spielfeld korrekt aufgenommen wurden oder nicht, kann dann je nach Spiel nicht mehr automatisch durch die Spielesoftware überprüft werden. Um dieses Problem zu lösen, stelle ich in Abschnitt 3.5 verschiedene Strategien vor.

3.4.1. *Log-Datei(en)* Konzept

Neben dem eigentlichen Spiel werden Daten zum Kommunikations- und/oder Lokalisierungsnetzwerk in einer oder mehrerer Log-Dateien aufgezeichnet.

Anmerkungen: Obwohl kein Entwurfsmuster im klassischen Sinne, ermöglicht der Einsatz des Konzepts der Log-Datei doch die einfachste Möglichkeit Daten im Spiel zu sammeln. Log-Dateien sind dabei auch unabhängig von jeglicher Spielelogik und können somit von jedem beliebigen ortsbezogenen Spiel eingesetzt werden.

Beispiel: In *Can You See Me Now?* [102] wurden während eines Spieles in Rotterdam mit Hilfe von Log-Dateien die Empfangsbedingungen der verwendeten Lokalisierungs- und Kommunikationsnetzwerke (GPS und WLAN) festgehalten. Die daraus erstellte thematische Karte kann dazu verwendet werden, entweder das Spielfeld für zukünftige Spiele besser auszuwählen oder zumindest auf Bereiche des Spielfeldes mit schlechten Empfangsbedingungen vorbereitet zu sein.

Ein Hinweis auf die Arbeit von Benford et al. (2006) [102] muss an dieser Stelle folgen. Die Autoren stellen fünf allgemeine Strategien vor, wie mit schlechten Empfangsbedingungen (*seams*) umgegangen werden kann. Da die Empfangsbedingungen entscheidend für die Güte der Qualität, der durch das Konzept der Log-Dateien erhobenen Daten ist, muss dieses Problem vom Designer entsprechend angegangen werden. Für eine genaue Beschreibung der Strategien sei dieser auf Kapitel 2 Abschnitt 2.3 verwiesen.

Sammelbare Typen von Daten: Da die Idee der Log-Datei darin besteht, Daten nebenläufig zum Spielgeschehen zu sammeln, d.h. ohne Interaktion durch den Spieler, können nur Daten über das im Spiel eingesetzte Kommunikations-, bzw. Lokalisierungsnetzwerk gesammelt werden.

Favorisierte Spielfeldstrukturierung: Log-Dateien sind vollkommen unabhängig von der Strukturierung des Spielfeldes.

Vergleichbares Entwurfsmuster nach [21]: Keines.

3.4.2. *Rätsel/Aufgaben*-Entwurfsmuster

Die Spieler müssen an spielrelevanten Orten ein oder mehrere Rätsel (Quiz), bzw. Aufgaben (Quests) lösen, um eine bestimmte Spielaktion auszuführen (siehe auch [5]).

Anmerkungen: Die Rätsel/Aufgaben müssen nicht zwingend einen direkten oder indirekten Ortsbezug aufweisen, sind aber mit einem solchen für den Spieler nachvollziehbarer und fügen sich so besser in das Spielgeschehen ein. Auch im Hinblick auf die Erhebung von raumbezogenen Daten sind nicht-geographische Informationen mit einem Ortsbezug (direkt/indirekt) von größerem Interesse als solche, die keinen Ortsbezug aufweisen.

Beispiel: Obwohl meines Wissens kein ortsbezogenes Spiel bisher das Entwurfsmuster *Rätsel/Aufgaben* als Hauptspielelement umgesetzt hat, haben die Spieler in MobiMissions [2] einzelne Missionen in Form von Rätseln/Aufgaben erstellt. Die Missionen bestehen aus einer Reihe von Photos und Text, meistens in Form einer Frage oder einer Aufgabe. In einer von Grant et al. (2007) in [2] vorgestellten Mission mussten Spieler ein Objekt auf einem Photo identifizieren. Obwohl diese Art von Missionen nur eine simple Umsetzung dieses Entwurfsmusters darstellt, zum großen Teil ohne jegliche Form des Ortsbezugs - siehe Kapitel 2 Abschnitt 2.3.3 -, zeigt die Existenz solcher Missionen doch die Bereitschaft der Spieler, Spiele mit solchen Elementen spielen zu wollen.

Sammelbare Typen von Daten: Das Entwurfsmuster kann besonders gut eingesetzt werden, um nicht-geographische Informationen durch die Spieler erheben zu lassen. Beispiele sind die Preise auf einer Speisekarte für ein Restaurant oder die Öffnungszeiten eines Bekleidungsgeschäftes. Natürlich lassen sich prinzipiell beliebige Wissensfragen durch das Entwurfsmuster in ein Spiel einbauen. Aber Fragestellungen, die von den Spielern vor Ort gelöst werden können, sind im Hinblick auf die Immersion der Spieler (unter dem Begriff *spatial/temporal grounding* nach Kiefer et al. 2006 [6] bekannt) zu bevorzugen.

Favorisierte Spielfeldstrukturierung: Prinzipiell wird jede Spielfeldstrukturierung durch dieses Entwurfsmuster unterstützt. Wenn man aber die Erhebung von nicht-geographischer Informationen als Designziel hat, ist die Verwendung eines strukturierten oder gegebenenfalls semi-strukturierten Spielfeldes empfehlenswert.

Vergleichbares Entwurfsmuster nach [21]: Gain Information (Game Design Patterns for Goals)

3.4.3. *Transporter*-Entwurfsmuster

Hier besteht die Aufgabe des Spielers darin, ein oder mehrere spielrelevante Artefakte von einem geographischen Ausgangsort A zu einem ebenfalls geographischen Bestimmungsort B zu bringen, oder anders ausgedrückt zu „transportieren" (siehe auch [5]).

Anmerkungen: Das Artefakt kann entweder virtuell oder physisch umgesetzt werden, wie z.B. virtuelle „Anhalter" (Hitcher) im *Hitchers* Framework nach Drozd et al. (2006) [103] oder physische Pokerkarten aus *CityPoker* (siehe [99] oder [4].

Beispiel: Ein Beispiel für die Umsetzung des Entwurfsmusters ist sicherlich das Spiel *Gophers* von Casey et al. (2007) [129], welches wiederum das Hitchers Framework aus [103] verwendet. In Gophers sind die virtuellen Artefakte Erdhörnchen, die den Spieler auf unterschiedliche Missionen schicken; die meisten eben vom Typ des *Transport*-Entwurfsmusters.

Sammelbare Typen von Daten: Durch das *Transporter*-Entwurfsmuster ist es vor allem möglich, Daten über die geographische Umwelt zu erheben, wie z.B. mögliche Wege zwischen dem Ausgangs- und dem Bestimmungsort.

Favorisierte Spielfeldstrukturierung: Hier bietet sich ein strukturiertes oder semi-strukturiertes Spielfeld an. Dadurch haben die Spieler zumindest einen festen Ort im Spielfeld, den sie immer anlaufen können. Bei einem unstrukturiertem Spielfeld könnte ansonsten das unter Umständen zeitaufwendige Suchen des Ausgangs- bzw. des Bestimmungsortes alle anderen Spielelemente in den Hintergrund drängen.

Vergleichbares Entwurfsmuster nach [21]: Delivery (Game Design Patterns for Goals)

3.4.4. *Erkundung*-Entwurfsmuster

Ein Spiel, welches dieses Entwurfsmuster umsetzt, schickt seine Spieler aus, das Spielfeld nach spielrelevanten geographischen Orten zu „erkunden“.

Anmerkungen: Die Standorte der aufzusuchenden Orte sind nicht notwendigerweise durch die Spielesoftware vorgegeben bzw. auf andere Art und Weise - wie zum Beispiel durch GPS Koordinaten beim Geocaching [1] - im Vorhinein vom Spieldesigner definiert.

Das Entwurfsmuster ähnelt dem des *Transporter*-Entwurfsmusters, kommt aber ohne definierten Ausgangsort aus und die Spieler müssen kein Artefakt überbringen (virtuelle wie physische). Aufgabe der Spieler könnte sein, die GPS Koordinaten von allen Restaurants in einem Spielfeld aufzunehmen. Somit sind nur die Bestimmungsorte abstrakt vorgegeben, aber der Spieler ist in seiner Entscheidung frei, von welchem Ort im Spielfeld er seine Suche beginnt.

Beispiel: Im Spiel*EyeSpy* gehen Bell et al. (2009) [156] noch einen Schritt weiter. Hier können die Spieler von jedem beliebigen Ort georeferenzierte Fotos machen, die dann von den Mitspielern anhand der Fotos gefunden werden müssen. Wenn ein Ort mit Hilfe des Fotos - die Lokalisierungsdaten stehen den Spielern nicht zur Verfügung - gefunden wird, bekommt der Ersteller des Fotos und der Finder Punkte. D.h. Ausgangs- und Bestimmungsort sind den Spielern zu Beginn des Spieles unbekannt, bzw. werden vom Spieldesigner nicht vorgegeben.

Sammelbare Typen von Daten: Mit diesem Entwurfsmuster lassen sich alle Typen an Daten sammeln, die in Abschnitt 3.3 beschrieben sind. Insbesondere die Erhebung von raumbezogenen Daten ist möglich.

[1] `http://www.geocaching.com/`

Favorisierte Spielfeldstrukturierung: Durch die Unbestimmtheit der Bestimmungsorte bietet sich hier ein semi- oder unstrukturiertes Spielfeld an. Will man den Output an raumbezogenen Daten maximieren, ist die Verwendung eines unstrukturierten Spielfeldes zu empfehlen, da somit alle raumbezogenen Daten von den Spielern selbst erhoben werden.

Vergleichbares Entwurfsmuster nach [21]: Exploration (Game Design Patterns for Goals)

3.4.5. *Produzent*-Entwurfsmuster

Spielern ist es erlaubt, selbst Spielelemente zu erstellen, die dann von allen anderen Spielern verwendet werden können.

Anmerkungen: Natürlich legt das entsprechende Spiel den Rahmen fest, indem diese Spielelemente gestaltet werden können. Meines Wissens wurden bisher im Kontext von ortsbezogenen Spielen nur Missionen als von den Spielern erstellbares Spielelement umgesetzt.

Beispiel: Beispiele von Spielen, die dieses Entwurfsmuster verwenden, sind *MobiMission* [2], *Gophers* [129] - welche beide auf dem *Hitchers* Framework [103] beruhen - sowie *Clandestine* [157]. In allen drei Spielen erstellen die Spieler Missionen für die anderen Spieler im Spiel. In den ersten beiden Spielen handelt es sich um Missionen, bei denen die Spieler anhand von Fotos Gegenstände erraten müssen. Häufig werden auch Missionen nach dem *Transporter*-Entwurfsmuster entworfen - wie z.B. in *Gophers*.

In *Clandestine* wird das Ganze dann noch in einen thematischen Rahmen eingepasst. Die Spieler sind in Teams organisiert, die wiederum Vampirclans repräsentieren. Die erstellten Missionen bekommen somit einen thematischen Bezug zum Spiel und können nicht beliebigen Inhalt haben, wie bei den anderen beiden erwähnten Spielen. Außerdem müssen die Missionen in *Clandestine*, im Gegensatz zu den zwei anderen, zwingend einen Ortsbezug aufweisen.

Sammelbare Typen von Daten: Dadurch, dass je nach Spielregeln die Erstellung beliebiger Spielelemente möglich ist, ist die Erhebung aller Typen von Daten möglich. Zusätzlich lassen sich alle zuvor vorgestellten Entwurfsmuster - außer des *Log-Datei*-Entwurfsmusters - von den Spielern verwirklichen.

Favorisierte Spielfeldstrukturierung: Um den Spielern möglichst große Freiheiten bei der Gestaltung ihrer Spielelemente zu geben, sollte das Spielfeld möglichst unstrukturiert sein. Sollte es je nach Spielelement oder Entwurfsmuster doch nötig sein, einen oder mehrere spielrelevante Orte festzulegen, sollten die Spieler die Möglichkeit haben, diese zu definieren.

Vergleichbares Entwurfsmuster nach [21]: Construction (Actions and Events Patterns)

3.5. Entwurfsmuster zur Qualitätssicherung

Die im letzten Abschnitt vorgeschlagenen Entwurfsmuster zur Erhebung von Daten mit Hilfe von ortsbezogenen Spielen garantieren natürlich noch nicht, dass die gesammelten Daten eine entsprechende Qualität aufweisen. Aus diesem Grund muss beim Design des entsprechenden Spieles auch immer die Qualitätssicherung im Fokus stehen. Dabei sollte das Spiel selbst Elemente enthalten, um dies zu gewährleisten. Oder der gesamte Spielablauf muss so gestaltet sein, dass die erhobenen Daten von den Spielern im Spiel selbst überprüft werden. Da die Qualitätssicherung, wie auch die Erhebung, ohne Expertenwissen möglich sein soll, muss der Sicherungsprozess entsprechend einfach gestaltet sein.

Meine Analyse der in Abschnitt 3.1 vorgestellten Spielesammlung, ergab die folgenden drei Entwurfsmuster:

3.5.1. *Peer Reviewing*-Entwurfsmuster

Die Daten jeweils eines Spielers werden von anderen Spielern überprüft. Dies kann entweder im selben Spiel geschehen, oder in einem nachfolgenden.

Anmerkungen: Dieses Entwurfsmuster stellt die direkteste Umsetzung einer Qualitätssicherung durch Experten dar. Anstelle der Expertenmeinung werden eben die gesammelten Bewertungen der mitspielenden Laien verwendet. Je nach Art und Weise der Einbindung in das Spielgeschehen, liegt die Schwierigkeit dieser Strategie darin, einen vorsätzlichen Missbrauch durch die Spieler zu verhindern. Bringt z.B. die falsche Bewertung von Daten im Spiel (eigentlich gute Daten als schlecht oder andersherum zu bewerten) einen Vorteil, könnten zumindest einige Spieler versucht sein, dies auszunutzen. Je nach Spiel muss versucht werden, ein solches Verhalten zu verhindern, indem

man es entweder bestraft oder den Spieler sogar ganz vom Spiel ausschließt. Für web-basierte Spiele haben Ahn und Dabbish (2008) in [20] verschiedene Strategien vorgestellt, wie dies erreicht werden kann. Im restlichen Kapitel werde ich an entsprechender Stelle auf diesen Aspekt im Kontext von ortsbezogenen Spielen eingehen.

Die Bewertung der Daten durch die Spieler kann mit unterschiedlichen Metriken bzw. Verfahren durchgeführt werden. Dabei ist alles, angefangen von binären Bewertungen (z.B. „korrekt“ und „nicht korrekt“) bis hin zu 5-Punkte oder n-Punkte-Skalen (vergleiche die „Sterne-Wertung“ auf z.B. Amazon [2]), denkbar. Wird ein Datensatz besser bewertet als der Mittelwert der verwendeten Bewertungsskala, wird er als von ausreichender Qualität angesehen. Neben dem einfachen Mittelwert sind auch hier zahlreiche andere Maße möglich. Um nur ein Beispiel zu nennen, könnte man in die Wertung auch Trust-Werte [191] einfließen lassen, die sich die Spieler entweder in der Community für das jeweilige Spiel erworben haben oder die aus einer dritten Quelle stammen. Die Wertung von Spielern mit höherem Trust können dann auch mit einer höheren Gewichtung in die Gesamtbewertung einfließen.

Ein weitere Punkt, der noch zu beachten ist, ist: Was passiert mit den als „schlecht“, d.h. von nicht ausreichender Qualität, bewerteten Daten? Eine Möglichkeit, welche einen Einfluss auf das Peer Reviewing Verfahren selber hat, ist, den Spielern schon während des Spiels die Möglichkeit zu geben, ihre eingestellten Datensätze überarbeiten zu können. Die Bewertung läuft dann in zwei bis n Runden ab:

1. Ein (erhobener) Datensatz wird von den Mitspielern bewertet.
2. Ist eine kritische Masse an Bewertungen vorhanden, wird der Datensatz als „gut“ oder „schlecht“ klassifiziert.
3. Wird er als „gut“ bewertet, endet das Verfahren.
4. Wird er als „schlechter“ Datensatz bewertet, kann der Spieler, der ihn initial eingestellt hat, überarbeitet werden.
5. Mit dem überarbeiteten Datensatz beginnt das Verfahren wieder bei Punkt 2.

Beispiel: Im Spiel *Gopher* [129] findet sich ein recht komplexes Beispiel des vorgestellten Entwurfsmusters. Nach dem Vorbild einer Jury bei Gerichtsverhandlungen werden bei *Gopher* 5% der teilnehmenden Spieler eingeladen,

[2] `www.amazon.de`

eine von einem Spieler als erfolgreich beendete Mission zu begutachten. Die Juroren haben dafür 24 Stunden Zeit und bekommen für ihre Teilnahme eine vorher unbekannte Anzahl an Punkten im Spiel.

Die Punktzahl richtet sich danach, wie nah die Bewertung der Juroren am Median der Gesamtbewertung liegt. Dadurch soll die Ehrlichkeit der Juroren gewährleistet werden. Zusätzlich erhält der Spieler, der die Mission durchgeführt hat, Punkte und zwar je nach dem, wie die Juroren die Mission bewertet haben. Zuallerletzt bekommen noch all diejenigen Punkte, die dem Spieler bei der Erfüllung der Mission geholfen haben.

Wie die Autoren selber eingestehen, ist dies ein eher komplexes System, das dem Versuch geschuldet ist, eine Art („in-game ecomomy“) Wirtschaftssystem im Spiel zu integrieren. Dies hatte zur Folge, dass die Spieler bei der Bewertung des Spieles gerade diese Komplexität bemängelt haben - wobei bei der Befragung der Spieler leider nicht nach einzelnen Spielelementen differenziert wurde - und dem Spiel als Ganzes eine zu steile Lernkurve bescheinigten (siehe [129]).

Ich werde in Abschnitt 3.8 eine weitere Möglichkeit vorstellen, wie eine Umsetzung des *Peer Reviewing*-Entwurfsmusters aussehen kann, welches weniger Komplex aufgebaut ist und den Aspekt der „böswilligen“ Spieler geeignet behandelt.

3.5.2. *Output-Input*-Entwurfsmuster

Die erhobenen Datensätze einer schon gespielten Spielrunde bzw. eines anderen Spieles werden in einer nachfolgenden Spielrunde bzw. einem anderen Spiel als Ausgangsdaten verwendet.

Anmerkungen: Dieses Entwurfsmuster entkoppelt die Erhebung und die Bewertung der Daten durch ein Spiel voneinander. Wie schon in der Beschreibung angedeutet, kann die Entkopplung entweder dadurch geschehen, dass der Output an erhobenen Datensätzen eines Spiels in einer nachfolgenden Spielrunde desselben Spiels verwendet wird, oder man verwendet den Output gleich in einem komplett anderen Spiel, wie in Abbildung 3.4 dargestellt.

Wird die erste Strategie verfolgt und sollen die Spielregeln nicht von einer Runde zur nächsten geändert werden, müssen zumindest in der aller ersten Runde vordefinierte oder zufällig erstellte Daten verwendet werden. Danach kann immer wieder auf von den Spielern erhobene Datensätze zurückgegriffen werden. Ein beispielhafter Regelsatz für ein solches Spiel könnte folgendermaßen aussehen:

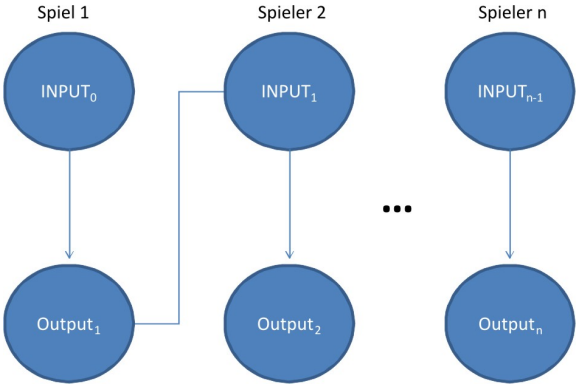


Abbildung 3.4.: Schematische Darstellung des *Output-Input*-Entwurfsmusters

Das Ziel des Spiels soll es sein, für n vorgegebene Orte im Spielfeld die n nächsten Gebäude zu finden, die zu einer bestimmten Kategorie (z.B. Restaurants, Hotels usw.) gehören. Das Spiel besteht aus mehreren Spielrunden, die jeweils zeitlich begrenzt sind:

1. In der ersten Spielrunde werden die initialen Orte zufällig im Spielfeld verteilt. Punkte werden in absteigender Reihenfolge nach der Entfernung der von den Spielern aufgenommenen Gebäude zum initialen Ort verteilt. D.h. der Spieler mit dem nächsten Gebäude bekommt auch am meisten Punkte.

2. Runde zwei verwendet dann als Ausgangsorte, die in Runde eins erhobenen Orte. Dadurch, dass der Ausgangsort auch gleichzeitig das nächste Gebäude ist, wird dieses mit sehr großer Wahrscheinlichkeit nochmals von demselben oder einem anderen Spieler aufgenommen und somit indirekt überprüft. Um das Ganze für die Spieler abwechslungsreicher zu gestalten, können statt den gleichen Kategorien für die aufzufindenden Gebäude auch ähnliche (statt Restaurant z.B. Lokal) verwendet werden.

3. Runde drei und nachfolgende werden wie Runde zwei gespielt.

Beispiel: Im Spiel *EyeSpy* [156] werden von den Spielern reale Orte per Photo oder Text aufgenommen, zusammen mit einer Georeferenzierung per

WLAN Triangulierung. Diese dienen dann als Missionen für andere Spieler. Die Mission besteht darin die aufgenommenen Orte zu finden, bzw. den Standort, von wo das Bild aufgenommen wurde. Dadurch wird der ursprünglich aufgenommene Datensatz durch im günstigsten Fall n weitere Spieler bestätigt. Sowohl der Ersteller der Mission als auch der- oder diejenigen, die sie spielen, bekommen bei Erfolg Punkte.

Wird ein Ort allerdings gar nicht aufgesucht - d.h. sein Datensatz nicht bestätigt -, kann dadurch aber noch keine Aussage über seine Unkorrektheit gemacht werden. Er kann auch einfach in einem von den übrigen Spielern schwer zugänglichen Gebiet liegen und somit unattraktiv sein. Somit kann *EyeSpy* nur in korrekte und unentschiedene Datensätze unterteilen.

3.5.3. *Wisdom of the Crowd*-Entwurfsmuster

Die Spieler werden durch die Spielregeln angeleitet, Entscheidungen im Sinne einer anonymen Mehrheit zu treffen.

Anmerkung: Die Entscheidungen der Spieler können sowohl die Datenerhebung betreffen als auch die abgegebenen Bewertungen.

Beispiel: Wie schon im Beispiel für das *Peer Reviewing*-Entwurfsmuster angedeutet, wird das *Widom of the Crowd*-Entwurfsmuster bisher nur im Zusammenhang mit der Bewertung von schon erhobenen Datensätzen, bzw. gespielten Missionen, verwendet, siehe z.B. [129]. Die einfachste Art und Weise, dass „Wissen der Vielen“ [23] einzusetzen, ist die, allen Spielern die Möglichkeit zu geben, alle Datensätze mit Hilfe einem binären Entscheidungsmaß zu beurteilen. Nach dem Spiel werden dann diejenigen Datensätze als korrekt angesehen, die mehr positive als negative Bewertungen aufweisen.

Ich werde in Kapitel 5 Abschnitt 5.2 aufzeigen, wie man das Entwurfsmuster auch schon während der Datenerhebung einsetzen kann, um eine Qualitätssicherung zu erreichen.

3.6. Weitere Eigenschaften

Neben den Entwurfsmustern zur Datenerhebung und zur Qualitätssicherung finden sich vor allem in den Arbeiten [129], [2], [128], [115], [20], [103], [154], [156], [155], [127] und [5] die Eigenschaften *Wiederspielbarkeit*, *Verteiltes Spielen* und *Asynchrones und langsames Spielen*, die wichtig für das

Design von orstbezogenen Spielen zur Datenerhebung anzusehen sind. Die genannten Eigenschaften bestätigen auch in meiner Analyse der gesammelten Spiele aus Abschnitt 3.1 allgemein ihre Relevanz.

An dieser Stelle soll aber nur die Auswirkung auf ortsbezogene Spiele, die zur Datenerhebung entworfen werden sollen, aufgezeigt werden.

3.6.1. Wiederspielbarkeit

Eine der wichtigsten Eigenschaften aller Typen von Daten, die ich in Abschnitt 3.3 besprochen habe, ist ihre Aktualität. Ob es nun durch Baustellen versperrte Straßen, neu eröffnete Restaurants oder geänderte Öffnungszeiten des Rathauses sind. Alle diese Informationen über die reale Umwelt sind vor allem dann interessant, wenn sie zeitnah zur Verfügung stehen.

Laut [189] und [5] brauchen zumindest in Deutschland die Landesvermessungämter eine längere Zeitspanne, angefangen von ein paar Monaten bis hin zu mehreren Jahren, bis eine Änderung im Datenbestand - z.B. eine veränderte Verkehrslage - sich in den Navigationsgeräten der Nutzer wiederfindet.

Um diese Zeitspanne zu verkürzen und die Möglichkeit zu haben, immer wieder neue Daten mit demselben Spiel zu erheben, muss ein orstbezogenes Spiel zur Datenerhebung ein hohes Maß an Wiederspielbarkeit aufweisen. Dabei ist zu bedenken, dass die Anzahl an Spielern für ein ortsbezogenen Spiel begrenzt ist. Das Spielfeld eines ortsbezogenen Spieles ist immer an einen bestimmten Ort gekoppelt. Damit können nur diejenigen Spieler an einem solchen Spiel teilnehmen, die sich dort für eine längere Zeit aufhalten, z.B. dort wohnen. Sieht man von regelmäßigen Besucherströmen in touristisch interessanten Orten ab, hat man also nur eine begrenzte Spielerzahl als Datenlieferanten zur Verfügung.

Anders als bei einem reinen Web-basierten Spiel zur Datenerhebung, wie das *ESP-Game* [19], ist es bei der Erhebung der Typen von Daten aus Abschnitt 3.3 auch nur sehr eingeschränkt möglich, entsprechende Daten aus Luftbildern oder ähnlichen Quellen online extrahieren zu lassen. Auch ist die Kenntnis der geographischen Gegebenheiten lokal ansässigen Spielern viel besser geläufig, als Spielern aus anderen Orten und somit die Qualität der erhobenen Daten tendenziell besser. In Kapitel 5 Abschnitt 5.2 werde ich diesen Aspekt noch etwas näher betrachten.

Die Wiederspielbarkeit kann auf unterschiedliche Art und Weise erreicht werden. Eine Möglichkeit ist es, das Spiel so zu entwerfen, dass es theoretisch unendlich lange gespielt werden kann. *Feeding Yoshi* [128] ist hierfür ein Beispiel. Durch das einzige Ziel des Spieles möglichst viele Punkte zu erreichen,

kann das Spiel solange laufen, bis die Spieler selbst das Ende des Spieles beschließen. Wenn z.B. eine Spieler praktisch unerreichbar viele Punkte gesammelt hat.

Eine andere Möglichkeit ist es, das Spiel zwar in abgeschlossenen Spielrunden zu gestalten, aber den Spielern für jede Runde abgeänderte Ausgangsbedingungen zu bieten. Beispiele hierfür sind missions-basierte Spiele wie z.B. *Gopher* [129], *MobiMission* [2] oder *EyeSpy* [156], bei denen eine Spielrunde sozusagen aus einer gespielten Mission besteht. Da der Nachschub an Missionen von den Spielern gesteuert wird, läuft dieses Spiel ebenfalls solange, wie die Spieler daran aktiv daran teilnehmen möchten.

Im Anwendungsbeispiel in Abschnitt 3.8 werde ich noch eine weitere Möglichkeit vorstellen, wie ein hohes Maß an Wiederspielbarkeit erreicht werden kann. Dabei geht es darum, dass man Parameter für ein Spielelement vor jedem Spiel von den Nutzern selbst festlegen lässt.

3.6.2. Asynchrones und langsames Spielen

Neben der Wiederspielbarkeit sind zwei weitere Eigenschaften von Interesse, um die Spieler zu motivieren möglichst permanent aktuelle Daten zu erheben: Asynchrones [119], [128]) und langsames Spielen [127].

Asynchrones Spielen bedeutet im Kontext dieser Arbeit, dass die Spieler unabhängig voneinander in ein Spiel ein- bzw. aussteigen können. Eine Möglichkeit dies zu erreichen, ist, das Spiel nicht in Echtzeit ablaufen zu lassen, sondern wie in [119] vorgeschlagen, die ausgeführten Aktionen der Spieler in diskreten Zeitpunkten zu sammeln. Die Spieler haben eine bestimmte Zeitspanne zur Verfügung, seien es Minuten oder Tage, ihre Spielaktionen zu wählen und das Spiel arbeitet diese dann nach Ablauf der Zeitspanne entsprechend ab. Diese Zeitpunkte können fix vorgegeben oder dynamisch sein, z.B. könnte bei einer Interaktion zwischen zwei Spielern, ein Spieler die Interaktion zu einem beliebigen Zeitpunkt beginnen und der andere Spieler hätte dann einen Tag Zeit, aktiv darauf zu reagieren. Falls er dies nicht tut, könnte ein bestimmtes vordefiniertes Verhalten angenommen werden.

Eine Frage, die sich im Zusammenhang mit einer asynchronen Spielweise stellt, ist bisher aus meiner Sicht noch zu wenig beachtet worden. Und zwar: Welche Art von Spielregeln verhindern, dass die dominierende Gewinnstrategie bei asynchronen Spielen diejenige ist, einfach mehr Zeit in das Spiel zu investieren als alle andere. Bell et al. (2006) [128] erkannten das Problem schon in ihrer Untersuchung zu *Feeding Yoshi*. Um dieses Problem anzugehen, schlugen die Autoren vor, die Anzahl der möglichen Spielaktionen pro -

in ihrem Fall - Spieltag auf eine bestimmte Anzahl zu begrenzen.

Will man aber möglichst viele Daten mit einem solchen Spiel erheben, ist eine solche Begrenzungsstrategie nicht erstrebenswert, da man ja gerade eine möglichst hohe Anzahl an Daten pro Spielrunde erhoben haben möchte. Im Anwendungsbeispiel (siehe Abschnitt 3.8) diskutiere ich eine Strategie, um diesem Problem zu begegnen. Diese findet auch in [119] Erwähnung.

Ein langsames Spiel bedeutet auf der einen Seite, dass die Spieler ihre Aktionen ohne Zeitdruck durchführen können und auf der anderen Seite, dass die Spielzeit selbst, eine längere Zeitspanne (mehrere Tage oder gar Monate) beträgt. Die kürzere Spielzeit bei einer synchroner Spielweise ist zum großen Teil dadurch geschuldet, dass die Spieler sich zur selben Zeit am selben Ort befinden müssen, um das Spiel zu spielen. Einen Termin für eine Spieldauer von maximal ein paar Stunden zu organisieren ist schon schwierig genug, aber bei einer Spielzeit über mehrere Tage beinahe unmöglich. Aus diesem Grund wird meist ein langsames Spiel mit einer asynchrone Spielweise verbunden.

3.6.3. Verteiltes Spielen

Durch ein asynchrones und langsames Spiel ist die Möglichkeit gegeben, das Spiel einfacher in den normalen Tagesablauf der Spieler zu integrieren. Trotzdem kann es aber vorkommen, dass die Spielerbasis - vor allem bei einem neuen ortsbezogenem Spiel - an dem Ort, wo ich das Spiel spielen möchte, nicht groß genug ist, um dort regelmäßig Spiele durchzuführen.

Kiefer et al. (2007) [4] stellen eine Möglichkeit vor, beliebig viele Spielfelder in geographisch getrennten Orten aufeinander abzubilden. Damit können Spieler aus einem Ort mit Spielern aus einem anderen Ort zusammenspielen, obwohl sie vielleicht 100 km voneinander entfernt liegen. In ihrer Arbeit präsentieren sie drei Eigenschaften eines Spielfeldes, die bei der Überprüfung der Kompatibilität zwei oder mehr Spielfelder zueinander ausschlaggebend sind. Die Eigenschaften sind: Das räumliche Ausmaß (*spatial scale*), die statischen (*static nature*) und die dynamischen Merkmale (*dynamic conditions*) des Spielfeldes. Die dynamische Merkmale lasse ich an dieser Stelle unbetrachtet, da sie wie in [4] beschrieben, nur sehr begrenzt durch den Designer beeinflussbar sind. Dynamsiche Merkmale sind beispielsweise das Wetter oder das Verkehrsaufkommen in einem Spielfeld.

Das räumliche Ausmaß - wobei laut [4] auch die Form des Spielfeldes zählt - und auch die statischen Merkmale - beispielsweise das Straßennetzwerk bzw. das Höhenprofil - beeinflussen hauptsächlich die Laufwege zwischen

den spielrelevanten Orten im Spielfeld. Je unterschiedlicher die Spielfelder in zwei Orten sind, z.B. eines 500 auf 500 Meter groß, mit einer eher flachen Topographie, und das andere 2000 auf 2000 Meter, zusätzlich mit einer Anhöhe in der Mitte, desto unfairer wird das darauf gespielte Spiel empfunden. Dies ist vor allem bei einem synchronen und auf Schnelligkeit ausgelegten ortsbezogenen Spiel, wie bei dem in [4] vorgestellten *CityPoker*, der Fall.

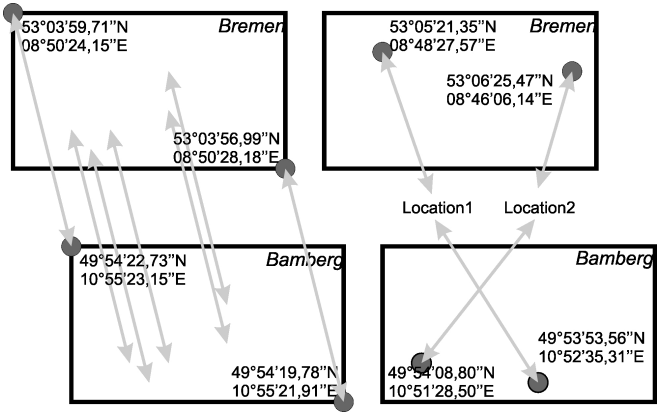


Abbildung 3.5.: direkte und indirekt Interaktion zwischen Spielern in einem verteilten orstbezogenen Spiel (aus [4] entnommen)

Wie ich in Abschnitt 3.2 ausgeführt habe, hängt der Grad der Strukturierung eines Spielfeldes damit zusammen, wo man im Spielfeld spielrelevante Aktionen durchführen kann. In Bezug auf die Abbildung von Spielfeldern ist aber noch eine andere Art der Interaktion wichtig, welche auch Abbildung 3.5 aufzeigt: Ob eine direkte Interaktion zwischen den Spielern möglich ist oder nicht.

Laut Definition ist eine direkte Interaktion zwischen Spielern nur bei semi- bzw. unstrukturierten Spielfeldern möglich. In beiden Fällen können die Spieler teils oder sogar ganz frei die Orte im Spielfeld aussuchen, an denen sie Aktionen ausführen. Sind diese Orte gleichzeitig die Standpunkte der Mitspieler, muss bei einem verteiltem Spiel sichergestellt sein, dass alle von den Spielern zugänglichen Orte im einen Spielfeld auch auf dem anderen Spielfeld zugänglich sind (siehe linke Spielfeldabbildung in Abbildung 3.5). Dagegen müssen bei einem vollständig strukturiertem Spielfeld nur die vorher festgelegten spielrelevanten Orte auf beiden Spielfeldern erreichbar sein (siehe rechte Spielfeldabbildung in Abbildung 3.5).

Bei einem ortsbezogenen Spiel zur Erhebung von raumbezogenen Daten ist aber die direkte Interaktion zwischen den Spielern erst einmal zweitrangig, da ja primär Daten über im Spielfeld befindliche Objekte (wie Gebäude, Straßen oder Plätze) gesammelt werden sollen und nicht über den Standort der gegnerischen Spieler. Zusätzlich sollen mit einem solchen Spiel die statischen Merkmale erhoben werden und können somit nicht schon vorher zur Berechnung von Ähnlichkeitsmaßen, wie in [4] vorgeschlagen, zwischen den Spielfeldern verwendet werden. Auch spielen die aus den statischen Merkmalen resultierenden Laufwege bei einem asynchronen und langsamen Spiel nur eine untergeordnete Rolle.

Deshalb lässt sich an dieser Stelle festhalten:

Je unstrukturierter das Spielfeld ist, desto besser (einfacher) lassen sich geographisch getrennte Spielfelder aufeinander abbilden und desto besser (quantitativ) ist auch das Spielfeld zur Datenerhebung geeignet.

Ich werde in Abschnitt 3.8 zeigen, dass, wenn man die statischen Merkmale außer Acht lassen kann/muss, bei einem unstrukturiertem Spielfeld nur das räumliche Ausmaß von Interesse für ein Abbildung der Spielfelder ist.

3.7. Design Framework

Fassen wir nun alle vorangegangenen Aspekte zusammen, die beim Design eines ortsbezogenen Spiels zur Datenerhebung eine Rolle spielen, ergibt sich folgendes in Abbildung 3.6 gezeigtes Design Framework.



Abbildung 3.6.: Schematischer Aufbau des Design Framework zur Erstellung von ortsbezogenen Spielen zur Datenerhebung

Auf der untersten Ebene des Framework befindet sich die Strukturierung des Spielfeldes (Abschnitt 3.2), da diese maßgeblich den Typus von Daten

(Abschnitt 3.3) vorgibt, den man mit dem fertigen Spiel erheben kann.

Darauf aufbauend werden entsprechende Spielelemente zur Motivation der Datenerhebung (Abschnitt 3.4) und zur Qualitätssicherung (Abschnitt 3.5) durch die Spieler eingebaut. Das Ganze wird dann in einen thematischen Rahmen eingebettet. Alternativ kann man sich auch ausgehend von einem thematischen Rahmen leiten lassen, um sich darauf aufbauend für eine passende Spielfeldstrukturierung und Spielelemente zur Datenerhebung und Qualitätssicherung entscheiden.

Den Schritt der thematischen Einbettung betrachte ich in der vorliegenden Arbeit nicht mehr. Die vorgestellten Entwurfsmuster sind allgemein gehalten und können deshalb einfach in jede beliebige Thematik übertragen werden. Zum Beipiel kann beim *Transporter*-Entwurfsmuster der Spieler die Rolle eines Flussschiffers im Mittelalter, der Güter entlang eines Flusses handelt [3], leicht gegen den eines Postboten in der Neuzeit eintauschen, ohne dass sich die Spielmechanik selbst ändert.

Fasst man die einzelnen Entscheidungen, die beim Design getroffen werden müssen, in einer Abbildung zusammen, ergibt sich das in Abbildung 3.7 gezeigte Flowchart.

Ich werde nun im Folgenden diesen Designprozess anhand eines Anwendungsbeispieles genauer betrachten.

3.8. Anwendungsbeispiel: CityExplorer

Grundsätzlich erlaubt das Design Framework die Entwicklung beliebiger Spielkonzepte. Um aber den Schritt der thematischen Einbettung zu vereinfachen, habe ich für die Umsetzung eines spielbaren Prototypes auf ein schon bekanntes nicht-orstbezogenes Spiel zurückgegriffen: Das Brettspiel *Carcassonne*. In diesem von Klaus-Jürgen Wrede [4] entwickeltem preisgekrönten Spiels - Spiel des Jahres von 2001 - werden von den Spielern abwechselnd quadratische Spielplättchen an das dadurch wachsende Spielbrett gelegt; d.h. am Anfang besteht das Spielbrett nur aus einem einzigen Plättchen und wächst dann rundenweise um die Anzahl an Spielern an. Die Spieler ziehen verdeckt jede Runde ein Spielplättchen, welches sie als nächstes legen müssen, aus dem Stapel an Restplättchen. Wenn alle Plättchen gelegt sind, endet das Spiel. Derjenige Spieler mit den meisten Punkten am Ende des Spieles gewinnt.

[3] wie im Spiel *Leinen los!* - `www.kinf.wiai.uni-bamberg.de/flupa`

[4] `http://www.carcassonne.de/`

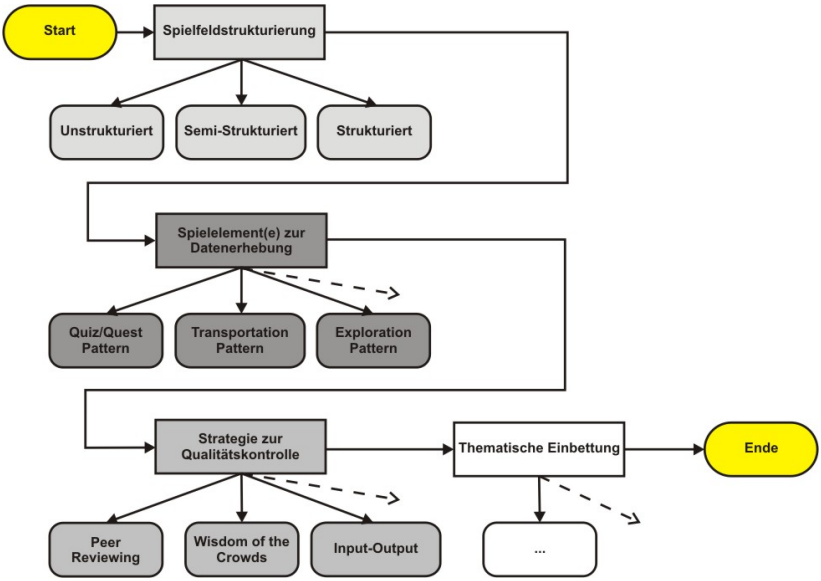


Abbildung 3.7.: Flowchart für den Designprozess eines ortsbezogenen Spieles zur Datenerhebung

Punkte erhalten die Spieler, wenn Sie ihre Spielsteine - von denen jeder nur eine begrenzte Anzahl zur Verfügung hat - auf bestimmte Merkmale der Spielplättchen platzieren (siehe rechtes Foto in Abbildung 3.8), wie Straßen, Städte, Kirchen und Wiesen. Werden diese durch das Legen eines anderen Plätzchens „abgeschlossen“ (siehe linkes Foto in Abbildung 3.8), bekommen die Spieler der zugehörigen Spielsteine entsprechend Punkte. Zusätzlich bekommt man seinen Spielstein zurück und kann ihn beim Legen des nächsten Plätzchens wiederverwenden.

Die drei Kern-Spielmechanismen von Carcassonne sind:

- Das Spielbrett ist anfangs leer.
- Das Spielbrett wird erst im Spielverlauf mit spielrelevanten Orten gefüllt.
- Es sind zu Spielbeginn bestimmte geographische Merkmale festgelegt, auf denen die Spieler ihre Spielsteine setzen können und für die sie dann Punkte erhalten.

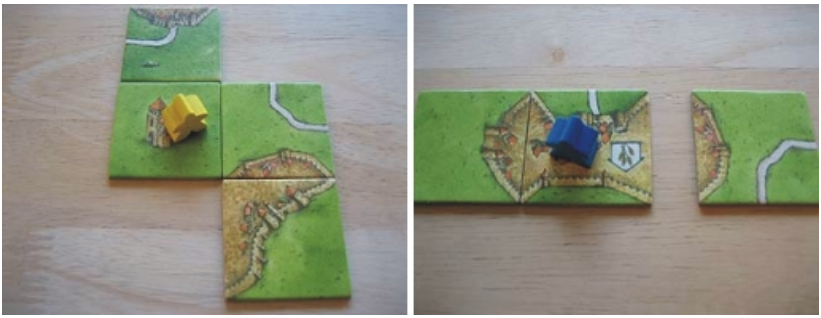


Abbildung 3.8.: Wachsendes Spielbrett und Anlegen eines neuen Spielplättchens in Carcassonne; Photos von `http://www.brettspiele-report.de/`

Mit Hilfe des Design Frameworks wurde Carcassonne folgendermaßen in ein ortsbezogenes Spiel umgesetzt:

3.8.1. Spielregeln

Entsprechend dem ersten Kern-Spielmechanismus von Carcassonne habe ich als Strukturierung des Spielfeldes *„Unstrukturiert“* gewählt. Durch das unstrukturierte Spielfeld bietet sich das *Erkundung*-Entwurfsmuster als Spielelement zur Datenerhebung an. Die Spieler „entdecken“ alle spielrelevanten geographischen Merkmale erst während des Spiels. Dies entspricht auch am Besten dem verdeckten Ziehen der Spielplättchen. Die Spieler „setzten“ ihre Spielsteine wie im Original an geographische Merkmale im Spielfeld.

Um die *Wiederspielbarkeit* zu gewährleisten, sollen die Merkmale nicht bei jedem CityExplorer Spiel gleich sein. Deswegen kann jeder Spieler bis zu drei Merkmalskategorien zu Beginn des Spieles festlegen, an die alle Spieler ihre Spielsteine legen dürfen. Um das Spiel nicht allzu komplex werden zu lassen - auch im Hinblick auf eine noch festzulegende Qulitätssicherung - soll eine Merkmalskategorie nur nicht-bewegliche, durch eine einzige GPS-Messung zu identifizierende geographische Orte bzw. Objekte beinhalten. Solche Merkmalskategorien können z.B. Restaurant, Brunnen oder auch Einbahnstrassenschild sein. Damit stehen dann den Spielern alle Orte im Spielfeld zum Setzen von Spielsteinen offen, die zu den drei genannten Kategorien gehören.

Trotz des Namens kann CityExplorer auch außerhalb einer städtischen Um-



Abbildung 3.9.: Die drei Spielphasen in CityExplorer: Kategorien festlegen, Spielsteine setzen und Endabrechnung, © Karte: Google Maps (`http://maps.google.de/`).

gebung gespielt werden. Gerade durch die freie Wahl der im Spiel verwendeten Kategorien für die Spielsteine, ist CityExplorer an jedes Spielfeld bzw. geographische Umgebung anpassbar.

Alle Spielsteine werden virtuell gesetzt. Das Setzen der Spielsteine wird dabei mit Hilfe von mobilen Geräten, die über eine GPS Lokalisierung verfügen, durchgeführt und enthält folgende Schritte:

1. Die Spieler machen ein Foto von dem Ort, dem Gebäude oder dem Objekt, an dem sie ihren Spielstein setzen möchten.

2. Die Spieler bewegen sich so nah wie möglich an den Ort, das Gebäude oder das Objekt heran, welches sie fotografiert haben.

3. Dort angekommen, wählen sie dann die entsprechende Merkmalskategorie, zu der der Spielstein gehören soll.

Bei Punkt 1 und 3 werden von der Spieleapplikation automatisch GPS Koordinaten aufgezeichnet, d.h. als raumbezogene Daten erhebt das Spiel ein georeferenziertes Foto und eine semantisch annotierte GPS Koordinate (durch die Kategorisierung). Ein Beispiel findet sich in Listing 3.1. Aus den zwei GPS Koordinaten kann dann später zusätzlich noch der Richtungsvektor vom Ort der Fotoaufnahme hin zum Fotoobjekt berechnet werden. Diese recht komplexen Daten ermöglichen mir in Kapitel 4 die Realisierung eines neuartigen Aggregationsverfahrens, um die Qualität der erhobenen Daten weiter zu verbessern.

Eine weitere Eigenschaft der Spielregeln, die später eine Rolle spielen wird, ist die erlaubte Mehrfachaufnehme von Merkmalen. Die Spieler können

prinzipiell beliebig viele Spielsteine an ein und dasselben Merkmal im Spielfeld setzen. Ich werde ebenfalls in Kapitel 4 zeigen, warum ein gewisser Grad an Redundanz bei der Aufnahme von räumlichen Daten von geographischen Merkmalen von Vorteil ist.

Um das strategische Element des Brettspiels so gut wie möglich zu erhalten und das Legen der Spielplättchen nachzuahmen, begrenze ich das räumliche Ausmaß des Spielfeldes und teilen es in quadratische Regionen ein. Da, wie im Original, keine geographischen Merkmale „abgeschloßen" werden können, führe ich die Regel ein, dass derjenige zum Ende des Spiels Punkte bekommt, der in einer Spielfeldregion die meisten Spielsteine gesetzt hat. In der jetzigen Version sind es zwei Punkte pro so gewonnene Region.

Ein weiteres strategisches Element ist der Gewinn einer Kategorie. Ein Spieler gewinnt eine Kategorie, indem er, auf das gesamte Spielfeld betrachtet, am meisten Spielsteine von dieser Kategorie während des Spieles gesetzt hat. Somit muss man nicht nur die Anzahl an Spielsteinen in den einzelnen Regionen im Auge behalten, sondern sich auch überlegen, von welcher Kategorie man Spielsteine in einer Region setzten möchte. Gegenwärtig erhält man für den Gewinn einer Kategorie einen Punkt in der Endabrechnung.

Das Spiel läuft nicht rundenbasierend ab, wie das Brettspiel, sondern die Spieler können jederzeit innerhalb einer von den Spielern frei wählbaren Gesamtspieldauer - welche von einigen Stunden bis hin zu mehreren Wochen oder Monaten dauern kann - Spielsteine im Spielfeld verteilen. Der aktuelle Spielstand wird dabei über eine Webseite verwaltet. Dadurch wird ein asynchrones Spiel ermöglicht, d.h. die Spieler können den Zeitpunkt, an dem sie das Spiel spielen wollen, selbst bestimmen. Abbildung 3.9 zeigt den Anfangszustand eines Spielfeldes, das Setzen eines Spielsteines und den Spielstand bei Spielende für ein Spiel in der Innenstadt von Bamberg.

Wird die Spieldauer entsprechend lange gewählt, wird auch eine langsame Spielweise unterstützt. Somit kann das Spielgeschehen dann, wie bei *Feeding Yoshi* [128], in den normalen Tagesablauf integriert werden.

Als Qualitätssicherungsstrategie habe ich das *Peer Reviewing*-Entwurfsmuster umgesetzt. Alle gesetzten Spielsteine können von allen Spielern überprüft werden. Im Vordergrund der Überprüfung steht dabei die Korrektheit der GPS Koordinate des Merkmales und dessen Kategorisierung, wobei hierfür das gemachte Foto als Kriterium herangezogen werden kann.

Das *Peer Reviewing*-Verfahren läuft dabei folgendermaßen ab:

1. Alle Spieler, außer dem Besitzer des Spielsteines, können einen unbewerteten Spielstein - unter Angabe eines Grundes - ablehnen oder



Abbildung 3.10.: Web-basierter Wizard für das Peer-Reviewing Verfahren, © Karte: Google Maps (`http://maps.google.de/`).

annehmen. Einen Screenshot des zugehörigen web-basierten Wizards zeigt Abbildung 3.10.

2. (OPTION 1:) Nimmt ein beliebiger Spieler einen Spielstein an, wird er als korrekt markiert (*approved marker*) und kann nicht mehr von anderen abgelehnt werden.

3. (OPTION 2:) Wird ein Spielstein von einem beliebigen Spieler abgelehnt, fällt er erst einmal aus der Wertung heraus und wird als inkorrekt markiert (*open marker*).

4. Ein abgelehnter Spielstein kann von seinem Besitzer einmal überarbeitet werden und neu in das System eingestellt werden.

5. Für einen neu eingestellten Spielstein beginnt das Verfahren bei Punkt 1 von vorne, mit der Einschränkung, dass zusätzlich zum Besitzer nun auch der erste Gutachter den Spielstein nicht bewerten kann.

6. (Abbruchbedingung:) Ist ein Spielstein angenommen oder zweimal abgelehnt worden, ist der Begutachtungsprozess beendet und das letzte Ergebnis endgültig.

Um die Spieler zu motivieren, die Überprüfung der Spielsteine durchzuführen - da der Nachteil des *Peer Reviewing*-Entwurfsmusters die eher lose Kopplung zum eigentlichen Spielgeschehens darstellt - bekommt derjenige Spieler mit den meisten Begutachtungen währende des gesamten Spieles drei Punkte in der Endabrchung.

Das spielergetriebene Begutachtungsverfahren bietet die Möglichkeit das Spiel an die bevorzugte Spielweise der Spieler anzupassen. Wird z.B. das mehrfache Setzen von Spielsteinen an einem Merkmal als unfair empfunden und soll deshalb unterbunden werden, können die Spieler einfach alle Spielsteine ablehnen, die ihrer Meinung nach ein Duplikat eines schon vorhandenen Spielsteines sind. Wie bei allen gemeinschaftlichen Entscheidungssystemen ist ein „böswilliger“ Spieler immer gut in den Griff zu bekommen, sind es jedoch mehrere, stößt das Verfahren an seine Grenzen.

3.8.2. Zusatzregeln für ein verteiltes Spiel

CityExplorer ermöglicht es n Spielfelder miteinander zu verknüpfen. Bei der Abbildung der einzelnen Spielfelder aufeinander wird als Kriterium allein das räumliche Ausmaß herangezogen. Dies hat, wie schon in Abschnitt 3.6.3 angedeutet zwei Gründe. Erstens finden im Spiel keine direkte Interaktion zwischen den Spielern statt. Zweitens ist das Spiel auf eine asynchrone und langsame Spielweise ausgerichtet. Daher ist eine genaue Abbildung der statischen Merkmale nicht notwendig.

Ein zusätzlicher Aspekt, der aber nur spezifisch für CityExplorer ist, muss dennoch beachtet werden. Die Kategorien, die die Spieler am Anfang eines Spiels festlegen, müssen für alle Spielfelder gleich sein. Dies muss aus Gründen der Fairness gewährleistet sein, damit jeder Spieler in jedem Spielfeld die Möglichkeit hat, die Kategoriewertung zu gewinnen.

Außerdem sollten pro Spielfeld mindestens drei Spieler teilnehmen, damit das Peer-Reviewing Verfahren durchgeführt werden kann. Von dieser Regel kann aber abgewichen werden, wenn sich genügend Spieler eines Spielfeldes im Spielfeld der anderen Spieler auskennen. Prinzipiell könnten zwar bei einem verteilten Spiel alle Spieler alle Spielsteine begutachten. Nur ist es natürlich für Spieler schwierig bis gar unmöglich die Korrektheit der Standorte von Merkmalen in einem vielleicht mehrere tausend Kilometer entfernten Ort zu überprüfen, wenn sie dort noch nie für längere Zeit gelebt haben.

Abbildung 3.11 zeigt ein frühes Screendesign für ein verteiltes Spiel zwischen Bamberg (Deutschland) und Shonandai (Japan - nahe Yokohama). Das

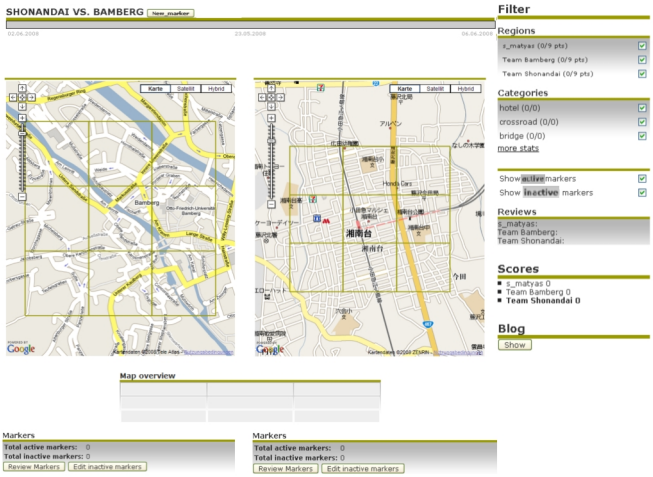


Abbildung 3.11.: Screendesign für ein verteiltes CityExplorer Spiel, © Karte: Google Maps (`http://maps.google.de/`).

räumliche Ausmaß ist identsich, jedoch sind die Zoomstufen in Google Maps[5] in beiden Städten unterschiedlich. Konzeptionell interessant ist an dieser Stelle nur der Punkt „Map Overview“, der die einzelnen Regionen der zwei Spielfelder aufeinander abbildet, d.h. es wird zum Beispiel die rechte obere Region in Bamberg auf die rechte obere Region in Shonandai abgebildet.

In Kapitel 5.1 werde ich genauer auf die Details zu durchgeführten Spielen von *CityExplorer* eingehen, unter anderem auch, wie ein verteiltes Spiel als Teamspiel durchgeführt werden kann.

3.8.3. Systemarchitektur

Das CityExplorer Spiel besteht aus zwei Komponenten: Einer mobilen Anwendung und einer Serveranwendung. Beide Komponenten kommunizieren entweder direkt, über HTTP-Requests, oder indirekt, über ein XML-basiertes Austauschformat, miteinander. Abbildung 3.12 zeigt den Aufbau, der in CityExplorer verwendeten Komponenten und deren Interaktionsmöglichkeiten.

[5] `www.maps.google.de`

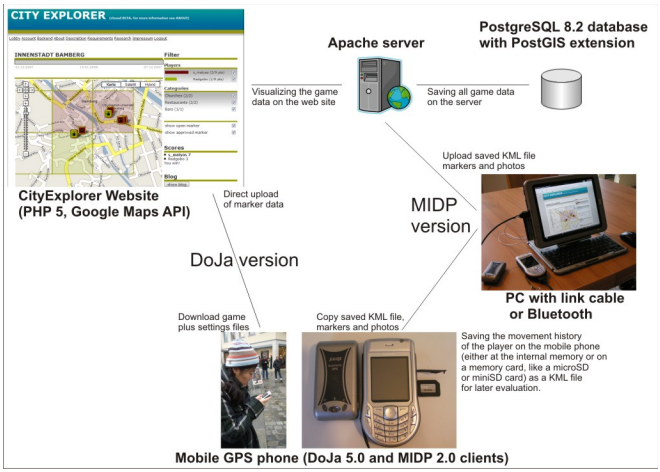


Abbildung 3.12.: Aufbau der Systemarchitektur von CityExplorer

Serverkomponente

Auf dem Server erfolgt die Verwaltung der Spielerdaten (Login-Daten und persönliche Blogs), der erhobenen raumbezogenen Daten (GPS Daten, semantische Daten und der Fotos) und das Durchführen der einzelnen Spielrunden. Letzteres beinhaltet das Anlegen von neuen Spielfeldern, das Konfigurieren und Starten einer neuen Spielrunde (Festlegen der Spielzeit, des Spielfeldes und der eigenen drei Spielsteinkategorien) und der Visualisierung des aktuellen Spielzustandes.

Für die Umsetzung - Abbildung 3.13 zeigt zu den oben genannten Aspekten Screenshots - wurde ein Apache Server [6] aufgesetzt und eine Webseite in PHP5 [7] entwickelt. Zur Visualisierung der Spielfelder wurde die Google Maps API [8] verwendet und über Java Script in die PHP Seite eingebunden. Die Datenhaltung erfolgt durch eine PostgreSQL (8.2) Datenbank [9], die durch die PostGIS Erweiterung [10] um Funktionen für die Haltung und Verarbeitung von raumbezogenen Daten erweitert wurde. PostGIS entspricht dem Open-

[6] http://www.apache.org/
[7] http://www.php.net/
[8] http://code.google.com/intl/de-DE/apis/maps/
[9] http://www.postgresql.org/
[10] http://postgis.refractions.net/

GIS „Simple Features Specification for SQL“ Standard [11] und wurde als kompatibel mit dem „Types and Functions“ Profil zertifiziert.

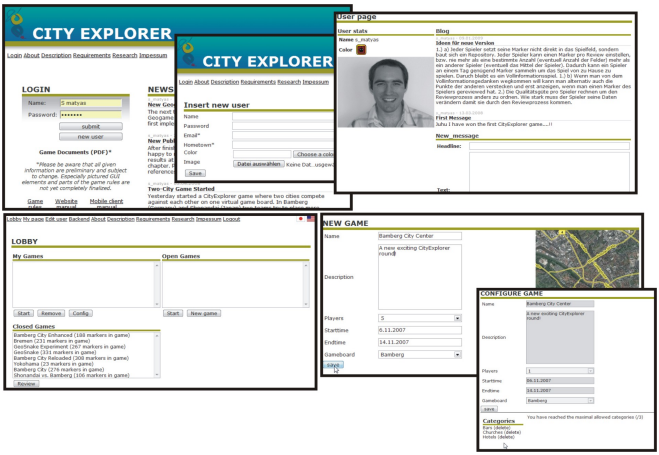


Abbildung 3.13.: Login-Screen, neuen Nutzer anlegen, persönliche Seite mit Blog, Spiel-Lobby, neues Spielfeld anlegen, Spiel starten (von links oben nach rechts unten), © Karte: Google Maps (http://maps.google.de/).

Mobile Anwendung

Bei der Entwicklung der mobilen Anwendung habe ich mich auf die zwei, zumindest 2007/2008, am weitest verbreitesten und damit am Besten unterstützten Software-APIs für mobile Geräte beschränkt: Die Java Micro Edition (J2ME)[12] (Version 2.0) und das .NET Compact Framework[13] (.NET CF Version 3.5). Mein Ziel war es, als primäre Gerätemodelle die der Nokia S60 Serie (2. Generation) und Geräte mit Windows Mobile 6 oder älterem Betriebssystem abzudecken.

Nokia Geräte bieten gegenüber anderen vergleichbaren Geräten zwei Vorteile. Erstens unterstützen die meisten folgende, für die mobile Anwendung nötige, J2ME Zusatz-API-Pakete: Java APIs for Bluetooth (JSR 82), Mobile Media API (JSR 135), PDA Optional Packages (JSR 75, enthält Klassen,

[11] `http://www.opengeospatial.org/standards/sfs`
[12] `http://java.sun.com/javame/index.jsp`
[13] `http://msdn.microsoft.com/en-us/netframework/aa497273.aspx`

um auf die Speicherkarte zuzugreifen) und die Wireless Messaging API (JSR 120). Zweitens, und das trifft vor allem für das bei den Anwendungsfällen verwendete Nokia 6630 zu (siehe Kapitel 5 Abschnitt 5.1), kann das Midlet mit Hilfe eines selbst erstellten Zertifikats signiert werden.

Dieses Zertifikat muss auf dem mobilen Gerät ebenfalls installiert werden, was ebenfalls nur bei den verwendeten Nokia Geräten möglich ist. Dadurch bekommt man erst die Möglichkeit das Midlet als „Trusted Third Party“ Anwendung zu installieren. Das hat den Vorteil, dass mehr Sicherheitseinstellungen am Gerät geändert werden können, wie z.B. das Speichern und Lesen von Daten auf der Speicherkarte. Normalerweise kann die Sandbox, in der eine J2ME Midlet auf dem Gerät läuft, nicht verlassen und nur sehr einfache Datenstrukturen persistent gespeichert werden. Da ich aber sowohl Fotos, die Bewegungsspur des Nutzers (im Durchschnitt wird alle zwei Sekunden die aktuelle Position gespeichert) als auch komplexere semantische Daten erhebe und das Ganze nur indirekt auf den Server hochladen kann, musste ich ungehinderten Zugriff auf das Speichermedium haben. Ohne Signierung des Midlets ist dies zwar auch möglich, aber der Nutzer muss dann bei jedem Zugriff (lesend und schreibend) einen entsprechenden Sicherheitsdialog beantworten. Diesen Dialog kann man mit dem Signieren umgehen.

Für den Anwendungsfall des verteilten Spielens von *CityExplorer* entstand zusätzlich noch eine Version für NTT DoCoMo Geräte, entwickelt unter zu Hilfenahme des i-appli Entwicklerkits für das DoJa-5.0 (DoCoMo Java) Profil [14]. Da alle offiziellen Dokumentationen nur in japanischer Sprache vorlagen, habe ich mich bei der Entwicklung auf die Übersetzungen der Mobile Developer Lab Gruppe gestützt [15].

Da auf dem japanischen Markt fast alle Mobiltelefonverträge mit einer Datenflatrate verkauft werden, habe ich für diese Version den direkten Kommunikationsweg zwischen mobilem Gerät und Server gewählt. Das bedeutet, dass das Gerät einen HTTP-Request an die Seite mit den relevanten Daten schickt und zwar einmal das Foto mit GPS Koordinate und zum Zweiten eine GPS Koordinate mit entsprechender Kategorisierung, wie in Abschnitt 3.8.1 beschrieben.

Für die J2ME und die .NET CF Anwendungen wurde der indirekte Kommunikationsweg über ein XML-Austauschformat umgesetzt. Alle gesetzten Spielsteine werden zuerst auf der Speicherkarte des Gerätes gesichert und müssen dann mit Hilfe eines Rechners über die CityExplorer Webseite zum

[14] `http://www.nttdocomo.co.jp/english/service/imode/make/content/iappli/`

[15] `http://sites.google.com/site/mobiledevlab/`

Server hochgeladen werden. Die Beschreibung eines Spielsteines sieht im verwendeten XML-Fromat folgendermaßen aus:

```
1  <marker>
2      <category>3</category>
3      <lat_poi>49.889866</lat_poi>
4      <lng_poi>10.891302</lng_poi>
5      <tag_accuracy>2.6</tag_accuracy>
6      <tag_satinview>4</tag_satinview>
7      <player>30</player>
8      <image>Bamberg_City_Enhanced-1212343749328
           salino.png</image>
9      <lat_image>49.889816</lat_image>
10     <lng_image>10.891419</lng_image>
11     <image_accuracy>2.5</image_accuracy>
12     <image_satinview>5</image_satinview>
13     <caption>salino</caption>
14     <game_id>24</game_id>
15     <regionx>2</regionx>
16     <regiony>2</regiony>
17 </marker>
```

Listing 3.1: Beispiel für einen gestzten Spielstein

Anzumerken ist hier, dass die Spieler optional die Möglichkeit haben, für ein gemachtes Foto ein passendes Stichwort („caption“ in Zeile 13) anzugeben. Dies kann, wie hier geschehen, der Name des Restaurants sein, an das der Spielstein gesetzt wurde. Die Kategorieinformation verbirgt sich hinter der Zahl in Zeile 2 („category“) und ist unabhängig vom verwendeten Stichwort. Zurzeit wird das Stichwort allerdings in CityExplorer noch nicht verwendet.

Da die .NET CF Anwendung nur als Prototyp entwickelt wurde und bei den Spielen zur Evaluierung von CityExplorer (siehe Kapitel 5) nur die J2ME Version zum Einsatz kam, will ich an dieser Stelle nur kurz auf die J2ME Implementierung zu sprechen kommen (siehe hierzu auch Matyas (2007) [5]). Bei der Portierung auf DoJa für die NTT DoCoMo Geräte wurden lediglich diejenigen Teile des Programmcode umgeschrieben bzw. neu implementiert, die wegen der anders strukturierten API nötig waren. Dies betraf vor allem die GUI Klassen und die neu hinzugekommene direkte Interaktion mit dem Server. Dabei wurde aber die grundlegende Struktur des Geogames Software

Frameworks nicht verändert, weswegen ich eine detaillierte Beschreibung der DoJa Implementierung an dieser Stelle verzichte.

Wie im oberen Diagramm von Abbildung 3.14 gezeigt, besteht das Framework aus einem Anwendungsteil, einem Daten- bzw. Objektverwaltungsteil sowie einem Kommunikationsteil (siehe hierzu auch die Definition des ADK-Modells nach Ferstl und Sinz (1993) [192].

Der Anwendungsteil (*application management* [5]) enthält alle relevanten GUI Klassen (vor allem zum Anzeigen des Spielfeldes) und Klassen für die Spiellogik. Die Spiellogik beinhaltet dabei nicht nur die Spielregeln selbst, sondern insbesondere Klassen und Methoden zur Behandlung von ortsbezogenen Ereignissen (*Location Manager*), wenn sich z.B. die Spieler einem bestimmten Ort im Spielfeld nähern. Dabei sind zu diesem Zeitpunkt aber noch keine Klassen implementiert, die sich auf ein bestimmtes Spiel beziehen, sondern nur solche Funktionalitäten, die für alle möglichen Geogames nach [134] relevant sind. Daneben finden sich hier auch noch Klassen für diverse Zusatzfunktionalitäten, wie beispielsweise eine Klasse zum Generieren von Tönen.

Im Daten- bzw. Objektteil (*data management* [5])sind diverse Klassen zum Lesen der in CityExplorer verwendeten Dateitypen (Bitmap- und Textdateien) bzw. Settingsdateien und zum Erstellen der zugehörigen Datenobjekte implementiert.

Der Kommunikationsteil (*communication management* [5]) regelt den Daten- bzw. Eventaustausch sowohl zwischen den über Bluetooth verbundenen GPS Empfängern als auch die Interaktion zwischen den Geräten der Spieler. Das Event-Management informiert dazu den Anwendungsteil über Änderungen z.B. der Position der Spieler oder im Spielstatus. Außerdem bewirken die so erzeugten Events Änderungen an den Datenobjekten, die vom Datenmanagement im Framework erstellt wurden.

Mit Hilfe des Frameworks wurden die Geogames *GeoTicTacToe* [134] und *CityPoker* [4] umgesetzt. Entsprechend zu diesen zwei Spielen wurden, wie im unteren Diagram von Abbildung 3.14 zu sehen, für *CityExplorer* nur die spielspezifischen Klassen implementiert, allen voran die *game manager*-Klasse, die die Spielregeln verwaltet. Abbildung 3.15 zeigt Screenshots der J2ME und der DoJa Implementierung.

3.9. Zusammenfassung

In diesem Kapitel habe ich auf Grundlage einer Analyse von existierenden ortsbezogenen Spielen die drei Bestandteile - und deren Ausprägungen - für ein allgemein gültiges Design Framework vorgestellt: Drei mögliche *Strukturierungen des Realwelt-Spielfeldes*, fünf *Entwurfsmuster zur Motivation spielgetriebener Datenerhebung* und drei Strategien zur *spielergetriebenen Qualitätssicherung*. Dabei bin ich vor allem auf den Zusammenhang mit den unterschiedlichen Typen von sammelbaren Daten eingegangen.

Weiterhin wurden *Wiederspielbarkeit*, *asynchrones Spielen*, *langsames Spielen* und *verteiltes Spielen* als Eigenschaften besprochen, die einmal zum Spielspaß beitragen und zum Anderen die Quantität der erhobenen Daten erhöhen.

Alle diese Eigenschaften, ein *unstrukturiertes Spielfeld*, das *Erkundung*-Entwurfsmuster und das *Peer Reviewing*-Entwurfsmuster wurden bei der Entwicklung des ortsbezogenen Spieles von mir umgesetzt bzw. verwendet. Mit der Entwicklung von CityExplorer habe ich gezeigt, dass das vorgestellte Design Framework alle notwendigen Elemente enthält, die zur Erstellung eines ortsbezogenen Spieles zur qualitätsgesicherten Datenerhebung gehören. Die Beschreibung der technischen Umsetzung zeigt auf, welche softwaretechnischen Besonderheiten zu beachten sind, will man ein ähnliches ortsbezogenes Spiel selbst realisieren.

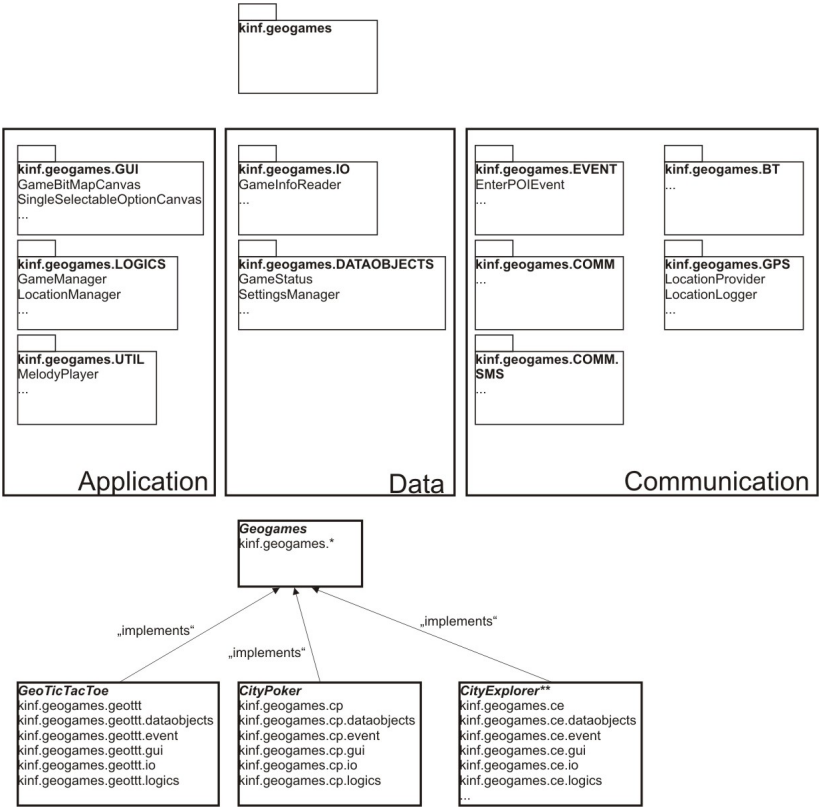


Abbildung 3.14.: Paketdiagram des Geogames J2ME Software Framework und die zugehörige CityExplorer Implementierung (entnommen aus Matyas (2007) [5])

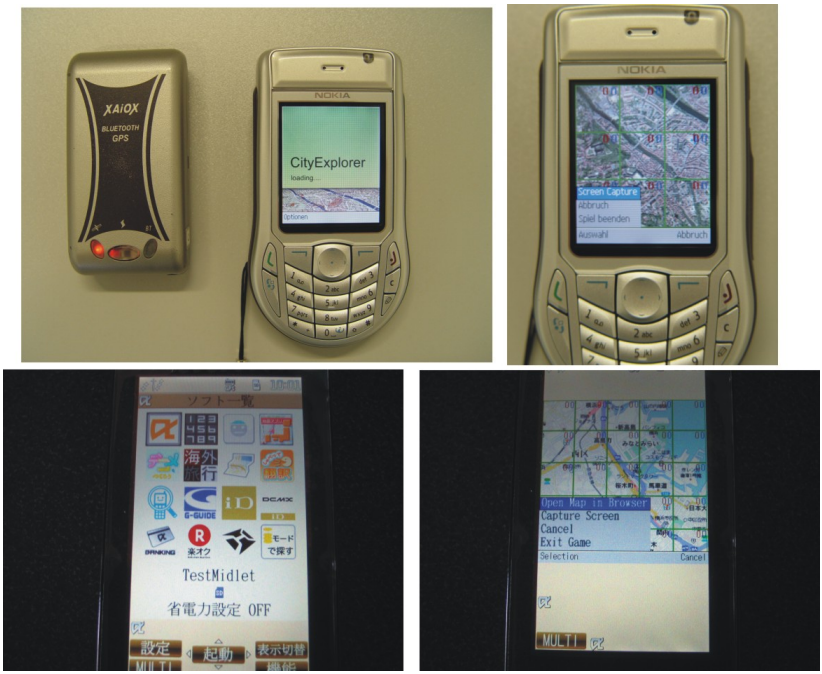


Abbildung 3.15.: Screenshots der J2ME Implementierung (obere Fotoreihe) und der DoJa (i-appli) Implementierung von CityExplorer (untere Fotoreihe), © Karte: Google Maps (`http://maps.google.de/`).

4. Semantische Integration raumbezogener Daten

Im folgenden Kapitel stelle ich meinen Lösungsvorschlag für die zweite Kernfragestellung der vorliegenden Arbeit vor:

Wie kann die Qualität von Sammlungen semantisch angereicherter raumbezogener Daten verbessert werden?

Das Ziel des im vorherigen Kapitel eingeführten Design Frameworks ist die Erstellung ortsbezogener Spiele mit dem geowissenschaftliche Laien raumbezogene Daten erheben. Gewisse Eigenschaften dieser Spiele, wie beispielsweise Wiederspielbarkeit, haben den Effekt dass Daten redundant erhoben werden. Diese Redundanz ist aber in diesem Fall durchaus gewollt, da sie dazu beiträgt, die Qualität der resultierenden Datensammlung weiter zu verbessern. Raumbezogenen Daten haftet je nach verwendeter Lokalisierungstechnologie ein gewisser Fehler an. Bei GPS ist dieser Fehler bezogen auf die Genauigkeit mit der die aufgenommene Koordinate der realen Position entspricht. Die Abweichung von der realen Position beträgt bei GPS in 50% der Fälle bis zu 5 Meter und in 95% der Fälle bis zu 15 Meter. Zieht man noch in Betracht, dass geowissenschaftliche Laien nicht immer über das nötige Wissen verfügen, ihre Hardware richtig einzusetzen bzw. die bestmöglichen Empfangsbedingungen zu erreichen, kann sich dieser Fehler noch vergrößern. Werden aber raumbezogene Daten für ein und denselbe Ort durch mehrere unterschiedliche Laien, bzw. Spieler, erhoben, kann durch eine Integration, bzw. Aggregation, der einzelnen Aufnahmen der Gesamtfehler verbessert werden.

Die Lösung der obigen Fragestellung baut auf folgenden Teilaspekten einer semantischen Integration raumbezogener Daten auf:

1. Zuerst stelle ich in Abschnitt 4.1 eine Konstellation von raumbezogenen Daten vor, die bisherige Integrations- bzw. Aggregationsverfahren nicht auflösen können.

2. Um diesen neuartigen Typus von Anwendungsfall lösen zu können, müssen die raumbezogenen Daten, die man integrieren möchte eine gewisse Struktur aufweisen. Diese Struktur findet sich in Abschnitt 4.1.1 in der formalen Definition der raumbezogenen Daten wieder, die mit Hilfe des in Abschnitt 3.8 vorgestellten ortsbezogenen Spieles erhoben werden.

3. In Abschnitt 4.1.2 wird der Begriff der *Semantischen Barriere* eingeführt und gezeigt, wie diese Barrieren zur Lösung des zuvor angesprochenen Anwendungsfalles eingesetzt werden können.

4. Die Berücksichtigung der *semantischen Barrieren* in meinem *SVAC*-Algorithmus, (Abschnitt 4.2), zeigt am Beispiel der Datensammlung der CityExplorer Studien (siehe Kapitel 5 Abschnitt 5.1), wie ein Ansatz aussehen kann, der einen Vorteil aus der komplexen Datenstruktur der erhobenen Daten zieht.

5. Im letzten Abschnitt 4.3 stelle ich als Ausblick eine andere Herangehensweise an die oben genannte Fragestellung vor. Werden beim *SVAC*-Algorithmus durch rein geometrische Operationen die *semantischen Barrieren* berücksichtigt, wird dies beim *RaDAG*-Algorithmus über semantische Regeln erreicht.

4.1. Semantischer Integrationsprozess

In Kapitel 2 Abschnitt 2.4 habe ich die in der Forschung bisher eingesetzten Verfahren vorgestellt, um redundante Daten zu einem neuen qualitativ besseren Datensatz zusammenzuführen. Wie dort kurz angesprochen, können alle diese Ansätze die in Abbildung 4.1 und Abbildung 4.2 gezeigten Anwendungsfälle nicht lösen.

Die folgenden zwei Beispiele illustrieren die Problematik im Detail:

In ersten Beispiel (Abbildung 4.1) werden zwei Tracks dargestellt, die von Spielern eines ortsbezogenen Spieles aufgezeichnet wurden. Einer hat dabei kurzzeitig den Weg auf dem linken Flussufer genutzt (weißer Track), während der andere immer auf dem rechten (oranger Track) geblieben ist. Die Daten für den Fluss (blauer Track) stammen aus einer dritten Quelle - bereitgestellt beispielsweise von einem Kanufahrer. Von beiden Spielern wurde ihre Tracks

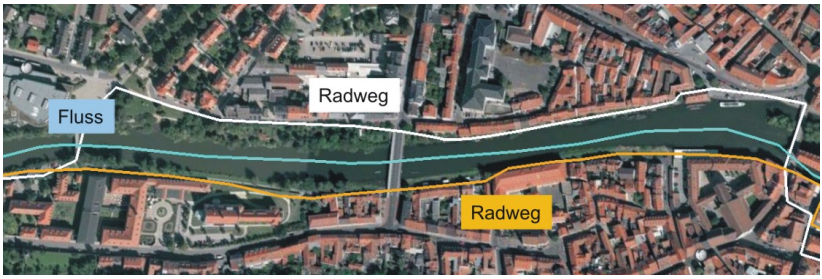


Abbildung 4.1.: Ungelöstes Integrationsproblem im 1-D-dimensionalen Fall, © Karte: Google Earth (`http://earth.google.com/intl/de/`).

korrekterweise als *Radweg* kategorisiert und der blaue Track war schon als *Fluss* semantisch annotiert.

Im zweiten Beispiel (Abbildung 4.2) werden drei, in einem ortsbezogenen Spiel erhobenen, POI dargestellt. Zwei der POI wurden als Restaurants klassifiziert (weißer und oranger POI) und einer als Modegeschäft (blauer POI), wobei dieser geographisch zwischen den zwei Restaurants liegt.

Wenden wir nun ein Verfahren aus Kapitel 2 Abschnitt 2.4 auf diese zwei Beispiele an, bekommen wir folgendes Ergebnis:

Sowohl ein rein geometrischer Ansatz, wie z.B. [168], als auch ein kombiniertes Maß, wie z.B. [186], würde beide Tracks zusammenführen und als einen neuen Datensatz einfügen, da die Ausgangsdatensätze sowohl geometrisch als auch semantisch ähnlich bzw. identisch sind. In beiden Fällen liegt das berechnete Ergebnis der Zusammenführung der Tracks mitten in der Regnitz. Dort kann selbstverständlich kein Fahrradweg verlaufen.

Ein Clusterverfahren für Punktdaten wie z.B. [3], welches auf geometrischer Nähe beruht und die semantische Information nur als eine Art Filter verwendet, würde die zwei Restaurants aus Abbildung 4.2 zu einem neuen Datensatz zusammenführen. Damit würde der neue Clustermittelpunkt aber direkt „auf“ dem POI des Modegschäftes liegen. Dies ist in der Realität natürlich nicht möglich.

Beide gezeigten Szenarien haben unabhängig von der geometrischen Form folgende Eigenschaften gemeinsam, die sie schwer für bisherige Lösungsansätze handhabbar machen:

- Zwei der drei Datensätze sind semantisch identisch oder ähnlich.

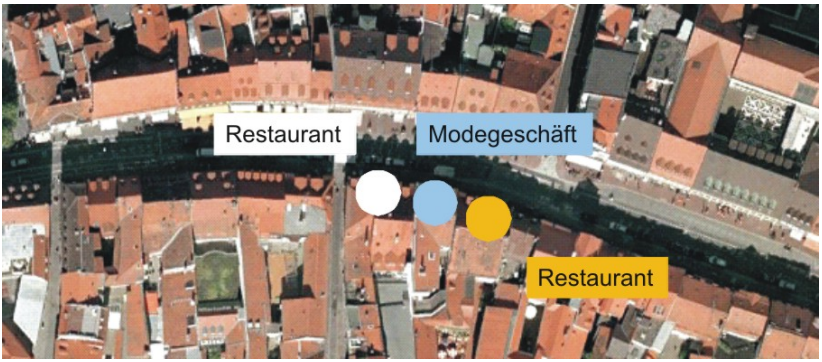


Abbildung 4.2.: Ungelöstes Integrationsproblem im 0-D-dimensionalen Fall, © Karte: Google Earth (`http://earth.google.com/intl/de/`).

- Zwei der drei Datensätze sind geographisch nah.
- Der dritte Datensatz liegt geographisch zwischen den beiden anderen Datensätzen und ist semantisch unähnlich zu diesen.
- Trotz ihrer geographischen und semantischen Nähe bzw. Ähnlichkeit, repräsentieren die zwei Datensätze unterschiedliche Realwelt-Objekte.

Normalerweise können Verfahren zur Lösung des Restaurant-Szenarios auf weniger Information zurückgreifen als zur Lösung des Fluss-Szenarios. Bei 1-D-dimensionalen Daten stehen neben Kennzahlen für die geometrische Nähe Informationen wie Richtung, Anzahl der Überschneidungen oder der Winkel am Schnittpunkt zur Verfügung, die dann bei der Entscheidung, ob Datensätze zusammengefasst werden sollen oder nicht, herangezogen werden können.

Bei 0-D-dimensionalen Datensätzen, wie sie für Clusterverfahren bisher eingesetzt werden (vergleiche beispielsweise [3]), kann nur auf zwei Informationen zurückgegriffen werden: Die geographischen Koordinaten und eine Kategorisierung (siehe hierzu auch die Definition eines *geospatial information objects* nach [25]).

Was den 0-D-dimensionalen Fall für mich noch interessanter macht, ist, das in Abbildung 4.3 abgewandelte Restaurant-Szenario. Wie im obigen Beispiel haben zwei unterschiedliche Spieler (oder derselbe Spieler in derselben oder einer anderen Spielrunde) zwei Datensätze erhoben. Beide wurden in

der identischen semantischen Kategorie *Restaurant* eingeordnet, unterscheiden sich allerdings in den erhobenen geographischen Koordinaten.

Nun stellt sich natürlich die Frage, ob beide Datensätze demselben Realwelt-Objekt entsprechen oder nicht? Diese Frage kann aber mit den zur Verfügung stehenden Informationen (Koordinate und Kategorie) nicht beantwortet werden. Die Abweichung könnte entweder einen Fehler in der verwendeten Lokalisierungstechnologie entsprechen oder eben doch zu zwei unterschiedliche Restaurants in der realen Welt gehören. Für eine endgültige Entscheidung ist immer ein menschlicher Experte nötig. Nur mit Hilfe von zusätzlicher Informationen ist ein automatisches Verfahren in der Lage, diese Unterscheidung eigenständig zu treffen.

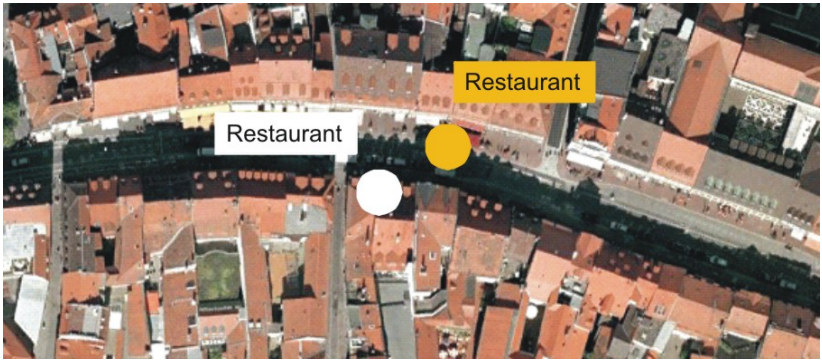


Abbildung 4.3.: Zusätzliches Integrationsproblem im 0-D-dimensionalen Fall, © Karte: Google Earth (`http://earth.google.com/intl/de/`).

Wie im vorangegangenen Kapitel (3) helfen uns ortsbezogene Spiele auch bei der Lösung dieser Problemstellung weiter.

4.1.1. Semantisch annotierte raumbezogene Daten

Ortsbezogene Spiele ermöglichen es mir, durch die Formulierung von speziellen Regeln, genau die Information zu erheben, die ich brauche, um es einem Verfahren zu ermöglichen, automatisch eine Entscheidung für das zweite Restaurant-Szenario zu berechnen.

Welche Information aber ist dies im vorliegenden Anwendungsfall aus Abbildung 4.3? Eine zusätzliche semantische Information in Form einer Be-

zeichnung, d.h. einem Namen, für das Restaurant schließe ich aus. Erstens muss der Name nicht eindeutig sein, es kann durchaus zwei gültige Bezeichner für ein Realwelt-Objekt geben und zweitens ist dieser zeitlich gesehen ebenfalls nicht eindeutig.

Bleibt eine Erweiterung der geographischen Information. Die Regeln von *CityExplorer* (siehe Abschnitt 3.8.1) wurden genau mit diesem Ziel formuliert. Die Spieler nehmen nicht nur die geographischen Koordinaten für den Standort des Objektes auf, sondern durch die Koordinate des aufgenommenen Fotos, auch die Blickrichtung eines Betrachters auf das Objekt selbst. Mit diesem Blickvektor ist es mir möglich ein Verfahren vorzuschlagen, welches berechnen kann, ob die zwei Spieler auf dasselbe Objekt „geschaut“ haben, als sie die Koordinaten dafür erhoben haben. Abbildung 4.4 verdeutlicht diesen Gedanken.



Abbildung 4.4.: Blickvektor als Unterscheidungsmerkmal von geographisch und semantisch ähnlichen POIs, © Karte: Google Maps (`http://maps.google.de/`).

Bevor ich mich aber den zugehörigen Algorithmus im Detail in Abschnitt 4.2 anschauen, will ich folgende formale Definitionen präsentieren, die ich nachfolgend verwenden werde.

In Anlehnung an [25] bezeichne ich die fehlerlose Struktur eines raumbezogenen Datensatzes als ein *object of interest* (OI). Ein OI ist ein Paar $OI \equiv_{\text{def}} \left\langle \overrightarrow{poi}, cat \right\rangle$ aus

poi Der poi oder auch **point of interest** (POI) ist eine 2-D-dimensionale räumliche Koordinate, welche die geographische Position eines POI enthält.

cat Ein Stück semantischer Information. Im Kontext von CityExplorer ist cat genau eine Ausprägung aus einer begrenzten Menge von Kategorien. Der Algorithmus ist aber für komplexere Mengen erweiterbar.

Durch die Aggregation von mehreren Spielsteinen aus *CityExplorer* lässt sich der zugehörige OI berechnen. Formal ist ein Spielstein (Marker oder mrk) ein 4-Tupel $mrk \equiv_{\text{def}} \left\langle \overrightarrow{pov}, \overrightarrow{apl}, va, cat \right\rangle$, bestehend aus:

POV Der *point of view* (POV) („Aufnahmestandort“) ist ein 2-D-dimensionaler geographischer Koordinatenvektor, welcher die Position des Spielers bei der Aufnahme des Fotos des POI angibt.

APL Die *assumed POI location* (APL) („vermutete POI Position“) ist ein 2-D-dimensionaler geographischer Koordinatenvektor, welcher eine Schätzung für die tatsächliche Position des POI ist.

VA Der *view area* (va) („Sichtbereich“) repräsentiert, ausgehend vom APL, die geographische Region, innerhalb der sich die Position des POI befindet. Die Form und Größe des Sichtbereiches kann abhängig von der Lokalisierungsmethode und der zugehörigen Fehlerverteilung gewählt werden.

cat Ein *category identifier* (cat) („Kategorie-ID“) assoziiert eindeutig einen Spielstein mit einer nutzerdefinierten Kategorie.

Die Form des Sichtbereiches wurde für *CityExplorer* als Trapez gewählt. Objekte, die die Spieler im Spiel aufnehmen, sind nicht immer vollständig zugänglich. Die Spielsituation, die links in Abbildung 4.5 dargestellt ist, verdeutlicht die Problematik. Natürliche Hindernisse, wie Zäune, Gräben usw. verhindern, dass Spieler direkt an das Objekt herankommen und den genauen Standort davon erheben. Wegen dieses Unzugänglichkeitsproblemes nehme ich an, dass der APL immer vor der eigentlichen Position des POI liegt. Und selbst wenn ein direkter Zugang möglich ist, wie rechts in Abbildung 4.5 zu sehen, verhindert die eingeschränkte Verfügbarkeit der Lokalisierungstechnologie innerhalb oder in der Nähe bestimmter Objekte, wie z.B. Gebäuden, dass Spieler direkt am Objekt stehen, wenn sie ihren Spielstein setzen (d.h. deren Kategorie angeben).

Die Form des Sichtbereiches sollte diese Diskrepanz zwischen POI und OI überbrücken. Da in *CityExplorer* Fotos gemacht werden, lag es nahe den Winkel des Objektives (α) - falls diese Information zur Verfügung steht - als Maßstab zu nehmen, wie weit sich das Trapez hinter dem APL aufspannt und

damit abzubilden, ob nur ein kleines Detail fotografiert wurde oder ein ganzes Gebäude.



Abbildung 4.5.: Unzugänglichkeit eines Spielortes und rekonstruierte Sichtbereiche zweier Spielsteine, © Karte: Google Maps (`http://maps.google.de/`).

Abbildung 4.6 gibt die Definition des Spielsteines und des OI grafisch wieder. $minDistance$ ist die Entfernung zwischen POV und APL. $maxDistance$ ist ein manuell festgelegter Wert, der angibt, wie groß maximal das Trapez des Sichtbereiches werden kann. Damit soll die durchschnittliche Größe der Objekte im Spielfeld abgedeckt werden. Alternativ, wenn solche Information zur Verfügung stehen, könnte man den Sichtbereich auch je nach Kategorie individuell wählen. So hätten dann Kategorien die als Gebäude zählen, wie Restaurants, einen größeren möglichen Sichtbereich als Kategorien wie Statuen oder auch Straßenschilder.

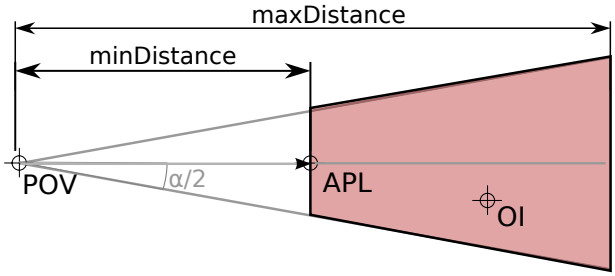


Abbildung 4.6.: Datenstruktur eines CityExplorer Spielsteines

Da $maxDistance$ fest ist, vergrößert sich der Sichtbereich je nach Ent-

fernung von POV und APL. Überschreitet $minDistance$ einen manuell festgelegten Schwellwert bzw. wird $minDistance$ identisch zu $maxDistance$, wird die Form des Sichtbereiches von einem Trapez in ein Rechteck geändert. Dies hat den Hintergrund, dass ich annehme, dass wenn die Entfernung zwischen POV und APL zu groß wird, der Spieler ein Objekt aus weiter Entfernung fotografiert hat. Die Größe des Rechtecks soll wiederum der durchschnittlichen Größe eines Objektes im Spielfeld entsprechen oder abhängig von der Kategorie festgelegt sein. Der APL bildet dabei den Mittelpunkt des Rechtecks.

4.1.2. Semantische Barrieren

Allgemein kann man die Sichtbereiche der Spielsteine (mrk) als speziell geformte Puffer um den OI herum ansehen. Bei der Integration/Aggregation von raumbezogenen Datensätzen sind Puffer eine der Methoden, wie man die geographische Nähe bzw. geometrische Ähnlichkeit, zwischen 2-n Datensätzen berechnen kann (siehe Kapitel 2). Neben dem Blickvektor - der zur Lösung des Problems aus Abbildung 4.3 herangezogen werden kann - ist der Sichtbereich unsere zweite Informationsquelle, die uns hilft das ursprüngliche Restaurant-Szenario aus Abbildung 4.2 aufzulösen.

Abbildung 4.7 zeigt eine abstrakte Darstellung des Restaurant-Szenarios, in dem nur die Sichtbereiche der gesetzten Spielsteine zu sehen sind. $mrk1$ und $mrk2$ stehen für die zwei Restaurants und $mrk3$ für das Modegeschäft in diesem Beispiel. Allgemeiner ausgedrückt ist cat bei $mrk1$ und $mrk2$ identisch und jeweils disjunkt zu $mrk3$. Die Sichtbereiche von allen drei Spielsteinen überlappen sich.

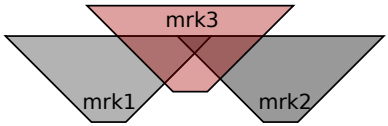


Abbildung 4.7.: Sichtbereiche für die Spielsteine aus dem Restaurant-Szenario: Im Gegensatz zu $mrk3$ besitzen $mrk1$ und $mrk2$ eine identische cat Ausprägung

Was nun verhindert werden soll, ist die Zusammenführung der Spielsteine $mrk1$ und $mrk2$, obwohl sie geographisch nah (überlappende Puffer) und se-

mantisch identisch (gleiche cat Ausprägung) sind. Wir können dies auf zwei Arten erreichen:

1. Wir verändern die geographische Nähe von Datensätzen mit gleicher semantischer Ausprägung anhand von geographisch nahen Datensätzen mit disjunkter semantischer Ausprägung.

2. Wir verändern die semantische Ähnlichkeit von geographisch nahen Datensätzen mit gleicher semantischer Ausprägung anhand geographisch naher Datensätze mit disjunkter semantischer Ausprägungen.

Das erste Vorgehen werde ich im nachfolgenden Abschnitt (4.2) vorstellen und in Kapitel 5 stelle ich meine praktische Evaluierung dazu vor. Die grundlegende Idee ist es, den Datensatz $mrk3$ als Barriere zu verwenden, die verhindert, dass die semantisch ähnlichen Datensätze $mrk1$ und $mrk2$ zusammengefasst werden, deswegen auch der Begriff *Semantische Barriere*. In Abschnitt 4.2.2 gehe ich auf die praktische Umsetzung dieser Idee ein.

Einen ersten Hinweis, wie das zweite Vorgehen umgesetzt werden kann, werde ich formal in Abschnitt 4.3 vorstellen.

4.2. Clustern von semantischen Sichtbereichen

Der *Semantic View Area Clustering*, oder kurz *SVAC*,-Algorithmus macht sich die Idee der *semantischen Barrieren* zu Nutze und ist ein Beispiel für einen Algorithmus, der das Vorgehen aus Abbildung 4.9 realisiert. Obwohl der Algorithmus primär für die raumbezogenen Datensätze entworfen wurde, die durch das *CityExplorer* Spiel erhoben werden, ist das grundsätzliche Vorgehen auch auf andere Datensammlungen - bestehend beispielsweise aus GPS Tracks - anwendbar. Ich werde an den entsprechenden Stellen im Algorithmus auf diesen Punkt eingehen.

Abbildung 4.8 gibt einen Überblick über die Schritte, die in einem allgemein formulierten Integrationsprozess durchlaufen werden (angelehnt an die Arbeit von Walter und Fritsch (1999) [169]). Die einzelnen Schritte werde ich im Hinblick auf den *SVAC*-Algorithmus im Folgenden erläutern.

4.2.1. Vorverarbeitug

Auf Grund der verwendeten Lokalisierungstechnologie bei der Erhebung von raumbezogenen Daten, weisen die Eingangsdaten in den meisten Fällen noch

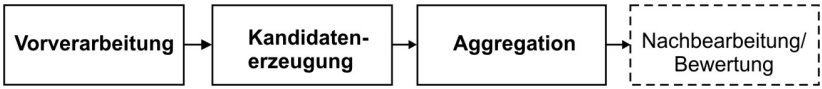


Abbildung 4.8.: Schritte in einem allgemeinem Integrationsprozess

ein sichtbares lokales Rauschen bzw. gegebenenfalls ein globaler Fehler auf. Walter und Fritsch [169] gehen in ihrer Arbeit darauf ein, welche Fehler bei der Digitalisierung von raumbezogenen Daten auftreten können und wie diese in einem Vorverarbeitungsschritt zu behandeln sind. Ziel der Vorverarbeitung ist es hier den globalen Fehler, in den zu integrierenden Datensätzen zu minimieren. Dieser kommt laut [169] hauptsächlich „durch eine ungenaue Transformation zwischen dem Digitalisierungstablett- und dem Modellkoordinatensystem zustande“. Zahlreiche Verfahren wurden unter dem Begriff *Rubber Sheeting* [193] vorgeschlagen, um diesem Fehler zu begegnen. Entsprechende Funktionalität findet sich in Standardsoftware, wie z.B. Esri's ArcGIS [1] Software, welche in diesem Schritt eingesetzt werden kann.

Bei raumbezogenen Daten, die mit mobilen Endgeräten im Feld aufgenommen wurden, treten aber vor allem lokale Fehler auf. Ungenügende bzw. ungünstige Empfangsbedingungen produzieren Ausreißer, Sprünge oder Punktwolken in den aufgezeichneten Tracks, meist durch die GPS Lokalisierungstechnologie. Ausreißer sind Trackpunkte, die von ihren - zeitlich gesehen - benachbarten Trackpunkten geographisch sehr weit entfernt sind. Sprünge kommen zustande, wenn der Empfang für eine bestimmte Zeitdauer komplett ausfällt, dabei aber weiter veraltete Daten aufgezeichnet werden. Punktwolken entstehen durch die Aufzeichnung von Daten für einen einzigen Ort. Bei einem Ausfall des GPS Empfangs sendet der Empfänger kontinuierlich seine letzte bekannte Position und generiert so unter Umständen massive redundante Daten. Ähnlich verhält es sich, wenn sich der Nutzer lange Zeit an am gleichen Ort aufhält. Die aufgezeichneten Trackpunkte sind annähernd normalverteilt mit geringer Abweichungen zueinander, d.h. der Track „springt“ in einem kleinen Bereich zufällig.

Hat man Zugriff auf die in den GPS Rohdaten enthaltenen Genauigkeitsangaben (wie z.B. Anzahl der empfangenen Satelliten oder auch *Dilution of Precision*-Werte [101]) (DOP) lassen sich die oben genannten lokalen Fehler einfach durch das Herausfiltern der Trackpunkte mit schlechten Genauigkeitswerten beheben. Stehen diese Werte nicht zur Verfügung, können fol-

[1] `http://www.esri-germany.de/products/arcgis/`

gende einfache Heuristiken angewendet werden.

Einzelne Ausreißer können z.B. dadurch entfernt werden, dass man alle Trackpunkte entfernt, deren Vorgänger und Nachfolger einen bestimmten Abstandswert überschreiten. Bei Sprüngen im Track sollte dieser in zwei getrennte Tracks aufgeteilt werden. Um Punktwolken zu beseitigen, hilft es den Track zu vereinfachen, d.h. die Anzahl von dicht aufeinander folgende (geographisch gesehen) Trackpunkte zu verringern, z.B. nur jeden dritten oder vierten Punkt zu behalten und alle anderen zu entfernen. Ansonsten können noch unterschiedliche Clusterverfahren [164] anwendet werden, um die Punktwolken auf einen Punkt (zum Beispiel den Schwerpunkt der Cluster) zu reduzieren.

Bestehen die Eingangsdaten lediglich aus Punktdaten, beispielsweise für POI oder die *CityExplorer* Spielsteine - wobei diese genau genommen aus einem Punktpaar bestehen (POV und APL) - vereinfacht sich der Vorverarbeitungsschritt stark. Stehen Genauigkeitswerte zur Verfügung, können schlecht aufgenommene Daten einfach nicht betrachtet werden. Sind keine solche Qualitätsmerkmale vorhanden, kann keine Vorverarbeitung statt finden, da auch keine Kontextinformationen zur Verfügung stehen. In *CityExplorer* stellt daher das Spielelement des *Peer-Reviewings* eine manuelle Vorverarbeitung da. Denn hier werden die Eingangsdatensätze von den Spielern selbst begutachtet und gegebenenfalls ausgebessert oder gleich ganz entfernt.

4.2.2. Aggregation

Sind die Eingangsdaten durch die Vorverarbeitung so gut als möglich von globalen und lokalen Fehlern bereinigt, kann der Aggregationsschritt durchgeführt werden. Dieser gliedert sich in zwei Phasen. In der ersten Phase werden passende Kandidaten für eine spätere Zusammenführung gefunden. Phase zwei führt die Aggregation auf den identifizierten Kandidatenmengen aus.

Betrachtet man 1-D-dimensionale raumbezogene Daten wie sie in GPS Tracks vorkommen - also im Prinzip 1-D-dimensionale Vektordaten -, sind laut [169] vier Möglichkeiten zu unterscheiden, wie Aggregationskandidaten gefunden werden können. Ein Element - diese sind hier Liniensegmente - aus einem Datensatz A kann entweder auf genau ein Element ($1 : 1$ Relation) oder auf n Elemente ($1 : n$ Relation) aus Datensatz B abgebildet werden. Die letztere Relation lässt sich natürlich auch invertieren und sagt aus, dass n Elemente aus Datensatz A auf ein Element aus Datensatz B abgebildet werden ($n : 1$ Relation). Werden n Elemente aus dem einen auf m Elemente des anderen Datensatzes abgebildet, spricht man von einer $n : m$ Relation. Ein

spezieller Fall der $n : m$ Relation ist die $n : m_1 + m_2$ Relation.

Die aufgezählten Relationen lassen sich auf 0-D und 2-D-dimensionale Datensätze übertragen. Ein Element stellt im 0-D-dimensionalen Fall ein POI und im 2-D-dimensionalen Fall ein Polygon dar.

Geometrische Ähnlichkeit

In Kapitel 2 habe ich die verschiedenen Möglichkeiten vorgestellt, wie man die geometrische Ähnlichkeit - die in meinem Anwendungsfall der geographischen Nähe entspricht - berechnen kann. Das gewählte Verfahren soll natürlich die Identifizierung aller im vorangegangenen Abschnitt genannten Relationen unterstützen und zusätzlich die Zusammenführung durch *semantisch Barrieren* getrennter Elemente verhindern.

Beide Bedingungen werden von einem Puffer-basierten Verfahren erfüllt, wie er z.B. in [169] eingeführt wird. Ein solches Verfahren realisiere ich auch in meinem *SVAC*-Algorithmus, wie folgendes Beispiel erläutert:

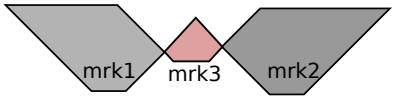


Abbildung 4.9.: Pufferbereiche nach dem Beschneiden der jeweiligen Sichtbereiche

Man kann hierzu das Beispiel aus Abbildung 4.7 logisch weiterführen: Um $mrk1$ und $mrk2$ geographisch zu trennen, wurden deren Sichtbereiche mit dem Sichtbereich von $mrk3$ geschnitten. Dadurch wird nur die Differenz der Sichtbereiche von $mrk1$ in $mrk3$ und $mrk2$ in $mrk3$, sowie $mrk3$ in $mrk1$ und $mrk3$ in $mrk1$, als Puffer in der Berechnung der geographischen Nähe berücksichtigt. Abbildung 4.9 gibt dies grafisch wieder. Entstehen durch die Berechnung der Pufferdifferenz Artefakte, wird nur der Teil des Sichtbereiches weiter betrachtet, der an den APL angrenzt. Der Sichtbereich von $mrk3$ bildet somit eine Barriere zwischen den anderen beiden Spielsteinen und verhindert, dass diese zusammengeführt werden. Da dies aufgrund der Betrachtung der semantischen Information geschieht, nenne ich raumbezogene Datensätze wie $mrk3$: *semantische Barrieren*.

Sind erst einmal die Puffer der geographischen Merkmale aufgebaut, d.h.

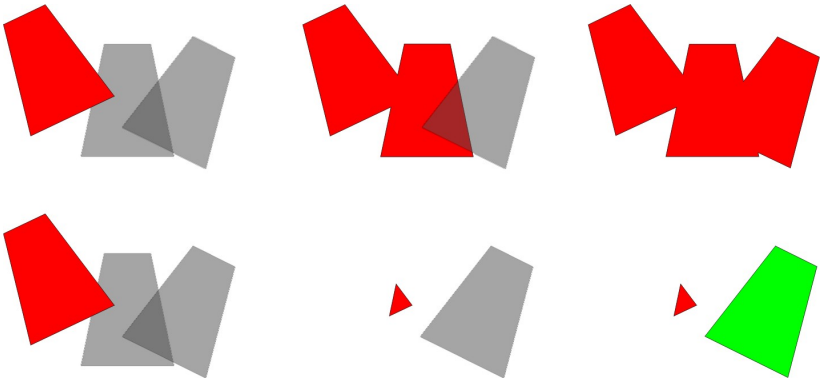


Abbildung 4.10.: Möglichkeiten der Clusterbildung aufgrund der geometrischen Nähe

die Sichtbereiche der Spielsteine mit denen der auftretenden *semantischen Barrieren* beschnitten, kann der Aggregationsschritt erfolgen. In diesem sollen dann alle geometrisch nahen und semantisch ähnlichen Objekte zusammengeführt werden.

Die geometrische Nähe wird anhand der Überlappung der Sichtbereiche festgestellt. Ein Cluster besteht initial aus einem zufällig ausgewählten Spielstein mrk. Alle möglichen Kandidaten (diejenigen mrk, dessen semantische Ähnlichkeit einen festgelegten Schwellwert nicht übersteigen) werden daraufhin überprüft, ob sie den gemeinsamen Sichtbereich des Clusters schneiden. Der gemeinsame Sichtbereich eines Clusters kann auf zwei verschiedene Arten definiert werden, die in Abbildung 4.10 dargestellt sind:

1. Der Sichtbereich des Clusters wird immer um die Sichtbereiche der geometrisch ähnlichen mrk erweitert, d.h. der Sichtbereich des Clusters entspricht der Vereinigung aller in ihm enthaltenen mrk. Das obere Beispiel in Abbildung 4.10 illustriert das Vorgehen. Am Anfang besteht der Sichtbereich des Clusters nur aus dem Sichtbereich des linken mrk und wächst dann in jedem Schritt sukzessive an.

2. Der Sichtbereich des Clusters entspricht der Schnittmenge der Sichtbereiche der im Cluster zusammengeführten mrk. Im unteren Beispiel in Abbildung 4.10 beginnen wir wieder mit dem Sichtbereich des linken mrk. Wird das Cluster um den mittleren mrk erweitert, besteht der Sichtbereich des resultierenden Clusters aus der Schnittmenge der

Sichtbereiche der zwei mrk. Der Sichtbereich des Clusters wird immer durch die Schnittmenge der ersten beiden im Cluster enthaltenen mrk definiert. Da in diesem Fall der Sichtbereich des rechten mrk keine Überlappung mit dem Sichtbereich des Clusters aufweist, bildet dieser ein neues Cluster.

Dabei unterscheidet der Algorithmus bisher noch keine verschiedenen Grade der Überlappung, sondern verwendet einen binären Operator zur Clusterbildung, d.h. ein neues Cluster c wird immer dann erweitert, wenn der Schnitt des Sichtbereiches des Clusters s_c und der eines semantisch ähnlichen mrk, s_m, nicht leer ist.

In Kapitel 5 stelle ich eine erste Evaluierung vor, in der unter anderem die zwei Ansätze zur Bestimmung des Sichtbereiches eines Clusters gegenübergestellt werden. An dieser Stelle möchte ich aber noch auf zwei Überlegungen hinweisen, warum meiner Meinung nicht ohne weiteres ein Ansatz dem anderen vorgezogen werden kann.

In Abbildung 4.11 werden zwei Konstellationen von Sichtbereichen dargestellt, die jeweils eine Methode begünstigen. Die erste Anordnung von Sichtbereichen illustriert den Fall, dass ein oder mehrere Spieler dasselbe Objekt, z.B. ein großes Gebäude, mehrmals oder jeweils einmal aus unterschiedlichen Richtungen kommend aufgenommen haben. Diese mrk können nur durch eine Vereinigung der jeweiligen Sichtbereiche bei der Clusterbildung zu einem einzelnen Cluster zusammengefasst werden. Wird die Schnittoperation angewendet, werden drei einzelne Cluster gebildet, egal in welcher Reihenfolge die mrk abgearbeitet werden.

Bei der zweiten Anordnung soll genau das Gegenteil erreicht werden. Der/-die Spieler hat/haben zwei unterschiedliche Objekte derselben oder zumindest mit hinreichend ähnlicher, semantischer Ausprägung aufgenommen. Dabei wurde das linke Objekt dreimal markiert und das rechte zweimal. Das Problem hierbei ist, dass beide Objekte geometrisch so nahe beieinander liegen, dass sich die Sichtbereiche der mrk teilweise überlappen. Hier würde die Vereinigungsoperation alle mrk in ein Cluster aggregieren, wohingegen die Schnittoperation zwei getrennte Cluster produziert. Auch muss bei der Verwendung der Vereinigungsoperation das vom Single-Linkage Clusterverfahren bekannte Phänomen des *Chaining* [194] beachtet werden, welches bei der Schnittoperation nicht auftritt.

Wird bei dieser Konstellation die Schnittoperation angewendet, spielt allerdings die Reihenfolge der Abarbeitung der mrk eine Rolle. Werden in diesem Fall die zwei mittleren mrk als erstes zusammengeführt, erhält man statt

den gewünschten zwei Cluster, drei. Dabei muss man anmerken, dass nur die Schnittoperation Reihenfolgeabhängig ist. Dieses Problem kann durch mehrmaliges Durchführen des *SVAC*-Algorithmus mit unterschiedlichen Start-mrk abgeschwächt werden.

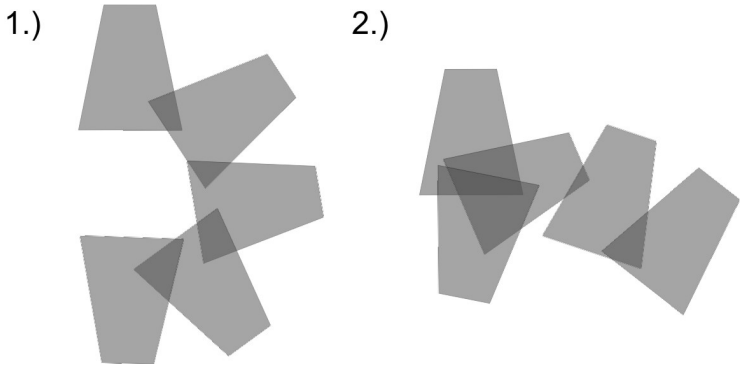


Abbildung 4.11.: Günstige Konstellationen für Vereinigung bzw. Schnitt als Methode zur Erstellung des Clustersichtbereiches s_c

Semantische Ähnlichkeit

Voraussetzung für eine Zusammenführung von mrk zu Clustern anhand von überlappenden Sichtbereichen ist erstens die Behandlung von *semantischen Barrieren* und zweitens die Feststellung der semantischen Ähnlichkeit. In Kapitel 2 habe ich einige Modelle zur Berechnung der semantischen Ähnlichkeit diskutiert. Des Weiteren stellen Budanitsky und Hirst (2001) [195] in ihrer Arbeit fünf verschiedene Ähnlichkeitsmaße vor, die speziell bei der Verwendung von WordNet (siehe [174] und [175]) in der Literatur Anwendung finden. Außer dem Maß nach Hirst and St-Onge [196] erbrachten die vier Maße nach Jiang und Conrath [197], Lin [198], Leacock und Chodorow [199] und Resnik [8] ähnlich gute Ergebnisse.

Für die vorliegende Implementierung des *SVAC*-Algorithmus habe ich mich für das Maß nach Resnik [8] entschieden. Prinzipiell lassen sich aber alle genannten Maße ohne Änderung im Algorithmus verwenden. Die Berechnung der Ähnlichkeiten zwischen den verwendeten Kategorien cat, die beim Setzen der mrk von den Spielern ausgewählt werden, geschieht unabhängig vom Rest des Algorithmus. Das Ergebnis ist eine Lookup-Tabelle, in dem der Algorithmus bei der Bearbeitung der *semantischen Barrieren* die entsprechen-

den Ähnlichkeitswerte überprüft.

Nach Resnik (1995) [8] basiert die Ähnlichkeit zwischen zwei Konzepten auf dem *Informationsgehalt* (information content), den ein übergeordnetes, beide subsumierendes und damit allgemeineres, Konzept enthält. Grundlage der Berechnung ist eine IS-A Taxonomie von Konzepten. Die Idee ist dabei ähnlich wie, wenn wir die verbindenden Kanten zählen würden, die von den beiden Konzepten zu den übergeordneten Konzept führen. Je weniger Kanten ich zwischen zwei Konzepten gehen muss, d.h. je kürzer der Weg zwischen beiden in der Taxonomie ist, desto ähnlicher sind sie.

Um nun das Problem zu umgehen, dass alle Kanten zwischen Konzepten die gleiche „Länge“ haben, versieht Resnik Konzepte in der Taxonomie mit Wahrscheinlichkeiten. Er führt dazu eine Funktion $p : C \rightarrow [0, 1]$ ein, wobei C die Menge aller Konzepte in der Taxonomie ist, die angibt, wie wahrscheinlich es ist, eine Instanz eines Konzeptes c in der Taxonomie vorzufinden. p ist dabei monoton steigend, wenn man in der Taxonomie aufsteigt, und 1 für das oberste Konzept.

Quantifiziert wird der Informationsgehalt, laut [8] aufbauend auf der Informationstheorie nach [200], eines Konzeptes c mit der negativen logarithmischen Wahrscheinlichkeit, $-log\ p(c)$. Daraus ergibt sich, dass der Informationsgehalt monoton sinkt, wenn man in der Taxonomie aufsteigt und der Informationsgehalt des obersten Konzeptes ist 0 (falls eindeutig), d.h. je abstrakter das Konzept, desto weniger Informationsgehalt besitzt es.

Um nun den Informationsgehalt desjenigen Konzeptes zu berechnen, welches die beiden Konzepte in einer Taxonomie zusammenfasst, definiert Resnik folgende Gleichung:

$$sim(c_1, c_2) = max_{c\ \in\ S(c_1,c_2)}[-log\ p(c)] \tag{4.1}$$

$S(c_1, c_2)$ ist eine Menge von Konzepten, die c_1 und c_2 zusammenfassen. Normalerweise findet sich immer ein eindeutiges Konzept mit maximalem Informationsgehalt, aber im allgemeinen Fall muss dies nicht der Fall sein. Wie von Resnik angesprochen, muss in der Praxis, wie auch in meinem Fall, nicht die Konzeptähnlichkeit untersucht werden, sondern die Wortähnlichkeit, d.h. der Name der Kategorie. Ein Wort kann verschiedene Bedeutungen haben und deshalb auch in unterschiedlichen Konzepten auftreten. Daher muss jede mögliche Kombination an Konzepten für ein Wort berechnet werden.

$$wsim(w_1, w_2) = max_{c_1,c_2}[sim(c_1, c_2)] \tag{4.2}$$

s_w stellt die Menge der Konzepte dar, die in der Taxanomie die Bedeutung

des Wortes w annehmen, wobei c_1 die Menge $s(w_1)$ durchläuft und c_2 die Menge $s(w_2)$.

Will man nun den Wert von $p(c)$ bestimmen, berechnet man im Falle von WordNet ([174] und [175]) zuerst die Häufigkeit, mit der ein bestimmtes Nomen in dem verwendeten Korpus vorkommt, mit der Formel:

$$freq(c) = \sum_{n \in words(c)} count(n) \tag{4.3}$$

wobei $words(c)$ die Menge der Wörter ist, die vom Konzept c zusammengefasst werden. Die Konzeptwahrscheinlichkeit ergibt sich dann nach:

$$\hat{p}(c) = \frac{freq(c)}{N} \tag{4.4}$$

wobei N die Gesamtzahl von Nomen im Korpus darstellt.

Wir werden in Kapitel 5 sehen, wie die semantische Ähnlichkeitsberechnung nach Resnik auf unser Anwendungsbeispiel übertragen werden kann. Im folgenden zeige ich, wie der Algorithmus in seiner Gesamtheit abläuft.

SVAC-Algorithmus

Wie oben erwähnt beginnt der Algorithmus mit der Berechnung der semantischen Ähnlichkeit zwischen den in *CityExplorer* von den Spielern gewählten Kategorien. Die resultierenden Ähnlichkeitswerte dienen als Grundlage für die in Abbildung 4.2.2 in Pseudocode dargestellten Behandlung der semantischen Barrieren.

```
CROPVIEWAREAS(Markers, threshold)
1  for ⟨m1, m2⟩ ∈ Markers × Markers
2      do if m1 ≠ m2 and If sim_C(m1, m2) < threshold
3          ▷ Crop marker view area, unless semantically similar
4            then m1.VA ← m1.VA − m2.VA
```

Abbildung 4.12.:

.

In dieser ersten Phase des Algorithmus wird jeder Sichtbereich, der sich mit einem anderen Sichtbereich eines semantisch unähnlichen Spielsteines,

mrk, überschneidet, beschnitten (siehe Zeile 2 und 3 in Abbildung 4.2.2). Ausgangspunkt ist ein zufällig ausgewählter mrk_1. Alle mrk_n, die dessen Sichtbereich überschneiden, werden auf semantische Ähnlichkeit überprüft, d.h. es wird ein lookup des entsprechenden Wertes in der zuvor aufgestellten Ähnlichkeitstabelle durchgeführt. Liegt dieser Wert unter einem vorher festgelegten Schwellwert, wird der Sichtbereich von mrk_1 beschnitten. Werden keine weiteren überschneidenden mrk_n mehr gefunden, wird mit dem nächsten Spielstein fortgefahren.

SVAC($Markers$)

```
1  O ← Markers                          ▷ Open/unprocessed marker list
2  C ← ∅                                ▷ Cluster list
3  while O ≠ ∅
     ▷ Create a new cluster with a marker from the open list
     ▷ Combine currentArea with the view area of the new marker
4      do m1 ← PICKANY(O); O ← O \ {m1}
5         currentArea ← GETVIEWAREA(m1)
6         c ← NEWCLUSTER(m1)
7         m2 ← FINDOVERLAPPING(currentArea, O, m1.cat)
            ▷ Find all markers with overlap in the same category.
8         while m2 ≠ NIL
9            do O ← O \ {m2}
10              currentArea ← COMBINEVIEWAREAS(
11                GETVIEWAREA(currentArea, m2))
12              c ← c ∪ {m2}
13              m2 ← FINDOVERLAPPING(currentArea, O, m1.cat)
14        C ← C ∪ {c}
15 return C
```

Abbildung 4.13.: Simple View Area Clustering

Innerhalb der anschließenden Aggregationsphase des *SVAC*-Algorithmus werden, wie ebenfalls oben schon beschrieben, die mrk anhand derer Sichtbereiche zu entsprechenden Clustern zusammengeführt. In Zeile 12 im Pseudocode von Abbildung 4.2.2 wird die Vereinigungsoperation verwendet, statt dessen kann entsprechend auch die Schnittoperation angewendet werden.

Um nun aus den Clustern der mrk den entsprechenden OI zu erhalten, kann zum Beispiel der Schwerpunkt des Koordinatenraums, der von den APL Koordinatenvektoren der mrk aufgespannt wird, berechnet werden. Als semantische Information cat wird am Besten die Vereinigungsmenge aller cat Ausprägungen der mrk verwendet.

Bewertung/Nachbearbeitung

Optional kann in einem allgemeinen Integrationsprozess noch eine abschließende Bewertungs- bzw. manuelle Nachbearbeitungsphase angeschlossen werden. Da hierfür allerdings in der Literatur noch keine geeigneten Verfahren bzw. Kennzahlen zur Bestimmung der Güte des Integrationsergebnisses zu finden sind, gehe ich an dieser Stelle nur auf die allgemeine Idee dahinter ein.

In der Bewertungsphase sollen Kennzahlen helfen das Ergebnis der Integration zu bewerten. Aus nicht-geographischen Forschungsfeldern sind zahlreiche Ansätze zur Berechnung solcher Kenngrößen bekannt, siehe beispielsweise [201], [202], [203], [204] oder auch [205]. Diese können zwar auch auf raumbezogene Datensätze angewendet werden, sind aber in diese Zusammenhang nicht ausreichend. Das Problem ist die Unschärfe bei raumbezogenen, bzw. geographischen Datensätzen. Anders als beispielsweise bei Addressdatensätzen oder ähnliche in einem betriebswirtschaftlichen Kontext auftretenden Daten, kann bei geographischen Daten nicht so einfach zwischen korrekten und falschen Daten unterschieden werden. Eine Person wohnt entweder in einer bestimmten Adresse oder eben nicht. Die Koordinaten eines bestimmten Restaurants sind dagegen nicht eindeutig festgelegt. Legt man diese an den (Haupt-)Eingang des Gebäudes oder doch besser „mitten ins“ Gebäude? Noch schwieriger wird es bei flächigen Phänomenen wie Regionen oder größeren Gebäudekomplexen, z.B. Fabrikanlagen.

Natürlich verringert sich das Problem der Unschärfe etwas, wenn als Repräsentation der geographischen Merkmale, neben Punktkoordinaten auch Mengen von Punktkoordinaten (Linienzüge oder Polygone) genutzt werden können. Leider unterstützen die meisten Systeme, die geographische Daten verwenden, wie z.B. Navigationssysteme, diese Art der Darstellung nicht. Oder die Festlegung eines Merkmals auf diese Art und Weise bleibt trotzdem unscharf, wie es bei eher abstrakten Regionen, z.B. Stadtteilen, der Fall ist.

Zwar gibt es auch einen offiziellen Standard, welcher Maße zur Beschreibung der Qualität von geographischen Daten vorgibt, - den ISO 19113 (siehe [206] oder direkt [207]) Standard - aber auch dort wird die gerade angesprochene Problematik nicht aufgelöst.

Eine detaillierte Behandlung würde an dieser Stelle den Rahmen der vorliegenden Arbeit sprengen, der interessierte Leser sei daher auf folgende Arbeiten zu diesem Thema verwiesen: [208], [209], [210], [211], [202], [205], [212], [201], [204], [203], [213] und [214]).

Aus meiner Sicht ist das wichtigste Kriterium für Datenqualität im Zusammenhang mit geographischen Daten immer noch die Fitness („fitness for use“) der Daten für den Nutzer. Kann diese eher abstrakte Kennzahl bei Ansätzen zur Datenerhebung durch Experten erst im Nachhinein bestimmt werden, ist dies bei einer gemeinschaftlichen Erhebung der Daten durch geowissenschaftliche Laien schon vor der ersten Nutzung der Daten durch einen Nutzer - der in den meisten Fällen eben auch ein geowissenschaftlicher Laie ist - möglich.

Ein einfaches Maß für die Qualität eines gemeinschaftlich erstellten geographischen Datensatzes wäre z.B. die Anzahl der Nutzer, aus deren Einzelerhebungen der Datensatz erstellt wurde - durch beispielsweise der zuvor vorgestellten Aggregationsmethode. Natürlich lassen sich neben diesem, im Sinne des „Wisdom of the Crowd“ [23] zu verstehenden Maßes, auch noch komplexere Maße definieren. Ein meiner Meinung nach viel versprechender Ansatz wäre es auf Grundlage von Trust-Werten [191] entsprechende Gütekriterien zu definieren.

4.3. Ausblick: Regelbasierte Datenaggregation

Wie in Abschnitt 4.1.2 erwähnt, gibt es neben einer geographischen Herangehensweise an das Phänomen der *semantischen Barriere*, durch das Beschneiden der Puffer-Regionen der raumbezogenen Datensätze (z.B. von GPS Tracks oder POIs), noch eine semantische. Es wird in diesem Ansatz durch die Aufstellung von semantischen Regeln, in Form von beschreibungslogischen Ausdrücken, auf die zu aggregierende Menge an Kandidaten geschlossen. Die so gefundenen Kandidaten werden dann mit Hilfe von geographischen Aggregationsalgorithmen zusammengefasst. Der im nachfolgenden vorgestellte Ansatz beruht auf Ausführungen aus meiner Arbeit in [215].

Um die semantischen Aggregationsregeln formulieren zu können, schlage ich eine einfache relationale Sprache vor, die räumliche SQL Relationen [216] mit Konstrukten aus formalen Ontologien erweitert. Dies erlaubt mir Typinformationen in den Argumenten der Relationen zu formulieren. Ein relationaler Ausdruck macht dabei eine Aussage über eine räumliche oder se-

mantische Relation, welche zwischen räumlichen Objekten eines bestimmten Typus gilt. Formal ist ein relationaler Ausdruck definiert als ein räumliches oder semantisches Relationensymbol und eine Sequenz von typisierten Argumenten.

Die vorgeschlagene Regelsprache erlaubt das Modellieren von sowohl Entfernungsrelationen (z.B. *small_distance*) als auch von gerichteten Relationen, die Ordnungsinformationen enthalten (z.B. *between*). Daneben stehen die traditionellen räumlichen SQL Relationen [216] ebenfalls zur Verfügung. (z.B. *intersects*). Geometrische und semantische Relationen werden qualitativ definiert und nicht mit Hilfe von harten metrischen Einheiten. Negationen von Regeln sind ebenfalls möglich (z.B. when not ... *between* ...).

Ein Satz von formulierten Regeln legt nicht nur die Kandidaten für die nach Abschnitt 4.2 nachfolgende Aggregation fest, sondern kann auch als Menge von Vorbedingung für einen bestimmten Aggregationsalgorithmus angesehen werden. So können für unterschiedliche Algorithmen verschiedene Mengen von Regelsätzen formuliert werden.

Für den Anwendungsfall aus Abbildung 4.1 könnte man beispielsweise nachfolgende Regelsätze vorsehen. Diese verhindern dann bei der Verwendung des Aggregationsalgorithmus nach Morris et al. (2004) [7] die Zusammenfassung der zwei Tracks der Spieler (weiß und orange) mit dem blauen Track, der den Fluss repräsentiert. Durch die Relation *between* wird die geometrische Nähe, bzw. semantische Ähnlichkeit, durch das semantische Schließen entsprechend verändert, um der semantischen Barriere Rechnung zu tragen.

Regeln (Vorbedingungen):

```
when track_1:BikingTrail
     intersects
     trail_2:BikingTrail

when track_1:BikingTrail
     small_distance
     track_2:BikingTrail

when not track_1:BikingTrail
         between
         Regnitz_river: WaterBody
```

```
        track_2:BikingTrail

when Regnitz_river_tour: InformationClass
     sufficient_similar
     Regnitz_river_tour: InformationClass
```

Alle Datensätze, die diesen Regelsatz erfüllen, werden als Kandidaten für eine nachfolgende Zusammenfassung mit dem vorher ausgewählten Aggregationsverfahren festgelegt. Wie man hier sehen kann, werden die zwei Tracks der Spieler nicht in diese Kandidatenliste aufgenommen, da die dritte Regel (*...between...*) nicht erfüllt wird. Zwischen den beiden Tracks befindet sich ja der Track für den Fluss.

Die typisierten Argumente für die Regeln sind in einer zugehörigen Ontologie vorgegeben. Aus dieser können weiterhin T-Box Aussagen abgeleitet werden. Zusammen mit dem spezifischen Regelsatz ist eine Formulierung von A-Box Aussagen möglich. Für den obigen Regelsatz sehe die T-, bzw. A-Box folgendermaßen aus:

T-Box:

$BikingTrail \subseteq Concept$
$WaterBody \subseteq Concept$
$InformationClass \subseteq Concept$

A-Box:

$BikingTrail(track_1)$
$BikingTrail(track_2)$
$WaterBody(Regnitz_river)$
$InformationClass(Regnitz_river_tour)$
$InformationClass(Regnitz_river_tour)$

Der Vorteil bei der Verwendung einer Beschreibungslogik als Regelsprache ist die Möglichkeit eine A- und T-Box Konsistenzüberprüfung durchzuführen. Falls die Regeln konsistent zur A- und T-Box sind, können diese dann auf den vorhandenen Datenbestand angewendet werden bzw. durch logisches Schließen entsprechende Schlussfolgerungen gezogen werden - d.h. zu aggregierende Datensätze identifiziert werden.

Um diesen theoretischen Ansatz in der Praxis anzuwenden, sind noch folgende Problemstellungen zu beantworten:

- Welche der zahlreichen DL-Sprachen ist am Besten geeignet, bzw. ausdrucksmächtig genug, um die oben genannten Sprachkonstrukte - vor allem die aufgestellten Regeln - formulieren zu können.
- Ist diese dann noch entscheidbar?
- Falls ja, gibt es dafür auch einen praktisch einsetzbaren Reasoner?
- Wie wird mit konfliktionären Regeln umgegangen?

4.4. Zusammenfassung

In diesem Kapitel habe ich im ersten Teil eine formale Definition der Datensätze gegeben, die während des *CityExplorer* Spiels - aus Kapitel 3 - erhoben werden. Diese Formalisierung erlaubt es mir den Begriff der *semantischen Barriere* einzuführen. Diese geometrischen bzw. geographischen Merkmale nehmen bei der Zusammenführung von raumbezogenen Daten eine besondere Rolle ein und wurden in der Literatur bisher noch nicht betrachtet.

Das theoretische Konstrukt der *semantischen Barriere* berücksichtige ich in einem praktisch umgesetzten Algorithmus zur Aggregation von raumbezogenen Datensätzen. Der *SVAC*-Algorithmus ist zwar bisher auf die Definition der Datensätze aus Abschnitt 4.1.1 angepasst, kann aber dank der Verwendung eines Puffer-basierten Ansatzes zur Bestimmung der geometrischen Nähe, auch auf andere Datensätze (1-D-dimensionale, wie in [215], oder 2-D-dimensionale) angepasst werden.

Die semantische Ähnlichkeit wird mit Hilfe des Verfahrens nach Resnik [8] berechnet. Die Behandlung der *semantischen Barrieren* erfolgt durch die Beschneidung der Puffer der zu aggregierenden geographischen Merkmale. Dadurch wird die fälschlicherweise vorhandene geometrische Nähe zweier semantisch ähnlicher Merkmale unterbunden, die durch ein semantisch unähnliches aber geographisch nahes Merkmal getrennt sind.

Ein Ausblick auf ein auf regel-basiertes Verfahren schließt das Kapitel ab. Im *RaDAG*-Ansatz schlage ich vor durch beschreibungslogisches Schließen von semantischen Regeln die Kandidaten für eine nachfolgende Aggregation zu erzeugen. Spezielle Regeln ermöglichen die Behandlung der *semantischen Barrieren*.

5. Evaluation und Diskussion

In den vorangegangenen zwei Kapiteln habe ich aufgezeigt, mit welchen Lösungsansätzen ich die zwei von mir aufgestellten Forschungsfragen bearbeitet habe. Durch die Verwendung von ortsbezogenen Spielen wird das Motivationsproblem angegangen und die Struktur der erhobenen Daten ermöglicht die Formulierung eines neuartigen Ansatzes für die Aggregationsphase des semantischen Integrationsprozesses. Durch die Zusammenführung der raumbezogenen Datensammlung werden Fehler in der Erhebung durch die geowissenschaftliche Laien, die nicht während des Spiels ausgebessert wurden, behoben bzw. verringert.

Die im nachfolgenden vorgestellten Studien dienen dazu, diese zwei Lösungsansätze zu evaluieren. Dazu habe ich folgende Versuchsreihen durchgeführt:

1. Das in Kapitel 3 Abschnitt 3.8 vorgestellte ortsbezogene Spiel *CityExplorer* wurde in drei Nutzerstudien unter anderem in Bezug auf den Spielspaß, die Möglichkeit der Einbindung in den normalen Tagesablauf der Spieler sowie die Quantität und Qualität der erhobenen Daten untersucht.

2. Eine zusätzliche Studie zu *CityExplorer* hat die Machbarkeit eines verteilten Spieles getestet. Spieler von der Universität Bamberg (Deutschland) traten hierbei gegen Studenten aus der Keio University in Yokohama (Japan) an.

3. In einer eigenen Studienreihe, bestehend aus zwei Studien mit dem dafür eigens entwickelten ortsbezogenen Spiel *GeoSnake*, wurde das *Wisdom of the Crowd*-Entwurfsmuster zur Sicherung der Qualität der so spielerisch erhobenen raumbezogenen Datensätze evaluiert.

4. Eine letzte Studie habe ich im Zusammenhang mit dem *SVAC*-Algorithmus durchgeführt. Ziel war es den Ansatz mit einem in der Literatur weit verbreiteten Verfahren zur Aggregation von 0-D-dimensionaler raumbezogener Daten zu vergleichen.

5.1. Studien zur Erhebung raumbezogener Daten mit ortsbezogenen Spielen: CityExplorer

Bei allen *CityExplorer* Studien kamen vier Nokia 6630 Mobiltelefone und mittels Bluetooth verbundene GPS Mäuse (XAIOX Wonde-XL) zum Einsatz. Auf den Telefonen wurde die J2ME Variante des Spieles verwendet. Die Windows Mobile Version von *CityExplorer* (siehe Abschnitt 3.8.3 in Kapitel 3) wurde aus Mangel an GPS Mäusen nicht verwendet.

5.1.1. Nutzerstudien zum nicht-verteilten Spiel

Alle drei Studien zum nicht-verteilten Spiel, d.h. es wurde auf einem einzigen Spielfeld gespielt, wurden in der Innenstadt von Bamberg durchgeführt. Jedes der Spiele folgte demselben Ablauf:

1. Das Spiel wurde mit einer Einführung in die Spielregeln und der Hardware eingeleitet (siehe Abbildung 5.1). Danach erhielten die Spieler die Telefone, GPS Geräte und eine CD mit der nötigen Software zum Verbinden der Telefone mit dem eigenen PC. Dies war zum Hochladen der während des Spieles gesetzten virtuellen Spielsteine (d.h. gemachte Fotos und die zugehörige XML-Datei, siehe Abschnitt 3.8.3) nötig.

2. Jedes Spiel lief über einen längeren Zeitraum von mehreren Tagen, in denen die Spieler die Hardware und das Spiel unüberwacht verwendeten. In diesem Zeitraum wurde die Kommunikation zwischen den Spielern über die CityExplorer Webseite, E-Mail oder zufällige Treffen beim Verweilen innerhalb des Spielfeldes aufrechterhalten.

3. Für jedes Spiel wurde ein vorher unbekannter Sachpreis für den Sieger in Aussicht gestellt.

Zur Evaluierung der Spielaspekte von CityExplorer wurden folgende Methoden eingesetzt:

- Ein Fragebogen zur Erfassung des Vorwissens der Spieler in Bezug auf demographische Daten, Spiele und GPS Technologie, der vor dem Spiel ausgefüllt wird, siehe Anhang A.1.1.



Abbildung 5.1.: Spieleinführung und Hardwareausgabe vor Beginn eines CityExplorer Spieles

- Im Spiel selbst haben die Spieler über persönliche Blogs die Möglichkeit ihre eigenen Erfahrungen mit *CityExplorer* festzuhalten und über einen gemeinsamen Spiel-Blog das Spielgeschehen für alle sichtbar zu kommentieren.

- Nach einem Spiel erfasst ein weiterer Fragebogen die Spieleindrücke und die Bewertung des Spiels durch die Spieler, siehe Anhang A.1.2. Vor allem die Werte für die Fragen 1 und 2 sollen Aufschluss darüber geben, wie die Integration der Datenerhebung in den Spielablauf den Spielspaß beeinflusst hat. Darüber hinaus gibt die Quantität der erhobenen Daten ein Indiz darauf, ob das Spiel auch auf längere Sicht „unterhaltsam“ (motivierend) für die Spieler war.

- Ein unstrukturiertes Interview wird mit allen Spielern gemeinsam am Ende jedes Spieles durchgeführt.

Beim Fragebogendesign habe ich auf den in den *Geogame* Nutzerstudien [6] verwendeten Fragebogen zurückgegriffen und um *CityExplorer* spezifische Fragen erweitert. In der Literatur hat sich bisher noch kein standardisiertes Design durchgesetzt - siehe beispielsweise [121] und [147] -, obwohl meiner Meinung nach die Arbeit von Jegers (2007) [108] ein gute Grundlage dafür bieten würde.

Abbildung 5.2 zeigt das Spielfeld, welches ich für die Spiele in Bamberg verwendet habe. Es erstreckt sich über die gesamte Innenstadt, sowie Gebiete

der Bergstadt von Bamberg. Die Aufteilung in 3x3 Spielfeldsegmente wurde hier aufgrund der Größe der Ausdehnung des Spielfeldes gewählt, damit die Segmente nicht zu klein werden. Die Spieler sollten in jedem Segment ausreichend Gelegenheit haben, entsprechende Merkmale der im Vorhinein festgelegten Kategorien zu finden. Die von den Spielern selbst frei gewählten Kategorien zeigt Abbildung 5.2, links für das erste, in der Mitte für das zweite und die rechte Spalte entspricht den Kategorien für das dritte Spiel.



Abbildung 5.2.: Spielfeld für alle Studien zu CityExplorer in der Bamberger Innenstadt, sowie die von den Spielern gewählte Kategorien für diese Spiele, © Karte: Google Maps (`http://maps.google.de/`).

Nutzerstudie 1 - Bamberg City

An der ersten Studie nahmen drei wissenschaftliche Mitarbeiter des Lehrstuhls der Kulturinformatik der Universität Bamberg und ein Informatikstudent (Master), ebenfalls von der Universität Bamberg, teil. Die Spieldauer wurde auf sieben Tage festgelegt. Der Altersdurchschnitt lag bei 27 Jahren und alle Teilnehmer waren männlich. Es wurden insgesamt 276 Spielsteine während des Spieles gesetzt. Abbildung 5.3 gibt eine Übersicht über das Endergebnis dieses Spieles.

Nutzerstudie 2 - Bamberg City Reloaded

An der zweiten Studie in Bamberg nahmen vier Informatikstudenten (Bachelor) und zwei Kulturwissenschaftler (Diplom) teil. Das durchschnittliche

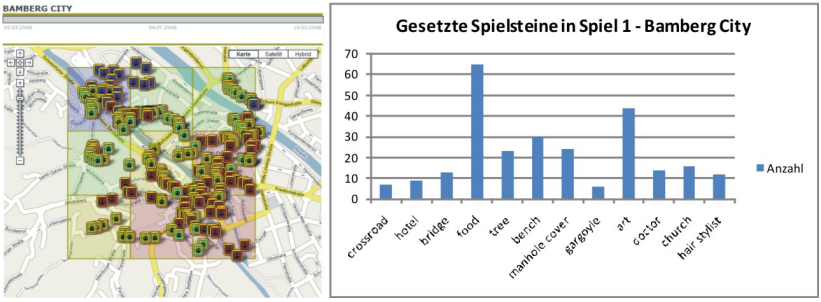


Abbildung 5.3.: Endzustand des ersten CityExplorer Spieles in Bamberg, © Karte: Google Maps (`http://maps.google.de/`).

Alter lag bei 27,5 Jahren, bei vier männlichen und zwei weiblichen Teilnehmern. Insgesamt wurden in neun Tagen Spielzeit 308 Spielsteine generiert, Abbildung 5.4 gibt eine Übersicht über das Endergebnis dieses Spieles.

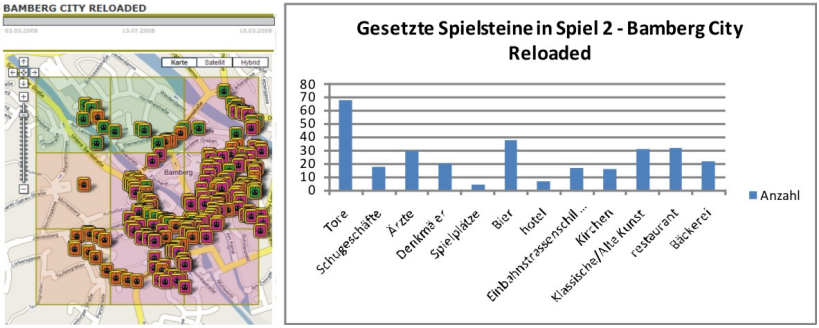


Abbildung 5.4.: Endzustand des zweiten CityExplorer Spieles in Bamberg, © Karte: Google Maps (`http://maps.google.de/`).

Nutzerstudie 3 - Bamberg City Enhanced

An der dritten Studie waren drei Informatikstudenten (Bachelor) und eine ehemalige Informatikstudentin beteiligt, mit einem Altersdurchschnitt von 26,5. In einer etwas kürzeren Spielzeit von vier Tagen wurden insgesamt 188 Spielsteine gesetzt. Ich werde auf dieses Spiel in der Diskussion der Ergebnisse (Abschnitt 5.1.2) noch genauer eingehen, da sich hier ein Aspekt besonders

anschaulich gezeigt hat, der für dieser Art von Spielen - zur gemeinschaftlichen Erhebung raumbezogener Daten - typisch zu sein scheint.

5.1.2. Diskussion der Spielaspekte: nicht-verteiltes Spiel

Zwölf der insgesamt vierzehn Teilnehmer der drei *CityExplorer* Nutzerstudien haben nach ihrem jeweiligen Spiel einen Fragebogen ausgefüllt und an der Interviewrunde teilgenommen. Tabelle 5.1 gibt die Ergebnisse der Auswertung der Fragebögen wieder. Zusätzliche Angaben waren, dass die Spieler zwischen 15 Minuten und 3,5 Stunden pro Tag das Spiel nach eigener Aussage gespielt haben. Außer einer Ausnahme waren alle Spieler mit der Nutzung eines GPS Navigationssystemes vertraut. Obwohl keiner der Teilnehmer einen ortsbezogenen Dienst auf einem mobilen Endgerät, z.B. dem eigenen Mobiltelefon, verwendet hat, hatten zehn von ihnen schon Erfahrungen mit zumindest einem anderen ortsbezogenen Spiel gesammelt, meistens eines mit einem unserer Geogames [134]. Die Zeit die die Teilnehmer mit Spielen (PC/-Konsole oder mobil) durchschnittlich pro Woche verbringen reichte von 0 bis 16 Stunden, wobei fünf der zwölf überhaupt nicht regelmäßig spielen (Angabe von 0 bzw. 0,5 Stunden pro Woche in beiden Kategorien, PC/Konsole oder mobil).

Die Ergebnisse für die zwei den Spielspaß betreffenden Fragen (Nummer 1 und 2 im Fragebogen) zeigen unterschiedliche Tendenzen auf. Die Spieler finden zwar den mobilen Aspekt (Spielsteine setzen, Fotos machen, Spielfeld erkunden) von *CityExplorer* sehr unterhaltsam, können aber mit dem web-basierten Aspekt (XML Dateien hochladen und Spielsteine bewerten) nicht allzu viel anfangen. Die Erklärung für den niedrigeren Wert, die teilweise in der eher prototypische Umsetzung der grafischen Nutzeroberfläche zu finden ist, wurden in den Interviewrunden wie folgt ausgeführt.

Grundsätzlich würden die Spieler eine integrierte Anwendung, die nur auf dem mobilen Endgerät läuft, bevorzugen. Die Spieler empfanden die Zweiteilung des Spieles in einen mobilen und einen web-basierten Teil eher als störend, da nach einer Spielsession im Spielfeld noch Zeit am PC zu Hause eingeplant werden musste, um die erhobenen Daten auf die Webseite hochzuladen. Dies deckt sich auch mit der Angabe im Fragebogen (Nummer 3, 4 und 6), dass zehn der zwölf Spieler sich extra Zeiträume für das Spiel reserviert hatten und diese auch außerhalb der normalen Tagesroutine lagen. Zwar würde das Spielprinzip auch eine zwanglose Einbindung in den normalen Tagesablauf anbieten, ähnlich wie in [128] vorgeschlagen, dem wirkt aber der web-basierte Teil des Spiels entgegen.

Question	Mean	Standard Deviation
1. The outdoor part of the game was fun to play.	4,4	0,5
2. The online part of the game was fun to play.	2,9	0,9
3. The game fitted between my everyday activities (like e.g. shopping, going to work/school or visiting friends).	2,75	0,9
4. I explicitly reserved time slots in my everyday schedule to play the game.	4,25	1,2
5. In order to play the game for free I would agree that the data (GPS positions, pictures and textual annotations) I collected throughout the game are used in a commercial context.	3,4	1,4
6. I played the game only in my free time e.g. after work, school or on the weekend.	4,5	1
7. The game has encouraged me to visit places, which I rarely go to.	4,3	1,1
8. The game has increased my knowledge about the geographic environment (location of places, the street network etc.) of the city the game took place.	3,8	1,4
9. In order to be competitive in the game you have to invest a great amount of time.	3,9	0,8
10. The game has encouraged me to be more on the way outdoors than I usually am.	4,25	1,2
11. I had the feeling that I was actually playing against a real team in the other city, although I did not meet them in person.	3,6	0,9

Tabelle 5.1.: Ergebnisse der Fragebogenauswertung - Likert Skala von 1 (Strongly disagree) bis 5 (Strongly agree) - bei n = 12

Die Idee hinter der indirekten Kommunikation zwischen mobilen Gerät und der Serveranwendung, wie in Kapitel 3 Abschnitt 3.8.3 beschrieben über XML-Dateien, ist, dass das Spiel dadurch keine Verbindungskosten verursacht. Da für die Nutzerstudien aber die Hardware für die Spieler kostenlos bereitgestellt wurde, kam dieser Aspekt in den Überlegungen der Teilnehmer nur am Rande zum Tragen. Beispiele wie *Botfighter* (siehe [79], [124] oder [125]) haben allerdings gezeigt, dass Spieler solche Kosten eher bei langfristiger Nutzung negativ sehen. Vor allem, wenn sie nicht einfach zu kalkulieren sind, wie bei SMS Kosten - wenn nicht ersichtlich ist, wie viele im Laufe eines Spieles gesendet werden [124] - oder GSM/3G Verbindungskosten.

Das web-basierte Reviewing-Verfahren wurde auch als zu umständlich empfunden, wobei nicht der Prozess als solches angesprochen wurde, sondern der dazu bereitgestellte Wizard, siehe Abbildung 3.10. Vor allem die Möglichkeit mehrere Spielsteine gleichzeitig zu begutachten und nicht wie in der jetzigen Version sequenziell, wurde als Verbesserungsmöglichkeit genannt. Insgesamt wurde auch die Integration des Reviewing-Prozesses als solches in den Spielablauf bemängelt. Die Extrapunkteregel, bzw. die Möglichkeit (unkorrekte) Spielsteine von der Wertung auszuschließen, - was einer stärkeren Integration entsprach als in anderen Beispielen aus der Literatur wie beispielsweise in [129] oder [2] - wurde als noch nicht ausreichend bewertet. Dies zeigte sich auch in der Prozentzahl der insgesamt überprüften Spielsteine recht deutlich. Hierauf werde ich in der Analyse der in allen Nutzerstudien gesammelten raumbezogenen Daten in Abschnitt 5.1.4 noch genauer eingehen.

Die Antworten auf die Fragen mit der Nummer 6,7,8 und 10 zeigen deutlich, dass *CityExplorer* gut geeignet ist, die Teilnehmer zu motivieren, mehr Zeit in ihrer jeweiligen geographischen Umgebung zu verbringen, als sie es normalerweise tun würden. Auch weckt so ein Spiel den Entdeckergeist der Spieler und die Spieler sind mehr geneigt auch Orte aufzusuchen, die sie normalerweise nicht aufsuchen würden. Im Spielkontext änderten die Teilnehmer bereitwillig ihr normales Verhalten.

Dies deckt sich mit Beobachtungen von denen Benford et al. (2006) in [83] berichteten. In *Uncle Roy All Around You* [123] müssen Spieler über Hinweise, die sie von professionellen Schauspielern im Spiel erhalten, den Standort von *Uncle Roy* herausfinden und zu ihm gelangen. Die Spieler kennen aber die Identität der Kontaktpersonen im Spiel nicht genau, sondern bekommen nur allgemeine Hinweise, z.B. über deren Aussehen („trägt ein weißes T-Shirt“).

Obwohl die meisten Menschen eher zurückhaltend gegenüber unbekannten Menschen auf der Straße sind, war es für die Spieler im Kontext des Spieles kein Problem dieses Verhalten abzulegen und spontan fremde Menschen anzusprechen.

Gegen Ende des Spieles müssen die Spieler sogar zu einem Fremden, von dem sie nicht sicher sein können, ob er zum Spiel gehört oder nicht, in ein Auto steigen und sich von ihm zu einem unbekannten Ort fahren lassen. Obwohl einige Spieler Bedenken nach dem Spiel geäußert haben, wurde diese Situation trotzdem als zum Spiel gehörend akzeptiert und deswegen von den Spielern mitgemacht.

Nieme et al. (2005) [82] haben zwei Eigenschaften identifiziert, die für ortsbezogene Spiele wichtig sind, um den oben angesprochenen Verhaltenswechsel bei den Spielern zu erreichen: Anonymität der Nicht-Spieler (*anonymity*) und Verantwortung (*accountability*) - wobei mit letzterem die Möglichkeit gemeint ist, dass die Nicht-Spieler die Spieler für ihre Aktionen auch bestrafen können. Da in *CityExplorer* nur in diesem Sinne anonyme Orte, bzw. Merkmale, des Spielfeldes in das Spiel einbezogen werden, ist das veränderte Verhalten der Spieler durchaus mit diesen Ergebnissen erklärbar.

Die Arbeit von Anderson et al. (2007) [153] stützt ebenfalls die These, dass *CityExplorer* auch als Motivator dienen kann, sich mehr physisch zu bewegen. Die Autoren untersuchten das Verhalten von Nutzern einer mobilen Anwendung (*Shakra*), die den Nutzern täglich ihre „Bewegungsaktivität“ präsentiert. Dies alleine führte laut den Ergebnisse der Arbeit schon zu einer gesteigerten physischen Aktivität der Anwender. Obwohl manche Nutzer dies laut den den Autoren als eine Art Spiel oder Wettkampf empfanden, sollte sich eine Einbindung in einen wirklichen Spielkontext vor allem auf langfristige Sicht, noch positiver auf die Motivation der Spieler auswirken. Damit würde die Anwendung auch etwas von ihrem sportlichen bzw. Trainingscharakter verlieren.

Ein oft genanntes Thema in allen drei Interviewrunden war die Auswirkung des Spieles auf die Wahrnehmung der Spieler bezüglich der geographischen Umgebung in dem das Spiel statt findet. Alle Spieler merkten an, dass sich ihre Sichtweise auf ihre tägliche Umgebung mit dem Spielen von *CityExplorer* verändert hatte. Dies traf verständlicherweise vor allem auf die Merkmale im Spielfeld zu, die zu den vor dem Spiel gewählten Kategorien gehörten. Beispielsweise merkte ein Spieler nach dem ersten Spiel an, dass er sehr verwundert war, wie viele Ärzte (Kategorie „doctor“) es in der Innenstadt von Bamberg gibt. Ähnliches berichteten zwei weitere Spieler nach dem zweiten Spiel in Bamberg. Beide hatten gedacht, dass es nur ein paar wenige Friseure

(Kategorie „hair stylist“) im Spielfeld gebe. Aber während des Spiels hatten sie dann das Gefühl, dass fast in jeder Straße in Bamberg ein Friseur ansässig ist.

Aber nicht nur die Merkmale die zu Kategorien gehörten, traten während des Spiels mehr in den Fokus der Spieler. Zwei der Teilnehmer des zweiten Spiels berichteten, dass sie beim Spielen zahlreiche Merkmale entdeckten, die sich gut als Kategorien geeignet hätten. Die Spieler hatten aber vor dem Spiel gedacht, dass es nicht genug Ausprägungen der Kategorien im Spielfeld geben würde, um sie tatsächlich zu verwenden.

Die Ergebnisse zur Frage 9 im Fragebogen und besonders der Spielverlauf des dritten CityExplorer Spieles zeigen ein noch bestehendes Problem nicht nur von *CityExplorer*, sondern von ortsbezogenen Spielen zur Erhebung von raumbezogenen Daten im Allgemeinen. Bell et al. (2006) [128] berichten in ihrer Arbeit, dass solche Spiele Probleme bekommen, wenn die Motivation zwischen den Spielern nicht sehr ausgeglichen ist. D.h. es Spieler gibt, die viel mehr Zeit in das Spiel investieren als alle anderen und dadurch sicher das Spiel gewinnen. Normalerweise sollte das auch so sein, aber da Spiele wie *CityExplorer* oder *Feeding Yoshi* konzipiert sind, Daten mit Ihnen zu sammeln, sollten sie so lange wie möglich alle beteiligten Spieler dazu motivieren weiter zu spielen. Des Weiteren sollte, zumindest wenn so ein Spiel Spielelemente zur Qualitätssicherung enthält, eine möglichst große Anzahl an Spielern möglichst lange spielen und nicht nur ein oder zwei, damit die Qualität der erhobenen Daten dieser Spieler auch überprüft wird.

Bell et al. verwendeten in ihrer Studie ein Spiel ohne strategische Komponente. Simpel ausgedrückt suchen die Spieler in einem unbegrenzten Spielfeld geschlossene und offene WLAN-Knoten. In den geschlossenen Knoten können verschiedene Früchte aufgesammelt werden und in den offenen können damit Yoshis gefüttert werden. Für das Füttern der Yoshis bekommen die Spieler Punkte. Zwar gibt es unterschiedliche Typen von Yoshis, die nur bestimmte Früchte fressen, aber im Prinzip gewinnt derjenige Spiele sicher, der mehr Zeit in das Spiel investiert und einfach mehr WLAN-Knoten „abgrast“.

Genau das ist auch bei den Nutzerstudien zu *Feeding Yoshi* passiert. Eine Spielergruppe - das Spiel wurde in Teams verteilt in verschiedenen Städten gespielt - dominierte, in dem sie einfach mehr spielte. Einen ähnlichen Spielverlauf zeigte auch das dritte Spiel in Bamberg, welches Abbildung 5.5 darstellt.

Gleich zu Beginn des Spieles, am Ende des ersten Tages, hatte Spieler 2 ca. 70 Spielsteine gesetzt und damit sechs der neun Spielsegmente für sich be-

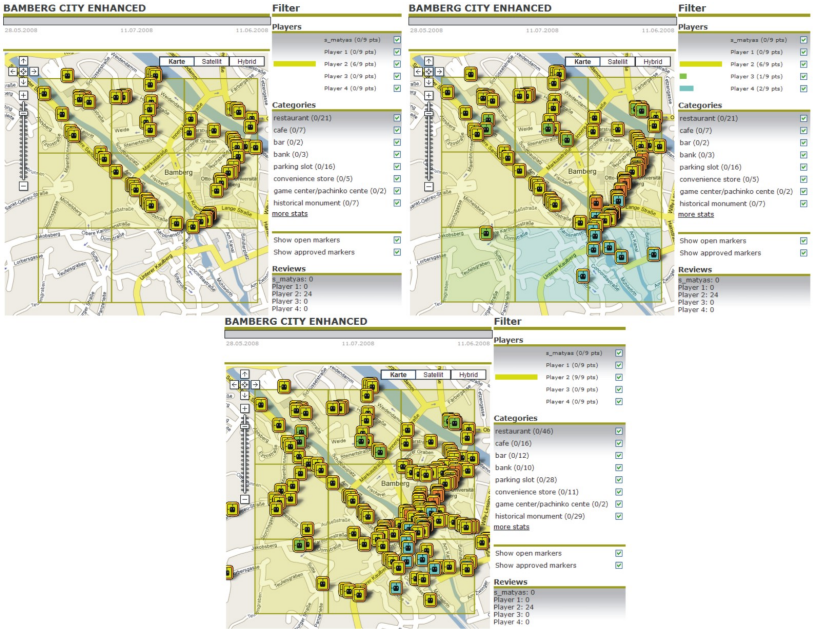


Abbildung 5.5.: Spielverlauf des dritten CityExplorer Spieles in Bamberg und typisches Beispiel für das *Überflutungsproblem* bei ortsbezogenen Spielen zur Datenerhebung (von links oben über rechts oben nach links unten), © Karte: Google Maps (`http://maps.google.de/`).

ansprucht. Wie in der Interviewrunde nach dem Spiel ersichtlich wurde, hatte das die Motivation der anderen Spieler so beeinträchtigt, dass sie nur noch weiterspielten, weil es sich um eine Studie handelte und nicht weil sie noch eine Aussicht auf einen Sieg sahen. Dementsprechend kamen in den folgenden Tagen (Abbildung 5.5 zeigt hier drei Ausschnitte aus dem Spielverlauf) auch fast alle weiteren Spielsteine von Spieler 2. Abbildung 5.6 zeigt die Endverteilung der gesetzten Spielsteine zwischen allen Spielern und Spieler 2.

CityExplorer ist so konzipiert das dieses *Überflutungsproblem*, wie ich es nenne, von den Spielregeln zumindest abgeschwächt wird. Natürlich sollen fleißige Spieler für ihren Eifer belohnt oder zumindest nicht bestraft werden. Deshalb habe ich das Spielfeld von *CityExplorer* in Segmente eingeteilt und die Kategorienwertung eingeführt. Dadurch macht es auch noch Sinn in

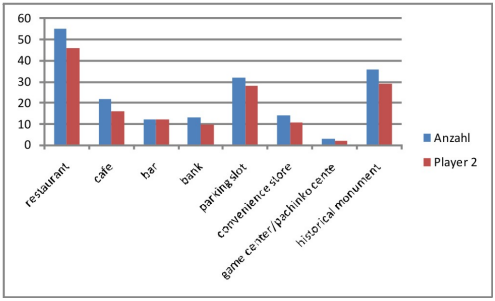


Abbildung 5.6.: Gesetzte Spielsteine in Spiel 3 insgesamt (blauer Balken) und von Spieler 2 (roter Balken)

einem Segment Spielsteine zu setzen, obwohl man möglicherweise an die Anzahl an Spielsteinen des Führenden in einem Segment nicht mehr herankommt. Trotzdem kann man sich durch diesen Schritt aber eine Kategoriewertung sichern, falls beispielsweise in den anderen Segmenten keine oder nicht mehr so viele Merkmale der gesuchten Kategorie vorkommen.

Aus diesem Grund war das in Abbildung 5.5 gezeigte Spiel noch lange nicht nach diesem ersten Tag verloren. Nur zwei Segmente und zwei Kategorien waren in diesem Sinne „überflutet“, mit 14 und 15, bzw. 21 und 16 Spielsteinen. Alle anderen Segmente und Kategorien waren noch offen. Die anderen Spieler nahmen aber natürlich richtigerweise an, dass Spieler 2 mit diesem Eifer weiterspielen würde und waren selber nicht bereit so viel Zeit einzusetzen.

Deshalb muss bei einem solchen Spielkonzept, wie Bell et al. schon vorgeschlagen und auch die Spieler in der Interviewrunde angemerkt haben, die Gesamtanzahl an setzbaren Spielsteinen pro Spielzug, beispielsweise an einem Tag, doch beschränkt werden. In *CityExplorer* habe ich dieses Spielelement nicht eingesetzt, weil natürlich ein oberes Limit für die mögliche Anzahl an sammelbaren Datensätzen festlegt, die man mit einer Spielrunde eines solchen Spieles erreichen kann. Wenn das Spiel aber am Ende von den Spielern nur als Arbeit angesehen wird, spielt es natürlich auf lange Sicht auch keiner. Hier ein ausgewogenes Maß oder vielleicht auch völlig andere Spielelemente zu finden, die das *Überflutungsproblem* nicht haben, erfordert weitere Studien.

5.1.3. Nutzerstudie zum verteilten Spiel

Wie in Abschnitt 3.8.2 von Kapitel 3 vorgestellt, ist *CityExplorer* auch verteilt spielbar, d.h. die Spieler müssen nicht notwendigerweise auf ein und demselben Spielfeld spielen, sondern können sich beispielsweise in geographisch getrennten Städten befinden. Dabei ist die Anzahl der kombinierbaren Spielfelder beliebig groß, wobei die Übersichtlichkeit, visuell wie kognitiv, ab einem gewissen n aber schwierig wird.

Für das verteilte Spiel wurden zwei Spielfelder ausgewählt, und zwar das schon in den drei vorherigen Studien genutzte Spielfeld in Bamberg und ein entsprechendes Spielfeld in der japanischen Stadt Fujisawa (nahe Yokohama, um die Shonandai Haltestation herum). Da *CityExplorer* auf einem unstrukturierten Spielfeld gespielt wird, ist demzufolge - siehe Abschnitt 3.6.3 von Kapitel 3 - nur das Ausmaß der zwei Spielfelder für die Fairness entscheidend.

Zwei Besonderheiten in Bezug auf das Spiel *CityExplorer* müssen dennoch beachtet werden und ziehen entsprechende Regeländerungen nach sich:

1. Spieler in der jeweils anderen Stadt können nur die Spielsteine der Spieler aus der eigenen Stadt begutachten. Wird ein Teamspiel durchgeführt, d.h. eine Stadt spielt gegen eine andere, muss pro Stadt ein neutraler „Schiedsrichter“ bestimmt werden, der die Spielsteine in seiner Stadt überprüft. Dies erhöht die Kosten zur *Spieldurchführung* („orchestration challenge“ nach [97]) nach Kapitel 2 Abschnitt 2.3.2, kann aber für die Motivation der beteiligten Spieler förderlich sein [147]. Die Extrapunkte für die meisten Begutachtungen werden in der Teamvariante nicht vergeben.

2. Die Kategorien müssen für alle Städte, die an dem verteilten Spiel teilnehmen, gültig sein. Es müssen in allen Städten ungefähr gleich viele Merkmale zu einer bestimmten Kategorie vorhanden sein.

Trotz der höheren Inszenierungskosten wurde diese Studie als Teamspiel durchgeführt. Zwei Teams aus jeweils vier Studenten in beiden Städten traten gegeneinander an. Das Bamberger Team bestand aus vier Spielern, die auch schon an einer der anderen Studien teilgenommen hatten und vier Studenten der Keio Universität (aus nicht-technischen Fachrichtungen, zwei männliche und zwei weibliche). Der Altersdurchschnitt der japanischen Spieler lag bei 20,7 Jahren. Insgesamt lief das Spiel drei Tage und es wurden von beiden Teams 106 Spielsteine (96 in Bamberg und 29 in Fujisawa, mit 16 abgelehnten Spielsteine) gesetzt. Abbildung 5.7 zeigt das Endergebnis dieses Spieles.

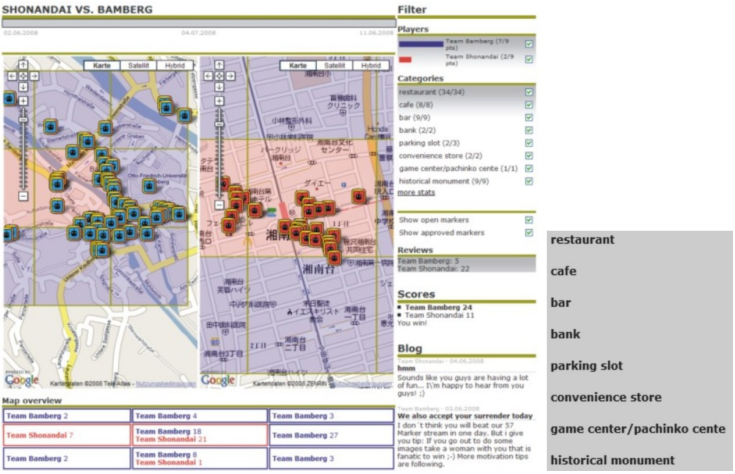


Abbildung 5.7.: Endzustand des verteilten Spieles zwischen Bamberg und Fujisawa und die im Spiel verwendeten Kategorien, © Karte: Google Maps (`http://maps.google.de/`).

Auf eine detaillierte Diskussion der Ergebnisse der Fragebogenauswertung verzichte ich an dieser Stelle, was den *dynamischen Bedingungen* („dynamic conditions“ aus [4]) zu Verschulden ist. Diese Bedingungen, die sich von Spieldurchgang zu Spieldurchgang ändern, die man aber nicht durch ein entsprechendes Spieldesign abbilden kann, sollten auf den beiden Spielfeldern ähnlich sein. In diesem Fall war es das Wetter, welches eine vollständige Durchführung der Studie verhinderte. Für das japanische Spielfeld war im Spielzeitraum eine Taifunwarnung aktiv, was den Spielverlauf entsprechend einschränkte. Deshalb ist eine genaue Bewertung auf Grund der wenigen vorliegenden Daten nicht möglich - drei der vier japanischen Spieler haben aber trotzdem den Fragebogen ausgefüllt (siehe Anhang A.1.3). Die Antworten zur Frage Nummer 11 bewegen sich hierbei im Mittelmaß, wobei die Spieler aber wohl eher das Potenzial dieser Spielweise bewertet haben, als das durchgeführte Spiel selbst.

5.1.4. Analyse der erhobenen raumbezogenen Datensätze

Neben der Quantität, die ich in Abschnitt 5.1.1 schon angesprochen habe, ist die Qualität der erhobenen raumbezogenen Daten fpr mich von Bedeutung.

Hier hat mich vor allem interessiert, welcher Prozentsatz der erhobenen Datensätze von den Spielern selbst begutachtet wurde und wie groß der dadurch behobene Fehler war.

Einen ersten Richtwert für meine Untersuchung lieferten die Ergebnisse der Arbeit von Grant et al. (2007) [2]. In deren *MobiMission* Spiel haben die Teilnehmer die Möglichkeit die Missionen der anderen Spieler zu kommentieren, bzw. bewerten. Zu 42% der Missionen wurden Kommentare von mindestens einem anderen Spieler abgegeben. Die Einbingung der Kommentar-/Bewertungsfunktion ist in *MobiMission* dabei nur sehr lose realisiert.

Bell et al. (2009) [156] haben vor kurzem eine noch engere Einbindung der Qualitätssicherung in das Spielgeschehen vorgestellt, welches weitestgehend dem *Output-Input*-Entwurfsmuster aus Abschnitt 3.5.2 entspricht. Trotz dieser direkten Einbindung des Reviewprozesses in das Spielgeschehen berichten Bell et al., dass in ihren beiden Nutzerstudien 43% - bei der Betrachtung nur der von den Spielern selbst erstellten Datensätzen geht der Wert auf 33% zurück -, bzw. 40%, der im Spiel verwendeten Datensätze durch die Spieler begutachtet wurde.

Wie ich schon in [217] und [98] vorgestellt habe, bin ich bei der Analyse der Datensammlung der *CityExplorer* Spiele auf ganz ähnliche Ergebnisse gekommen. Von den in der gesamten Datenbasis vorhandenen 922 Datensätzen (dazu zählen auch Spielsteine aus einem kleineren Feldversuch in der Innenstadt von Yokohama) wurden 42% davon mindestens einmal begutachtet - zieht man nur die Spielsteine in Betracht, die in den hier vorgestellten Nutzerstudien erhoben wurden, geht der Wert auf 39% zurück. Je nach dem welchen Wert man zugrunde legt, bleiben damit 58%, bzw. 61%, der Datensätze qualitativ nicht abgesichert. Hier muss allerdings festgehalten werden, dass eine manuelle Untersuchung zumindest der gemachten Fotos und der gewählten Kategorien zu den Spielsteinen ergeben hat, das diese zwei Information für alle Spielsteine korrekt sind - siehe Kapitel 4 Abschnitt 4.1.1 für die formale Definition eines Spielsteines in *CityExplorer*. Auch wurden bei den verbesserten Datensätzen nur die Position von POV (Position, von dem das Photo gemacht wurde) und APL (Position des Merkmals, an dem der Spielstein gelegt wurde) ausgebessert, Photos und Kategorieinformation waren immer korrekt.

Auf Grundlage der durch die *CityExplorer* Nutzerstudien erbrachten Ergebnisse, sowie der Erkenntnisse aus den zwei fremden Arbeiten - [2] und [156] -, scheint eine Begutachtungsrate von um die 40% ein realistischer Wert zu sein, den man durch diese Art von ortsbezogenen Spielen erwarten kann.

Um nun die Qualität der nicht-bewerteten Datensätze abschätzen zu können,

habe ich zwei weitere Qualitätsmaße berechnet: das innere Qualitätsmaß (IQM - Inner Quality Measure) und das äußere Qualitätsmaß (OQM - Outer Quality Measure).

Zwei Arten von Positionsangaben werden von Spielern nach Abschnitt 4.1.1 in Kapitel 4 erhoben: Der Standpunkt, von wo das Foto aufgenommen wurde POV und der Standort des aufgenommenen Merkmales APL. Das IQM wird für beide Positionsangabe wie folgt berechnet: Sei M die Menge der APL und I die Menge der POV aller mrk in der Datensammlung. pi_m ist dann die initiale APL Position eines mrk und pi_i die initiale Position des POV desselben mrk. pr_m ist entsprechend die durch das Review-Verfahren festgelegte APL Position, sowie pr_i die des POV. Die Funktion $dist$ ist definiert als die Distanz zwischen der initialen und der im Review festgelegten Position eines mrk.

Damit ergeben sich folgende zwei Gleichungen:

$$IQM_m = pi_m \; dist \; pr_m \tag{5.1}$$

$$IQM_i = pi_i \; dist \; pr_i \tag{5.2}$$

Wenn wir nun die beiden Gleichungen auf die begutachteten mrk in der Datensammlung der vier *CityExplorer* Spiele (die drei Spiele in Bamberg und das verteilte Spiel) anwenden, bekommen wir laut Tabelle 5.2 folgende Ergebnisse (die genauen Werte für IQM_m und IQM_i für die einzelnen Spielsteine aus Spiel 1 und 2 finden sich im Anhang in Abschnitt A.2):

	Spiel 1	**Spiel 2**
APL (Mittelwert)	14,36	14,15
POV (Mittelwert)	12,42	15,03

Tabelle 5.2.: Mittelwert für die IQM Werte für Spiel 1 und 2 in Bamberg

Die Werte zeigen deutlich, dass die Teilnehmer der vier Nutzerstudien erstens im Mittel entsprechend den technischen Möglichkeiten genau gemessen haben und zweitens selbst kleine Abweichungen von weniger als 5 Meter korrigiert wurden. Die berechneten Werte sind deshalb so bemerkenswert, weil der gemessene mittlere Fehler in der Positionsgenauigkeit (PDOP - Position Delution of Precision nach [101] bei GPS, selbst bei optimalen Bedingungen,

gemitteltet über eine Messreihe von 24h, um die 10 Meter liegt. Unter normalen Bedingungen wird eine mittlere Genauigkeit häufig mit einem Wert zwischen 10-15 Metern beziffert.

Zur Berechnung des OQM werden lediglich die Angaben zum APL der eingestellten Spielsteine verwendet. Hier soll die Genauigkeit der Positionsangaben in Bezug auf eine externe Referenz berechnet werden. M ist wiederum die Menge an APL Angaben, Q definiert die Menge von Koordinatenangaben aus einer Menge von extern - durch Experten - referenzierte POI. p_m ist die APL Koordinate eines bestimmten mrk und p_p das entsprechende Gegenstück in der Menge Q. Die Funktion $dist$ berechnet den Abstand zwischen der in *CityExplorer* aufgenommenen Koordinate und der aus einer extern stammenden Referenzierung des semantisch gleichen Merkmals.

OQM ist dann folglich definiert als:

$$OQM_{mp} = p_m \; dist \; p_p \tag{5.3}$$

Zur Berechnung des OQM für die *CityExplorer* Datensammlung habe ich mir aus insgesamt fünf Kategorien zufällig eine begrenzte Anzahl an Datensätzen, p_m, - sieh Abbildung 5.8 für die genauen Zahlen - gewählt. Anzumerken ist hier, dass ich die Stichprobe aus der gesamten Datenbasis entnommen habe, nicht nur aus der Teilmenge der begutachteten Spielsteine.

Zu den ausgewählten Spielsteinen habe ich mir dann über die Google Maps Suche [1] die entsprechenden Gegenstücke gesucht, hier als p_p bezeichnet. Die Variable $dist$ berechnet die Entfernung zwischen den zwei Datensätzen. Die Ergebnisse des OQM_{mp}, aufgeschlüsselt nach Kategorien, sind in Abbildung 5.8 dargestellt.

Tabelle 5.3 fasst die Kennzahlen zum OQM für die gesamte Stichprobe (n = 93) zusammen.

Total	
Mittelwert	33,02

Tabelle 5.3.: OQM Werte für die CityExplorer Datensammlung

Nimmt man die beiden Ergebnisse zusammen, stellt man fest, dass die Genauigkeit der Angaben im Mittel etwa doppelt so hoch liegt, wie die mittlere Positionsgenauigkeit bei GPS. Dadurch, dass ich für die Stichprobe sowohl

[1] `http://maps.google.de/`

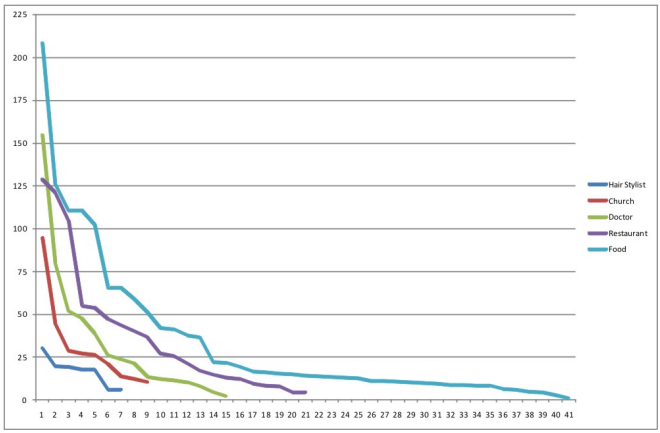


Abbildung 5.8.: OQM Verteilung für die Stichprobe aus den Kategorien „hair stylist“ (n = 7), „church“ (n = 9), „doctor“ (n = 15), „restaurant“ (n = 21) und „food“ (n = 41), insgesamt n = 93

begutachtete als auch nicht-begutachtete Datensätze verwendet habe, ist dieser Wert als gut zu bewerten. Vor allem wenn man noch die Qualität der Referenzmenge Q in Betracht zieht, die nicht immer einwandfrei ist. Als Beispiel soll Abbildung 5.9 dienen.



Abbildung 5.9.: Vergleich von p_m und p_p für Spielsteine - gelbe Stecknadel - aus M und Google Maps POI aus Q - roter Pin, © Karte: Google Maps (`http://maps.google.de/`).

Die grünen Markierungen in Abbildung 5.9 zeigen jeweils den tatsächlichen Standort der beiden Merkmale „Rathauschenke“ und „Weinfass“. Wie man erkennen kann, sind bei diesen zwei Datensätzen die in *CityExplorer* erhobenen APL genauer als die in Google Maps vorhandenen Angaben. Da

der OQM die Distanz zwischen der gelben und der roten Markierung berechnet, fließt der Fehler in den Google Maps Daten in die Berechnung mit ein. Trotzdem ist das OQM-Maß als grober Richtwert sinnvoll und zeigt die hohe Qualität der mit *CityExplorer* erhobenen raumbezogenen Daten.

5.1.5. Analyse der Geogames GPS-Tracks

Neben den raumbezogenen Datensätzen in Form der Spielsteinen in *CityExplorer*, wird das Entwurfmuster der *Log*-Dateien (siehe Kapitel 3 Abschnitt 3.4.1) noch in zwei anderen ortsbezogenen Spielen, die ebenfalls mit dem Geogames Software-Framework (siehe Kapitel 3 Abschnitt 3.8.3) entwickelt wurde. Während der Spiele *GeoTicTacToe* [134] und *CityPoker* [99] werden alle zwei Sekunden eine GPS Koordinate aufgenommen und in einer Log-Datei gespeichert. Um die CPU-Belastung niedrig zu halten, wird auf dem mobilen Gerät keine Qualitätskontrolle der GPS Koordinaten durchgeführt. Auf diesem Weg wurden insgesamt 93047 GPS Koordinaten gesammelt.

Zur Bestimmung des Anteils an qualitativ „guten“ GPS Koordinaten, habe ich folgendes Qualitätskriterium angesetzt: Alle GPS Koordinaten, die mit weniger als vier Satelliten erhoben wurden, wurden als „schlecht“ markiert und aus den Log-Dateien aussortiert.

Obwohl drei Satelliten ausreichend sind, um eine GPS Position zu berechnen, erhöht die Anzahl an Satelliten die Genauigkeit der Positionsangabe. Weitere Kennzahlen, wie die Dilution of Precision [101] Werte (DOP), können helfen, qualitätsmindernde Einflüsse zu erkennen. Da allerdings nicht alle GPS Geräte alle DOP-Werte ausgeben, habe ich mich für den kleinsten gemeinsamen Nenner zwischen den unterschiedlichen Protokollen entschieden, d.h. wirklich nur die Satellitenanzahl als Kriterium anzusetzen.

Insgesamt habe ich Daten von 48 Spielrunden aus unterschiedlichen Spielevents der beiden Spiele, *GeoTicTacToe* und *CityPoker*, untersucht. Das Alter der Spieler lag zwischen 5 und 55 Jahren und die Spiele fanden auf vier unterschiedlichen Spielfeldern statt, in der Innenstadt von Bamberg, an der Universität Bamberg (WIAI Fakultät in der Feldkirchenstrasse, kurz FEKI), in Bremen und in Coburg. Eine genaue Beschreibung der Events findet sich in in Matyas (2007) [5]. Als GPS Geräte wurden die auch in *CityExplorer* verwendeten, XAIOX Wonde-XL eingesetzt.

Mit dem einfachen Qualitätsmaß, der Satellitenanzahl, bleiben von den 93047 Datensätzen noch 45565 übrig. Dies entspricht einer Ausfallrate von 49% - die genaue Aufteilung findet sich im Anhang in Abschnitt A.3. Abbildung 5.10 zeigt die bei einem der Spielevents erhobenen GPS Tracks für die

Innenstadt von Bamberg.



Abbildung 5.10.: GPS Track erhoben während der Girls Day CityPoker [6] Spiele in der Innenstadt von Bamberg, © Karte: Google Earth (`http://earth.google.com/intl/de/`).

Obwohl 51% der GPS Koordinaten als von nicht ausreichender Qulität eingestuft werden mussten, entspricht die verbliebene Datensammlung doch einem Wegenetz von ca. 157 Kilometern. Wenn man sich die *CityPoker* Spiele in der Innenstadt von Bamberg genauer anschaut, die eine Dauer von ca. 2 Stunden hatten, wurden dort im Schnitt pro Spielrunde ca. 7 Kilometer an „guten“ GPS Tracks erhoben. Nimmt man das gesamte Wegenetz von Bamberg als Maßstab, 262 Kilometer, bräuchte man ungefähr 37 Spielrunden, um dieses mit einem solchen ortsbezogenen Spiel komplett erheben zu lassen. Bei zwei Spierunden pro Tag, könnte dies schon in 18,5 Tagen erreicht werden. Bei den bisherigen *CityPoker* Spielen traten immer zwei Teams gegeneinander an, wobei aber pro Team nur ein GPS Gerät verwendet wurde. Somit würde eine Spielgemeinschaft von 74 Spielern ausreichen, um diese Anzahl

an Spielen zu erreichen, wenn wir davon ausgehen, dass keiner der Spieler das Spiel zweimal spielt, was laut [6] eine eher konservative Einschätzung ist.

Obwohl Abbildung 5.10 zeigt, dass die Datensammlung ohne nachgeschalteten Integrationsprozess noch nicht für professionelle Anwendungen, beispielsweise als Datengrundlage für Navigationssoftware, geeignet ist, kann man in Abbildung 5.11 erkennen, dass die Daten zumindest für thematische Karten im Maßstab 1:25000 hinreichend genau sind.

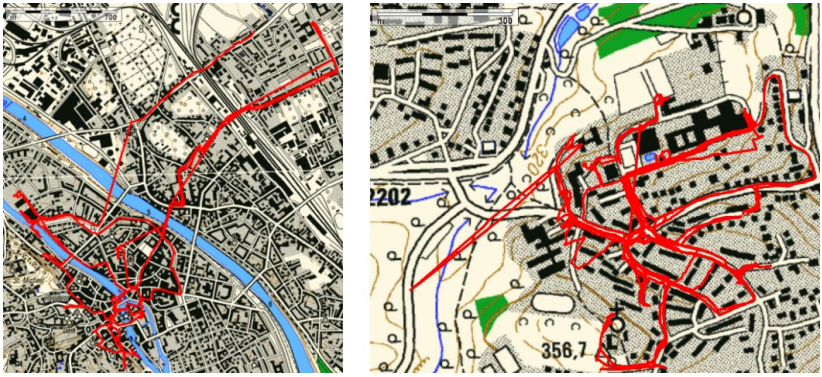


Abbildung 5.11.: Thematische Karten im Maßstab 1:25000 des Landesvermessungsamt Bayerns (TK25) für Bamberg und Coburg; in Rot mit den beiden Geogames erhobene GPS Tracks

5.2. Wisdom-of-the-Crowd Datenerhebung: GeoSnake

In Kapitel 3, Abschnitt 3.5.3, habe ich das *Widom of the Crowd*-Entwurfsmuster vorgestellt, bei dem es darum geht, Spielmechanismen in ein ortsbezogenes Spiel einzubauen, die es ermöglichen, Mehrheitsentscheidungen von Spielern zu bekommen, obwohl die Spieler diese anonym und ohne Kenntnis der Entscheidungen der übrigen Mitspieler treffen.

Bisher wurde dieses Entwurfsmuster nur im Sinne eines Bewertungssystems, beispielsweise realisiert als Punkteskala, innerhalb eines Peer-Review-Prozesses (siehe z.B. [129]) verwendet. Wie ich im Folgenden zeige, kann

dasselbe Prinzip auch schon während der Datenerhebung selbst angewendet werden. Der Vorteil dieser Methodik liegt darin, dass die Qualitätssicherung schon in den Datenerhebungsprozess integriert ist und nicht als Extra-Spielmechanismus eingebaut werden muss.

Die Prämisse lautet folgendermaßen:

Spieler werden aufgefordert raumbezogene Daten in einer geographischen Umgebung so zu erheben, d.h. vor allem die GPS Koordinaten für ein Objekt, bzw. Merkmal, so zu wählen, wie die Mehrheit der übrigen Spieler es auch tun würde.

Ausgangspunkt ist die Problematik, dass normalerweise alle Arten von geographischen Merkmalen, seien es Objekte, wie Statuen, Gebäude oder gar ganze Regionen, wie Plätze oder Stadtteile, bzw. ganze Städte, mit Punktkoorindaten repräsentiert werden [25]. Nun stellt sich aber die Frage, wie man dafür am Besten eine 0-D-dimensionale Repräsentation wählt, d.h. welche Punktkoordinaten sind sinnvoll.

Ein Gedanke, der bei der Konzeption der im nachfolgenden vorgestellten Experimentreihe eine Rolle spielt, ist die Art und Weise, wie zur Zeit raumbezogene Daten von geowissenschaftliche Laien erhoben werden. Ein Anwendungsfall, ist der, dass raumbezogene Datensätze häufig am Bildschirm manuell „eingezeichnet“ werden. Das bedeutet, dass der Anwender auf Grundlage z.B. eines digitalen Luftbildes, die darauf enthaltenen Merkmale per Mausklick manuell digitalisiert. Diese Arbeitsweise wird zum Beispiel von der Community-Seite OpenStreetMaps [2] unterstützt.

Ein anderer Anwendungsfall, wie geowissenschaftliche Laien raumbezogene Daten erheben, ist der in Abbildung 5.12 gezeigte. Die Daten werden im Kontext einer anderen Tätigkeit erhoben, wobei das häufigste Szenario das Aufnehmen von Fotos ist. Die gemachten Fotos werden dabei entweder in Echtzeit, durch den Foto oder ein mitgeführtes GPS Gerät, georeferenziert oder später wiederum manuell am Bildschirm. Die von den Anwendern bereitgestellten Georeferenzierungen beziehen sich aber meistens nicht auf das fotografierte Motiv (also in meinem Fall dem OI) sondern dem Standort, von wo aus das Foto gemacht wurde (in meinem Fall wäre das der POV).

Abbildung 5.12 gibt zahlreiche Lokalisierungen für das geographische Merkmal „Altes Rathaus“ in Bamberg wieder. Wie man sieht, streuen die vorgeschlagenen Standorte über eine große geographische Ausdehnung hinweg.

[2] `http://www.openstreetmap.de/`

Das Gebäude selbst besitzt aber eigentlich nur die grün markierte Ausdehnung.



Abbildung 5.12.: Suchergebnis auf der Community-Webseite Locr (`http://de.locr.com/`) für raumbezogene Datensätze zum Stichwort „Altes Rathaus Bamberg“

Die Frage, die sich hier stellt, ist die, ob es eine Möglichkeit gibt, die Streuung in den erhobenen Datensätzen so gering wie möglich zu halten. Dies würde ebenfalls die Ergebnisse des Integrationsverfahrens aus Kapitel 4 Abschnitt4.1 verbessern, ganz gleich welcher Aggregationsalgorithmus dort verwendet wird.

Eine Experimentreihe, eingebettet in einen spielerischen Kontext, soll Aufschluss darüber geben, ob das *Wisdom of the Crowd*-Entwurfsmuster geeignet ist, geowissenschaftliche Laien dazu zu bringen, qualitativ bessere Datensätze zu erheben, als sie es bisher tun.

Das Experiment soll so nah wie möglich an einen echten Spielkontext heran reichen, dabei aber trotzdem noch in einer kontrollierteren Umgebung ablaufen, als beispielsweise die Nutzerstudien zu *CityExplorer*. Damit soll sichergestellt sein, dass auch nur das Entwurfsmuster selbst den Ausschlag zur Qualitätssteigerung gibt. Das sehr einfach gehaltene ortsbezogene Spiel *GeoSnake* wurde unter den genannten Bedingungen von mir entworfen.

Das Spiel *GeoSnake* beruht auf der Spielidee des Videospielklassikers *Snake* [3]. In *Snake* steuert der Spieler eine Schlange, deren Aufgabe es ist, die im Spielfeld verteilten, bzw. immer an unterschiedlichen Stellen auftauchenden Früchte aufzusammeln, bzw. zu fressen. Durch das Aufsammeln der Früchte

[3] `http://en.wikipedia.org/wiki/Snake_(video_game)`

wächst die Schlange in ihrer Länge jeweils um einen festen Wert. Die Herausforderung liegt darin, immer weiter Früchte aufzusammeln, ohne den eigenen wachsenden Körper zu berühren. Geschieht dies dennoch, verliert man ein Leben und muss von vorne beginnen.

Die Regeln für die ortsbezogene Variante, *GeoSnake*, sehen nun wie folgt aus - im Folgenden Regelvariante 1 genannt:

1. Die Aufgabe der Spieler ist es, ähnlich dem Aufsammeln der Früchten im Original, im Spielfeld verteilte Orte abzulaufen.

2. Um einen Ort erfolgreich abzulaufen, müssen die Spieler eine GPS Koordinate für diesen Ort aufnehmen.

3. Die Spieler dürfen ihren zurückgelegten Weg nicht überschneiden und auch keinen Weg „doppelt“ ablaufen, d.h. alle Wege sind faktisch wie Einbahnstraßen zu behandeln. Übertragen auf Snake bedeutet dies, dass der Körper der Schlange kontinuierlich entlang des vom Spieler zurückgelegten Weges wächst - ähnlich den „Lichtrennern“ aus dem Film/-Spiel Tron [4].

Für jeden erfolgreich abgelaufenen Ort erhalten die Spieler einen Punkt. Für jeden Verstoß gegen Regel 3 bekommen die Spieler einen Punkt abgezogen, können aber ansonsten normal weiterspielen. Gewonnen hat derjenige Spieler mit den meisten Punkten am Ende eines gesamten Experimentdurchlaufs.

5.2.1. Hypothesen

Die von mir durchgeführte Experimentreihe bestand aus einer Vor- und einer Hauptstudie. Durch die beiden Experimente werden folgende Hypothesen untersucht:

1. Das Lokalisieren eines Ortes am Bildschirm führt zu einer größeren Fehlerstreuung als mit einem GPS Gerät vor Ort.

2. Das Lokalisieren eines Ortes mit einem GPS Gerät vor Ort mit Hilfe einer Umsetzung des *Widom of the Crowd*-Entwurfsmusters führt zu einer kleineren Fehlerstreuung als ohne.

[4] `http://de.wikipedia.org/wiki/Tron_(Computerspiel)`

Zur Untersuchung der zweiten These, muss das *Widom of the Crowd*-Entwurfsmuster noch in die Spielregeln von GeoSnake integriert werden. Dazu muss für die zweite Regel im ursprünglichen Regelsatz eine entsprechende Veränderung vorgenommen werden - im Weiteren Regelvariante 2 genannt:

Regel 2: *Die Spieler sollen für einen Ort diejenige GPS Koordinate wählen, von der sie meinen, dass sie auch von den anderen Spielern von GeoSnake gewählt wurde.*

Entsprechend muss natürlich auch die Punktevergabe angepasst werden. Ein Ort gilt als erfolgreich abgelaufen, wenn die GPS Koordinate so gewählt wurde, dass sie nahe an den meisten anderen Koordinaten der übrigen Spieler liegt. Würde man also mit den Koordinaten der Spieler nach dem Spiel Cluster berechnen, bekommen alle Spieler Punkte, deren Koordinaten das größte Cluster bilden.

5.2.2. Vorstudie

In der Vorstudie spielten insgesamt 12 Teilnehmer *GeoSnake*, wobei von 11 der Fragebogen korrekt ausgefüllt wurde. Der Altersdurchschnitt lag bei 26 Jahren. Die Aufteilung nach Geschlecht betrug genau 50%, d.h. 6 männliche Teilnehmer und 6 weibliche. 8 der 12 Spieler hatten schon an anderen Studien mit ortsbezogenen Spielen teilgenommen.

Die Studie gliederte sich in zwei Teile - einem Teil der an einem stationären PC bearbeitet wurde und einem, der mobil durchgeführt wurde -, die in pseudo-zufälliger Reihenfolge (siehe Tabelle 5.4) von den Spielern bearbeitet wurden.

In einem Teil mussten die Spieler *GeoSnake* am Bildschirm mit Hilfe des Programms Google Earth [5] spielen. Als Spielregeln wurde hier Regelvariante 1 eingesetzt. In Anhang A.4.1 finden sich die genauen Anweisungen an die Spieler für diese Variante. Das Spielfeld umfasste 12 Orte, die von den Spielern per Maus virtuell abgelaufen werden mussten - Abbildung A.7 in Anhang A.4.3 zeigt die geographische Konstellation dieser Orte.

Die Teilnehmer spielten im anderen Teil des Experiments *GeoSnake* in der ortsbezogenen Variante Dazu bekamen die Spieler ein Mobiltelefon (Nokia 6630) - mit der vorinstallierten Spielesoftware -, welches mit einem Bluetooth GPS Gerät (XAIOX Wonde-XL) verbunden war. Die Aufgabe bestand

[5] `http://earth.google.de/`

ebenfalls darin 12 Orte im Spielfeld abzulaufen, nur diesmal physisch. Dabei sollte bei den ersten sechs Orten Regelvariante 1 verwendet werden und ab dem siebten Ort auf Regelvariante 2 (*Wisdom of the Crowd*-Variante) gewechselt werden. Die Reihenfolge, in dem die Orte abgelaufen wurden, war dabei den Spielern selbst überlassen. Wie Tabelle 5.4 verdeutlicht, wurde allerdings von Teilnehmer zu Teilnehmer die Regelvariante, mit der angefangen wurde, gewechselt - nach dem 4. wurde die Reihenfolge wiederholt. Für diesen Teil der Studie wurden andere Orte verwendet, als für *GeoSnake* am PC-Bildschirm. Die geographische Konstellation ist in Abbildung A.6 im Anhang A.4.2 abgebildet.

Regelvariante		
1S	1V	2V
1S	2V	1V
2V	1V	1S
1V	2V	1S

Tabelle 5.4.: Pseudo-zufällige Reihenfolge der Experimentteile und Regelvarianten (1S = Regelvariante 1 am Bildschirm, 1V und 2V = Regelvariante 1 und 2 mit Mobiltelefon)

Obwohl der Versuchsaufbau nicht vorgab, welchen Ort die Teilnehmer mit welcher Regelvariante bearbeiten sollten, zeigt Tabelle 5.5, dass die pseudozufällige Reihenfolge dazu geführt hat, dass alle Orte ungefähr gleich oft mit einer der beiden Varianten bearbeitet wurden. Leider machten Fehler in der Hardware, bzw. schlechte Empfangsbedingungen an den ausgewählten Orten, viele GPS Koordinatensätze nutzlos für eine nachfolgende Auswertung. Die Zahlen in Klammern geben die Gesamtzahl an aufgenommenen Datensätzen an und die Zahl davor die fehlerlosen GPS Koordinatensätze für diesen Ort.

Zusätzlich war noch angedacht, das *Wisdom of the Crowds*-Entwurfsmuster auf textuelle Beschreibungen der Orte in *GeoSnake* anzuwenden. Das würde Bedeuten, dass die Spieler bei einer Spielvariante, welches das Entwurfsmuster in dieser Hinsicht umsetzt, weniger unterschiedliche Beschreibungen zu einem Ort abgeben. Der Regelsatz für das Spiel GeoSnake müsste deshalb um eine zusätzliche Regel erweitert werden:

- Um einen Ort erfolgreich abzulaufen, müssen die Spieler dort nicht nur eine GPS Koordinate aufnehmen, sondern sollen zusätzlich den Ort mit zwei Stichworten beschreiben.

Ort	Regelvariante 1	Regelvariante 2
Otto-Friedrich Universität	6	6
Obere Pfarre	6(7)	3(5)
St. Stephans Kirche	5(6)	1(6)
Sandstraße	4(6)	5(6)
Altes Rathaus	6	6
Zentraler Omnibusbahnhof (ZOB)	6	6
Maxplatz	3(5)	7
Am Kranen	7	4(5)
Gabelmann	4(5)	7
Geyerswörth	5	5(7)
Bamberger Dom	6(7)	4(5)
Schönleinsplatz	6	6

Tabelle 5.5.: Häufigkeit der besuchten Orte bezüglich der Regelvarianten; die Zahl in Klammern gibt die maximale Anzahl an abgegebenen GPS Koordinaten an und die Zahl vor der Klammer gibt die Anzahl an fehlerlosen GPS Koordinaten an.

Die Vorstudie war ursprünglich dazu gedacht den allgemeinen Versuchsaufbau zu testen, brachte aber einige andere interessante Erkenntnisse hervor, die ich in Abschnitt 5.2.5 bespreche, unter anderem auch in Bezug auf die textuellen Beschreibungen.

5.2.3. Hauptstudie

An der Haupstudie nahmen insgesamt 15 Probanten teil, wovon 12 den Fragebogen korrekt ausgefüllt haben. Der Altersdurchschnitt lag hier ebenfalls bei 26 Jahren. Es nahmen insgesamt 5 männliche und 10 weibliche Spieler teil. Der Versuchsaufbau wurde im Vergleich zur Vorstudie leicht angepasst.

Diesmal wurde die Studie in drei Teile aufgeteilt, die von den Teilnehmer bearbeitet werden mussten, einen am PC und zwei weitere mobil mit Mobiltelefon (Nokia 6630) und Bluetooth GPS Gerät (Columbus V-900). Um die Vergleichbarkeit der mit *GeoSnake* erhobenen Datensätze zu verbessern, wurde sowohl im PC-Teil als auch im mobilen Teil der Studie diesmal dieselben Orte verwendet. Die geographische Konstellation der Orte für den PC-Teil ist im Anhang A.4.4 in Abbildung A.9 abgebildet.

Wie schon in Abschnitt 5.2.2 kurz erläutert, war eines der Probleme in der Vorstudie das teilweise ungleiche Vorhandensein von gültigen GPS Koordinaten zu den einzelnen Orten. Dies lag zum einen an der verwendeten Hardware, den Bluetooth GPS Geräten, aber zum anderen auch daran, dass die Spieler die Reihenfolge der Abarbeitung zwischen allen 12 Orten frei wählen konnten.

Das rechte Beispiel in Abbildung 5.13 zeigt einen solchen Fall. Hier wurde für die Regelvariante 2 nur eine fehlerlose GPS Koordinate aufgezeichnet, wohingegen für Regelvariante 1 fünf vorhanden waren. Dadurch konnte kein sinnvoller Vergleich der Ergebnisse durchgeführt werden.

Der linke Fall in Abbildung 5.13 zeigt ein Problem, dass im Vorhinein nicht zu erwarten war: Bei Orten, die über keine nennenswerte geographische Ausdehnung verfügten, ist der Unterschied in den zwei Regelvarianten einzig von den GPS Empfangbedingungen abhängig. Da es hier nur eine „richtige“ Position für den Ort gibt, im Beispiel der Gabelmann in der Innenstadt von Bamberg, kann auch nur eine Entscheidung von den Spielern getroffen werden. Hingegen bei Plätzen, großen Gebäuden oder Straßen besteht eine gewisse Entscheidungsfreiheit für die Spieler darüber, welche GPS Koordinate sie wählen, um den Ort damit zu identifizieren.



Abbildung 5.13.: Beispiel für die Streuung der erhobenen GPS Koordinaten mit den Regelvarianten 1 (rote Markierung) und 2 (blaue Markierung) für die Orte „Geyerswörth“ und „St. Stephans Kirche“ in der Vorstudie zu GeoSnake, © Karte: Google Maps (`http://maps.google.de/`).

Aus den genannten Gründen wurden für den mobilen Teil in der Hauptstudie die Orte in zwei geographische Konstellationen aufgeteilt und von Or-

ten ohne nennenswerte Ausdehnung abgesehen. Dadurch wird sichergestellt, dass alle Orte gleich oft mit einer der beiden Regelvarianten von den Teilnehmer abgearbeitet werden. Die Teilnehmer spielten also *GeoSnake* zweimal in der mobilen Ausführung. Welcher Ortsverteilung mit welcher Regelvariante abgelaufen werden musste, wurde wiederum pseudo-zufällig ermittelt. Tabelle 5.6 gibt den zugehörigen Versuchsablauf wieder - nach dem 4. Teilnehmer wurde wiederum von vorne begonnen.

Regelvariante - Spielfeld		
1S	1V-V1	2V-V2
1S	2V-V1	1V-V2
1V-V1	2V-V2	1S
1V-V2	2V-V1	1S

Tabelle 5.6.: Pseudo-zufällige Reihenfolge der Experimentteile und Regelvarianten (1S = Regelvariante 1 am Bildschirm, 1V und 2V = Regelvariante 1 und 2 mit Mobiltelefon, V1 und V2 = Spielfeld 1 bzw. 2)

5.2.4. Variablen und Evaluierungsmodelle

Um die im Abschnitt 5.2.1 formulierten Hypothesen zu überprüfen, nehme ich als abhängige Variable die Streuung der GPS Koordinaten an, also die Abweichung der in *GeoSnake* erhobenen GPS Koordinate zu der für einen Ort gemittelten GPS Koordinate. Unabhängige Variablen sind die Regelsätze. Dabei unterscheide ich zwischen der Regelvariante 1 und Regelvariante 2 im mobilen Teil (*rule 1* und *rule 2* im Weiteren) und dem PC-Teil mit Regelvariante 1 (im Folgenden *rule 3*). Als weitere unabhängige Variablen sind noch Lerneffekte, d.h. die zeitliche Reihenfolge der angelaufenen Orte und das Vorwissen der Teilnehmer zu nennen.

Um die Auswirkungen der verwendeten Regelvarianten, bzw. Datenerhebungsmethoden, zu evaluieren, wurde folgendes generalisierte lineare Modell [218] mit Hilfe des Programms R [219] ausgewertet - für den kompletten Programmcode siehe Anhang A.6:

$$\mu = E(distance|rule, time) = h\left(\beta_0 + \beta_1 rule_1 + \beta_2 rule_2 + \beta_3 time\right) \tag{5.4}$$

Die Zielvariable ist die Euklidische Distanz ($distance$) der erhobenen GPS Koordinaten zum jeweiligen Mittelpunkt, der zu einem Ort insgesamt aufgenommen GPS Koordinaten. Diese soll idealerweise einen positiven Effekt haben. Um dies zu erreichen, nehmen wir als vermutete Verteilung der Distanzwerte eine Gammaverteilung an - dasselbe Modell habe ich auch mit einer Normalverteilung gerechnet. Die genutzte Linkfunktion ist eine kanonische (invers) - die Varianz des Fehlers wird als konstant angenommen. Im obigen Modell wird die Referenzgruppe für die kategorische Behandlung der *Regel* ($rule$) mit der Regelvariante 1 am Bildschirm (*rule 3*) angenommen (für die Überprüfung der zweiten Hypothese aus Abschnitt 5.2.1 wird die Referenzgruppe auf *rule 1* gelegt, d.h. Regelvariante 1 im mobilen Teil). Da ein Lerneffekt möglich erscheint, wird die Zeit ($time$) als Kontrollvariable zum Modell hinzugefügt.

5.2.5. Ergebnisse und Diskussion

Bevor ich auf die Evaluierung meiner in Abschnitt 5.2.1 aufgestellten Hypothesen zu sprechen komme, möchte ich noch auf drei Beispiele eingehen, die die Fragestellung, bzw. das Ergebnis, verdeutlichen sollen.

Abbildung 5.14 zeigt die erhobenen GPS Koordinaten für drei unterschiedliche Ortstypen. Im oberen Screenshot (entnommen aus der Haupstudie zu *GeoSnake*) wurde ein 1-D-dimensionales Merkmal, die „Kleberstraße“, als ein Ort im Spiel ausgewählt. Wie man hier sieht, führt die Erhebung nach dem *Wisdom of the Crowds*-Entwurfsmuster - Regelvariante 2 (blaue Markierungen) - zu einer sichtbar geringeren Streuung der GPS Koordinaten als bei der Erhebung am PC (gelbe Markierung), sowie vor allem bei der Erhebung nach der Regelvariante 1 (rote Markierung). Gleiches gilt für das ausgedehnte 0-D-dimensionale Merkmal „Honer“ (ein Einkaufsgebäude mit zwei geographisch weit auseinander liegenden Eingängen - ebenfalls aus der Hauptstudie) und das 2-D dimensionale Merkmal „Geyerswörth“, diesmal aus der Vorstudie (eine unscharf definierte Region).

Das Auffordern der Spieler über die Spielregeln sich dem Verhalten der anderen Spieler anzupassen (Regelvariante 2), führt dazu - dies bestätigte sich auch in Interviews mit den jeweiligen Spielern -, dass diese in den allermeisten Fällen, eine besondere Landmarke innerhalb des Merkmals als Repräsentation wählen. Bei der „Kleberstraße“ war es ein Eiscafe, in „Geyerswörth“ die Touristeninformation und beim „Honer“ war es der Haupteingang.

Sind die Spieler völlig frei in der Wahl ihrer Repräsentation, wie bei der Regelvariante 1, wird meist der Weg des kürzesten Widerstandes gegangen.



Abbildung 5.14.: Beispiel für die Streuung der erhobenen GPS Koordinaten mit den Regelvarianten 1 (rote Markierung) und 2 (blaue Markierung), sowie der Markierungen (gelb), die im PC-Teil gesetzt wurden, für die Orte „Kleberstraße“ (oben), „Geyerswörth“ (unten links) und „Honer“ (unten rechts), © Karte: Google Maps (`http://maps.google.de/`).

D.h. bei Straßen wird gleich am Beginn der Straße die GPS Koordinate aufgenommen und bei Plätzen und Gebäuden wird einfach die erste Möglichkeit wahrgenommen, die im Ablauf des Spieles erreicht wird.

Wie zu erwarten war, lieferte die Vorstudie nur bedingt gute Ergebnisse. Tabelle 5.7 und 5.8 fassen die Resultate zusammen. Immerhin konnte die zweite Hypothese mit einem signifikanten Ergebnis bestätigt werden, wenn auch mit einem niedrigen Signifikanzniveau von 0.1. Der Vergleich der Regelvarianten 1 und 2 zur PC-Variante (Regel 3) brachte kein signifikantes Ergebnis. Auch war hier die mittlere Streuung geringer als bei der Erhebung mit der mobilen Hardware, wobei hier die vielen fehlerhaften Daten eine Rolle gespielt haben.

Für die Hauptstudie sieht die Evaluierung folgendermaßen aus:

Tabelle 5.9 zeigt das Ergebnis zur Untersuchung der ersten Hypothese aus Abschnitt 5.2.1. Zu Beachten ist, dass in R [219] das GML mit einem kano-

	Koeffizent	Std. Fehler
rule 3 (Intercept)	3.973e-02***	1.113e-02
rule 1	1.212e-03	9.633e-03
rule 2	-1.010e-02	0.003
time	2.736e-05	6.376e-04

Tabelle 5.7.: Quelle: GeoSnake Spiel 1. Anmerkung: *** <0; ** <0.05; * <0.1

	Koeffizent	Std. Fehler
rule 1 (Intercept)	2.964e-02	5.468e-03
rule 2	1.131e-02*	6.271e-03
time	2.736e-05	6.376e-04

Tabelle 5.8.: Quelle: GeoSnake Spiel 1. Anmerkung: *** <0; ** <0.05; * <0.1

nischen Link bei einer Gammaverteilung die Richtung des Effekts der Zielvariable andersherum interpretiert werden muss. Somit ist *Regel 2* signifikant besser als *Regel 3*, aber *Regel 1* nicht signifikant besser als *Regel 3*, obwohl die mittlere Streuung bei *Regel 1* geringer ist. Auch ist kein signifikanter Lerneffekt (*time*) festzustellen. Die erste Hypothese kann also nur zu einem Teil - für die Regelvariante 2 (*rule 2*) - verifiziert werden.

	Koeffizent	Std. Fehler
rule 3 (Intercept)	0.016***	0.004
rule 1	-0.002	0.004
rule 2	0.006*	0.003
time	-0,0002	0.0002

Tabelle 5.9.: Quelle: GeoSnake Spiel 2. Anmerkung: *** <0; ** <0.05; * <0.1

Tabelle 5.10 zeigt das Ergebnis der Untersuchung zur zweiten Hypothese aus Abschnitt 5.2.1. *Regel 2* ist signifikant besser als *Regel 1*, womit die zweite Hypothese verifiziert werden kann. Allerdings muss man hier beachten, dass keine zufälligen Effekte mehr in die Berechnung mit einfließen, wobei aber bei einer Normalverteilung derselbe Effekt zu beobachten ist - mit der Betrachtung der zufälligen Effekte *intercept* und *time*.

	Koeffizent	Std. Fehler
rule 1 (Intercept)	0.014***	0.002
rule 2	0.008**	0.003
time	-0,0002	0.0002

Tabelle 5.10.: Quelle: GeoSnake Spiel 2. Anmerkung: *** <0; ** <0.05; * <0.1

5.3. Studie zum Integrationsprozess: SVAC-Algorithmus

Mein in Kapitel 4, Abschnitt 4.2, vorgestellter *SVAC*-Algorithmus bietet die Möglichkeit die bisher vernachlässigte Kontextinformation in Form der *semantischen Barrieren* bei der Datenintegration auszunutzen.

Ein anderes Hindernis ist die unzureichende Verfügbarkeit von Datenbeständen, die einen Algorithmus wie *SVAC* möglich machen. 0-D-dimensionale raumbezogene Datensätze, wie sie in Datensammlungen z.B. von ortsbezogenen Diensten vorkommen, bestehen eben nur aus einem Koordinatenpaar, einer (oder mehrerer) Ausprägung einer semantischen Kategorie und einem Namen als Identifikator (siehe auch [25]). Mit der Nutzung von ortsbezogenen Spielen wie *CityExplorer* ändert sich dieser Sachverhalt aber. Durch einen entsprechenden Regelsatz kann jede Art von raumbezogenen Daten gesammelt werden, je nachdem, wie sie in einem nachfolgenden Aggregationsalgorithmus verwendet werden sollen.

Beim Design von CityExplorer wurden die Anforderungen des *SVAC*-Algorithmus an die Datengrundlage mit berücksichtigt. Die erhobene Datensammlung wird nun im Folgenden verwendet, um den Algorithmus auf seine Güte zu überprüfen.

5.3.1. Versuchsaufbau: Datengrundlage und Vergleichsalgorithmus

Die Datensammlung aus *CityExplorer* umfasst insgesamt 839 Spielsteine, wenn man die in Abschnitt 5.1.1 vorgestellten Studien zum nicht-verteilten Spiel für die Innenstadt Bamberg plus die Datensätze aus dem verteilten Spiel zusammen nimmt, die in Bamberg erhoben wurden. Aus diesem Datensatz wurden manuell Referenzcluster gebildet. Dieser „Goldene Standard“ [220]

dient als Bewertungsbasis für jeden auf diesen Datensatz angewendeten Aggregationslagorithmus.

Um das manuelle Clustern etwas zu vereinfachen, wurden folgende Kategorien, bzw. die dazu gehörenden Datensätze, nicht mit betrachtet: *tree*, *bench*, *manhole cover* und *Tore*. Ohne diese Datensätze besteht die Datensammlung noch aus 753 Spielsteinen. Außerdem wurden Kategorien zusammengefasst, die auf deutsch und auf englisch den identischen Bezeichner tragen, wie z.B. „Ärzte“ und „doctor“. Die genaue Aufteilung der Instanzen in den Kategorien findet sich in Tabelle A.10 im Anhang.

Auf Grundlage dieser Häufigkeitsverteilung habe ich auch die semantische Ähnlichkeit zwischen den Kategorien nach [8] berechnet, wie in Kapitel 4 Abschnitt 4.2.2 beschrieben. Als Ontologie habe ich wie Resnik WordNet ([174] und [175]) eingesetzt. Die Tabellen A.11 bis A.17 geben die berechneten Werte wieder.

Um einen weiteren Vergleich, neben dem manuell erstellten „Goldenen Standard“ [220], zu haben, habe ich noch die Variante des k-means Algorithmus nach Ashbrook und Starner (2003) [3] auf die Datensammlung angewendet. Die semantischen Daten in Form der Kategorien werden, wie bei diesen Aggregationsansätzen üblich, als Filter angewendet. D.h. nur Datensätze von derselben Kategorie werden jeweils betrachtet.

5.3.2. Ergebnisse und Diskussion

Zum Vergleich der zwei Aggregationsalgorithmen habe ich für beide jeweils drei Kennzahlen berechnet, den *Precision* und *Recall*, sowie den *F-Score* nach [221].

Die Precision-/Recall-Werte wurden dabei zuerst für jedes Cluster i einzeln berechnet:

R_i sei das i-te Cluster in der Referenzclusterung (ein manuell erstelltes Cluster im oben erwähnten „Goldenen Standard“) mit r_i GPS Koordinatenpaare. K_i sei das i-te Cluster im k-means Clustering mit k_i GPS Koordinatenpaare. Entsprechend ist V_i das i-te Cluster im *SVAC* Clustering mit v_i GPS Koordinatenpaare.

$$precision_{ki} = \frac{K_i \cap R_i}{K_i} \tag{5.5}$$

$$precision_{vi} = \frac{V_i \cap R_i}{V_i} \tag{5.6}$$

Der Recall-Wert berechnet sich dann entsprechend.

$$recall_{ki} = \frac{K_i \cap R_i}{R_i} \tag{5.7}$$

$$recall_{vi} = \frac{V_i \cap R_i}{R_i} \tag{5.8}$$

Um nun den globalen Precision-/Recall-Wert zu bekommen, nehme ich die gemittelte Summe der Einzelwerte, wobei $count$ gleich der Anzahl der Cluster R_i ist.

$$Precision = \frac{\sum_{i=1}^{count} precision_{ki/vi}}{count} \tag{5.9}$$

$$Recall = \frac{\sum_{i=1}^{count} recall_{ki/vi}}{count} \tag{5.10}$$

Der *F-Score*-Wert berechnet sich daraufhin wie folgt (entsprechend [221] allerdings mit Hilfe der Formel für den $F_\beta - Score$).

$$F_{R_i,K_i} = \frac{(1+\beta^2) * precision_{ki} * recall_{ki}}{\beta^2 * precision_{ki} + recall_{ki}} \tag{5.11}$$

Nun wird der beste F-Score zu einem Referenzcluster R_i genommen.

$$F_{R_i} = max_{i=1,\ldots,count} F_{R_i,K_i} \tag{5.12}$$

Der globale *F-Score*-Wert ist das gewichtete Mittel aus den ausgewählten besten $F - Score$-Werten für die Cluster aus R.

$$\textit{F-Score} = \sum_{i=1,\ldots,count} \frac{R_i}{count} * F_{R_i} \tag{5.13}$$

Für die Studie wurde folgende Parameterwerte festgelegt:

1. Der Schwellwert für die semantische Ähnlichkeit zur Beschneidung der Sichtbereiche im ersten Teil des *SVAC*-Algorithmus (siehe Kapitel 4 Abbildung 4.2.2) wird auf 0.1 gesetzt.

2. Der Radius für den k-means Algorithmus nach [3] wird auf 80m gesetzt.

3. Die Variable β zur Berechnung des $F - Score$ erhält den Wert 3.

Die aus Matyas et al. (2009) [222] entnommene Tabelle zeigt die berechneten Ergebnisse für die vorgestellte Studie. Anzumerken ist hier, dass die verwendete Datensammlung nur einen einzigen Fall enthält, wo die Besonderheit einer *semantischen Barriere* auftritt, d.h. wo zwei semantisch ähnliche und geographisch nahe Spielsteine nicht demselben Real-Welt Merkmal entsprechen und ein dritter Spielstein dazwischen liegt, der die Zusammenführung der zwei verhindert.

Trotzdem zeigen die Ergebnisse ein gewisses Potential für den Aggregationsalgorithmus *SVAC*, siehe Tabelle 5.11. Gegenüber dem k-means Algorithmus ist eine Verbesserung des Precision-Wertes um 0.05 Punkte zu beobachten. Allerdings sinkt der Recall-Wert auch gleichzeitig um 0.1 Punkt, für *SVAC* mit der Schnittoperation. Wird die Vereinigungsoperation beim *SVAC* eingesetzt steigt die Precision nur um 0.04 Punkte aber dafür sinkt der Recall nur um 0.8 Punkte. Der $F - Score$ ist bei beiden Varianten besser als beim k-means, wobei bei der Vereinigungsoperation nur ein marginaler Anstieg von 0.01 Punkten zu beobachten ist.

Algorithmus	Precision	Recall	F_β
k-means	0.85	0.82	0.80
svac_catsim (Schnitt)	0.90	0.72	0.85
svac_catsim (Vereinigung)	0.89	0.74	0.81

Tabelle 5.11.: Clustervergleich Ergebnisse

Die Erklärung beider Phänomene ist, beim Nichtvorhandensein des Phänomens der *semantischen Barriere* relativ eindeutig. Durch die Sichtbereiche kann *SVAC* Cluster, die nur aus einem einzigen Spielstein bestehen, besser erkennen. Abbildung 4.4 zeigt eine häufig in der *CityExplorer* Datensammlung anzutreffende geographische Konstellation. Allerdings ist bei der Verwendung von Sichtbereichen das korrekte Erkennen von größeren Clustern schwieriger, was deshalb den Recall-Wert verringert. Dies erklärt auch den niedrigeren Recall-Wert bei der Schnittoperation gegenüber der Vereingungsoperation. Beim Schnitt werden häufiger, je nach gewähltem Schwellwert für den Ähnlichkeitswert, größere Cluster in mehrere kleinere zerteilt. Auch spielt die Form und die Größe der Sichtbereiche eine Rolle bei diesen Werten.

5.4. Ausblick: Simulationsumgebung

Der semantische Integrationsprozess, wie ich ihn in Kapitel 4 Abschnitt 4.1 vorgestellt habe, ist nicht auf eine Art der Vorverarbeitung und/oder Aggregationsalgorithmus festgelegt. In meiner Arbeit habe ich einen neuartigen Aggrgegationsalgorithmus vorgestellt, der mit der in den *CityExplorer* Spielen erhobenen Datensammlung, evaluiert werden konnte.

Um nun anderen Wissenschaftlern die Möglichkeit zu gegeben, erstens ihre schon vorhandenen oder zukünftigen Algorithmen auf meine Datensammlung anzuwenden und zweitens ihre Algorithmen miteinander zu vergleichen, habe ich prototypenhaft eine web-basierte Simulationsumgebung umgesetzt, die dies möglich machen soll.

Die aktuelle Implementierung unterstützt folgende Basisfunktionalitäten:

- Das Einbinden von Aggregationsalgorithmen in Form von JAR-Dateien.
- Das automatische Erkennen von notwendigen Parametern zur Durchführung der Algorithmen.
- Die Visualisierung der Aggregationsergebnisse - auf Grundlage von OpenStreetMaps-Kartendaten.
- Die Speicherung der Ergebnisse in einer PostgreSQL/PostGIS Datenbank zur weiteren Bearbeitung.

Technisch gesehen basiert die Implementierung auf dem Apache Struts 2-Framework [6], welches auf einem Tomcat-Server läuft.

Abbildung 5.15 zeigt die Integration des Aggregationsalgorithmus nach [7] in die Simulationsumgebung. Auf dem rechten Screenshot ist zu sehen, wie die Parameter *rTRESH* und *cTRESH* automatisch aus der JAR-Datei des Algorithmus erkannt wurden. Auf dem linken Screenshot sieht man die Anwendung des Algorithmus und die von der Umgebung zur Verfügung gestellten Visualisierung. Die beiden zusammengeführten GPS-Tracks wurden aus den Log-Dateien der *CityExplorer* Spielen extrahiert (siehe Abschnitt 5.1.5).

Damit die Simulationsumgebung den gesamten semantischen Integrationsprozess unterstützt, müssen noch folgende Funktionalitäten umgesetzt werden:

- Allgemeine Formalisierung und Implementierung eines Datenobjektmodells, welches nicht nur 1-D-dimensionale raumbezogene Daten (GPS Tracks) unterstützt, sondern auch 0-D- und 2-D-dimensionale.

[6] `http://struts.apache.org/2.1.8.1/index.html`, Version 2.0.11.2

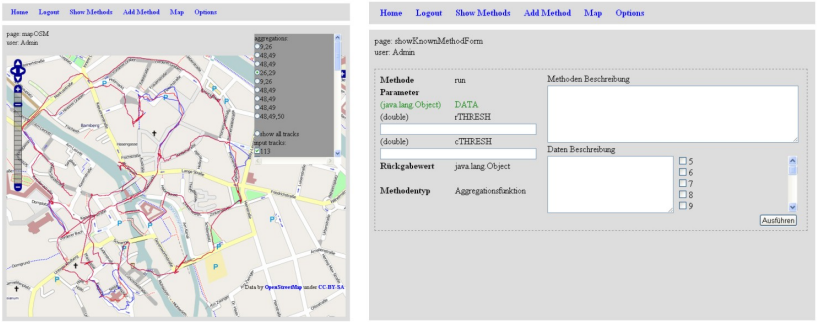


Abbildung 5.15.: Simulationsumgebung für den Semantischen Integrationsprozess - Beispielhafte Einbindung des Aggregationsalgorithmus nach [7]

- Die Unterstützung eines XML-basierten Austausch-/Import-, bzw. Exportformates, für alle in den ortsbezogenen Spielen erhobenen raumbezogenen Daten.
- Eine manuell einstellbare, automatische Vorverarbeitung.
- Eine entsprechende Auswertungsfunktionalität, die über die reine Visualisierung des Integrationsergebnisses hinausgeht. Diese könnte beispielsweise auf denselben Kennzahlen beruhen, die auch für den abschließenden Schritt im semantischen Integrationsprozess (siehe Abbildung 4.8 von mir angesprochen wurden.

6. Zusammenfassung und Ausblick

Am Anfang meiner Dissertation stand die Idee den Ansatz von Ahn und Dabbish (2004) [19] auf Fragestellungen aus der GIScience anzuwenden [14]. Für die Frage, wie man geowissenschaftliche Laien dazu bringt, raumbezogene Daten, oder in diesem Zusammenhang ihr lokales Wissen, freiwillig Pries zu geben, gibt meine Arbeit damit eine mögliche Antwort. Mein Design Framework für ortsbezogene Spiele - mit dem damit entworfenen Prototyp *CityExplorer* - ermöglicht es, diese so zu entwerfen, dass raumbezogene Daten spielerisch und qualitätsgesichert von Spielergemeinschaften erhoben werden. Die Einbindung in einen spielerischen Kontext ermöglicht es auch, Daten von solcher Struktur erheben zu lassen, dass damit bisher ungelöste Problemstellungen bei einer nachgeschalteten Datenintegration angegangen werden können. Dies habe ich mit der Präsentation eines neuen Aggrgegationsalgoithmus gezeigt, der das Problem der *semantischen Barrieren* löst, welches ich in meiner Arbeit identifiziert habe. Damit wird eine semantische Integration der erhobenen Daten in eine dadurch nochmals qualitätsgesicherte Datensammlung möglich.

6.1. Ergebnisse

Um diesen Weg der raumbezogenen Daten von der gemeinschaftlichen Erhebung bis hin zur Integration in einen qualitätsgesicherten Datenbestand, abzubilden, habe ich folgende Teilprobleme gelöst:

Design Framework zur spielerischen Erhebung raumbezogener Daten

Ausgehend von einer Analyse schon entwickelter Spiele, die im geographischen Raum - sprich der realen Welt - gespielt werden (Abschnitt 2.3), habe ich in Kapitel 3 ein Design Framework präsentiert, welches die Entwicklung ortsbezogener Spiele speziell zur Datenerhebung erlaubt. Dies wird durch die

Einbindung eines oder mehrerer der von mir definierten Entwurfsmuster entsprechender Spielelemente erreicht. Für eine integrierte Qualitätssicherung der Daten, habe ich ebenfalls Entwurfsmuster vorgeschlagen.

Prototypische Umsetzung eines ortsbezogenen Spieles

Das Design Framework wurde mit der Entwicklung des ortsbezogenen Spieles *CityExplorer* auf seine Vollständigkeit und Praxistauglichkeit überprüft (Abschnitt3.8). Mehrere Nutzerstudien mit *CityExplorer*, die ich in Abschnitt 5.1 beschrieben habe, haben gezeigt, dass die aus den Entwurfsmustern resultierenden Spielelemente den Spielspaß und -fluss nicht beeinträchtigen.

Spielinterne Qualitätssicherung

Die im Design Framework vorgeschlagenen Entwurfsmuster zur Qualitätssicherung wurden erstens durch eine Anylse der Datensammlung, der mit *CityExplorer* erhobenen raumbezogenen Daten, überprüft. Die Ergebnisse zeigen, dass die Nutzer Fehler in der Größenordnung des durchschnittlichen GPS-Fehlers produzieren, was für geowissenschaftliche Laien bei der Verwendung von handelsüblicher Hardware positiv zu bewerten ist. Mit einer Bearbeitungsquote von an die 40% zur Qualitätsverbesserung, liegt mein Ansatz im Bereich anderer Ansätze, wie ich in Abschnitt 5.1.4 berichtet habe. Somit kann die Quote bisheriger Web 2.0 Ansätze wie OpenStreetMaps deutlich gesteigert werden - wie in Kapitel 1 gezeigt liegt die Quote dort bei ca. 10%.

Die Durchführung und Analyse einer weiteren Nutzerstudie mit einem weiteren von mir entworfenen ortsbezogenen Spiel, *GeoSnake*, hat gezeigt, dass vor allem das *Wisdom of the Crowd*-Entwurfsmuster geeignet ist, die Erhebung von qualitativ hochwertige raumbezogene Daten zu ermöglichen. Ohne den Einsatz des Entwurfsmusters ist die Genauigkeit der erhobenen raumbezogenen Daten signifikant geringer als mit diesem. Ebenso liefert das Entwurfsmuster bessere Daten als ein rein web-basierter Ansatz, wie er in Spielen aus anderen Forschungsbereichen vorgeschlagen wird, z.B. dem Semantic Web (Abschnitt 2.1).

Semantische Barrieren

Ich führe in Kapitel 4 Abschnitt 4.1.2 den Begriff der *semantischen Barrie-*

re ein. Dieser beschreibt ein bisher in der Literatur nicht behandeltes Problem - was ich in Abschnitt 2.4 verdeutliche -, welches bei der Aggregation von raumbezogenen Daten auftreten kann: Zwei raumbezogene Datensätze, die geographisch nah und semantisch ähnlich sind, sollen nicht zusammengeführt werden, da geographisch zwischen ihnen ein dritter Datensatz liegt, der aber semantisch unähnlich zu den beiden ist. Würde man die beiden nun aggregieren, würde sich das Ergebnis mit dem anderen Datensatz semantisch in Konflikt stehen. Daher soll dieser mittlere Datensatz eine *semantische Barriere* zwischen den beiden anderen aufbauen. Die Spielregeln von *CityExplorer* erlauben es nun raumbezoegene Daten mit einer Strukturierung zu erheben, die es mir möglich macht, eine Lösung für dieses Problem vorzuschlagen.

Semantischer Integrationsprozess

Angelehnt an die Arbeit von Walter und Fritsch (1999) [169] stelle ich in Kapitel 4 einen allgemeinen semantischen Integrationsprozess vor. Für den Schritt der Aggregation redundanter raumbezogener Datensätze in einer Datensammlung schlage ich einen neuartigen Algorithmus vor, den *Semantic View Area Clustering*-Algorithmus (*SVAC*). Der Algorithmus nutzt die spezielle Datenstruktur der mit *CityExplorer* erhobenen Daten aus, um das Problem der *semantischen Barrieren* zu lösen. Die Idee den Kontext in der Berechnung der Ähnlichkeit mit einzubeziehen, lässt sich aber auch auf anders strukturierte Daten übertragen. Eine Evaluierung von *SVAC* zeigt, dass er eine bessere Genauigkeit hat, als das bekannte Clusterverfahren nach Ashbrook und Starner (2003) [3] - siehe Abschnitt 5.3.1.

6.2. Ausblick

Neben den in Kapitel 4 und 5 angesprochenen Arbeiten, die in Rahmen meiner Dissertation weiter angegangen werden könnten, möchte ich nachfolgend noch weitere interessante Fragestellungen oder Aspekte aus dem Kontext meiner Arbeit vorstellen, die meiner Meinung nach eine weitere Betrachtung wert sind.

6.2.1. Spielerische qualitätsgesicherte Datenerhebung

Ortsbezogene Spiele wurden für eine wissenschaftliche Betrachtung erst mit der Abschaltung der *Selective Availability* (SA) des GPS Signals am 02. Mai

2000 interessant. Obwohl GPS nicht die einzig zur Verfügung stehende Lokalisierungstechnologie ist, siehe beispielsweise Küpper (2005) [223] für einen Überblick, so ist sie doch die am einfachsten zu nutzende und diejenige, mit der besten Verfügbarkeit. Wird GPS eingesetzt, beschränkt sich damit natürlich das Spielgeschehen auf die geographische Umgebung außerhalb von Gebäuden oder sonstigen Orten ohne freie Sicht zum Himmel. Wie Kiefer et al. (2005) in ihrer Arbeit [4] diskutieren, kommen durch die Verlagerung des Spielgeschehens in die reale Welt dynamische Elemente (*dynamic conditions* [4]) ins Spiel, auf die während des Spieles kein bis nur einen geringen Einfluss genommen werden kann. Das dynamische Element mit dem größten Einfluss ist meiner Meinung nach das Wetter, wie folgender Kommentar eines (*CityExplorer*) Spielers aus Shonandai (Japan, siehe Abschnitt 5.1.3) zeigt:

> I would want to play the game under a good weather. (translator´s remark: there was typhoon coming...)

Natürlich sind die Auswirkungen des Wetters auf Spiele, die nur an einem Ort stattfinden, weniger gravierend, als bei verteilten Spielen. Sie finden dann einfach nicht statt. Aber gerade, wenn diese neue Art des Spielens den Schritt vom Forschungsgegenstand in die Hände des normalen Spielers finden soll, ist dies eine nicht zu unterschätzende Herausforderung an das Design solcher Spiele. Insbesondere auch, wenn das entworfene Spiel nicht nur in den Sommermonaten raumbezogene Daten liefern soll, sondern auch darüber hinaus. Hier sehe ich vor allem die Verbindung von ortsbezogenen und web-basierten Spielelementen als viel versprechende Möglichkeit, die ich in ihren Grundzügen schon in *CityExplorer* eingesetzt habe.

In *CityExplorer* sind bestimmte Spielaktionen nur web-basiert möglich und andere nur vor Ort mit dem mobilen Gerät. Hier ist die Frage, ob und welche Art von Spielmechanismen, -elementen, sowohl vor Ort als auch web-basiert vom Spieler ausgeführt werden können, ohne das sich das Spielerlebnis ändert? In event-basierten Spielen wie *Can You See Me Now?* [102] oder *Uncle Roy All Around You* [123] wird schon ein Mix aus ortsbezogenem und web-basiertem Spielelementen eingesetzt, allerdings werden dort die Rollen der offline und online Spieler getrennt. D.h. man kann entweder das Spiel aus der einen oder aus der anderen Perspektive erleben, aber bisher wurde noch nicht versucht es dem Spieler zu erlauben, nahtlos zwischen diesen zwei Ansichten zu wechseln.

Ein damit verbundener Aspekt findet sich in den Forschungsarbeiten zum asynchronen Spielen - siehe z.B. [127], [224], [225] und [119]. Die Möglichkeit das Spiel jederzeit unterbrechen und wieder fortsetzen zu können - und

damit in den normalen Tagesablauf zu integrieren -, ist eine wichtige Voraussetzung, ortsbezogene Spiele zur Datenerhebung auch in der Praxis zu etablieren. Um diesen Spielmechanismus allgemein greifen zu können, bedarf es aber noch weiterführender Forschung.

Neben einer Umsetzung des *Widom of the Crowds*-Entwurfsmusters in ein größer angelegtes ortsbezogenes Spiel, ist die Verbindung von solchen Spielen mit den Nutzergemeinschaften aus sozialen Netzwerken eine weitere interessante Richtung, in die man zukünftig gehen kann. Arbeiten wie [68] diskutieren schon, wie die Beziehungen in sozialen Netzwerken in Modelle eingebunden werden können, die vor allem Aussagen über die Qualität von ortsbezogener Information machen. So berechnete Trust-Werte könnten beispielsweise auch im *Peer-Reviewing*-Entwurfsmuster zum Einsatz kommen, um die einzelnen Bewertungen mit einem Vertrauensmaß zu gewichten.

6.2.2. Semantischer Integrationsprozess

Anderson (2003) hat in seiner Übersicht zur Künstlicher Intelligenz [226] den Ansatz des Computerspieles *The Sims* [1] vorgestellt. Die Idee dabei ist, nicht wie in anderen Computerspielen, die computergesteuerten Charaktere (Nonplayer characters, NPC) selbst Intelligent zu machen, sondern die Umwelt, in der diese Charaktere agieren. Die Objekte in der Spielumwelt werden dazu „annotiert“ [227], d.h. sie halten alle Information, die zu ihnen gehört, sowie auch welche Interaktionen mit ihnen möglich sind. Der virtuelle Charakter muss dann selbst kein Wissen mehr über die Umwelt haben, sondern kann einfach bei jedem Objekt „nachfragen“. Dies ermöglicht die Implementierung einer relativ simplen Künstlichen Intelligenz in einer intelligenten Umwelt.

Meiner Meinung ist die Idee der „Intelligenten Objekte“ auch für andere Forschungsgebiete interessant. In meiner Arbeit hat die Spielregel, wie die Spieler ihre virtuellen Spielsteine setzen müssen, dazu geführt, komplexe raumbezogene Datensätze zu generieren. Die Komplexität der mrk - siehe Abschnitt 4.1.1 - ermöglichte mir dann einen relativ einfachen Aggregationsalgorithmus im Verlauf des semantischen Integrationsprozess umzusetzen. Die Frage, die man hier beantworten muss, ist, welche Informationen ein Datensatz für einen entsprechenden Algorithmus aufweisen muss. Wird ein spielerischer Ansatz zur Erhebung dieser Information angewendet, muss dafür eine entsprechende Spielregel entworfen werden. Welche Spielregeln für welche Arten von Informationen geeignet sind, ist eine ebenfalls noch

[1] `http://thesims.ea.com/`, Erstveröffentlichung 2000

unbeantwortete Frage.

Man könnte aber auch noch einen Schritt weiter gehen. Das *Input-Output*-Entwurfsmuster bietet die Möglichkeit die Clusterbildung nicht automatisch durchführen zu lassen, sondern ebenfalls innerhalb eines Spieles. Die Input-Datensätze des zweiten Spieles wären dann die raumbezogenen Daten, die einem vorgelagerten ortsbezogenen Spiel erhoben würden. Eine andere Möglichkeit wäre, dass man die automatisch errechneten Cluster, durch die Spieler überprüfen und gegebenenfalls korrigieren lässt. Ein Idee, die auch schon im *Video Store Clerk* Spiel [2] erfolgreich eingesetzt wurde, siehe Kapitel 2 Abschnitt 2.1.

[2] `http://www.videostoreclerk.com/`

A. Anhang

A.1. Fragebogen zu CityExplorer

A.1.1. Fragebogen 1 - Vor einem CityExplorer Spiel

How old are you?

Please state your gender:

I have played at least one video game in my live (PC, home console - like the Nintendo Wii or PlayStation 2/3 - or handheld gaming systems - like the Nintendo DS or Sony PSP).

- Yes
- No
- No idea

I play video games on a regular basis.

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree
- No idea

Please give a rough estimation how much time per week you play video games:

Did you played a location-based game before?

- Yes (Name:)
- No
- No idea

I have played at least one game on my mobile phone in my live.

- Yes
- No
- No idea

I play mobile phone games on a regular basis.

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree
- No idea

Please give a rough estimation how much time per week you play mobile games:

Did you use any kind of location-based service with your PC before - like for example a web-based routing planner or Google Maps/Earth?

- Yes
- No
- No idea

Are you familiar or actually own some kind of GPS navigation system (e.g. a car navigation system or a GPS tourist guide)?

- Yes, familiar
- Yes, own

- Not even familiar
- No idea

Did you use any kind of location-based service with your mobile phone before - like for example a friend finder or an SNS, where you could locate yourself on a map?

- Yes
- No
- No idea

A.1.2. Fragebogen 2 - Nach einem CityExplorer Spiel

Please give a rough estimation of how much time per day you spend to play the game:

The outdoor part of the game was fun to play.

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree
- No idea

The online part of the game was fun to play.

- Strongly agree
- Agree
- Neutral
- Disagree

- Strongly disagree
- No idea

The game fitted between my everyday activities (like e.g. shopping, going to work/school or visiting friends).

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree
- No idea

I explicitly reserved time slots in my everyday schedule to play the game.

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree
- No idea

In order to play the game for free I would agree that the data (GPS positions, pictures and textual annotations) I collected throughout the game are used in a commercial context.

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree

- No idea

I played the game only in my free time e.g. after work, school or on the weekend.

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree
- No idea

Would you use personal categories in the game like „My Home“ or „My School“?

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree
- No idea

The game has encouraged me to visit places, which I rarely go to.

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree
- No idea

The game has increased my knowledge about the geographic environment (location of places, the street network etc.) of the city the game took place.

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree
- No idea

In order to be competitive in the game you have to invest a great amount of time.

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree
- No idea

The game has encouraged me to be more on the way outdoors than I usually am.

- Strongly agree
- Agree
- Neutral
- Disagree
- Strongly disagree
- No idea

It´s ok for me that my movement history is recorded during the time I play the game and latter anonymously uploaded to a website.

- Strongly agree

- Agree
- Neutral
- Disagree
- Strongly disagree
- No idea

A.1.3. Ergebnis der Fragebogenauswertung zum verteilten Spiel (Japan)

Question	Mean	Standard Deviation
1. The outdoor part of the game was fun to play.	3,6	0,6
2. The online part of the game was fun to play.	3	1
3. The game fitted between my everyday activities (like e.g. shopping, going to work/-school or visiting friends).	2	1
4. I explicitly reserved time slots in my everyday schedule to play the game.	2,3	1,5
5. In order to play the game for free I would agree that the data (GPS positions, pictures and textual annotations) I collected throughout the game are used in a commercial context.	1,6	1,2
6. I played the game only in my free time e.g. after work, school or on the weekend.	4,3	0,6
7. The game has encouraged me to visit places, which I rarely go to.	4	1

8. The game has increased my knowledge about the geographic environment (location of places, the street network etc.) of the city the game took place.	4	1
9. In order to be competitive in the game you have to invest a great amount of time.	5	0
10. The game has encouraged me to be more on the way outdoors than I usually am.	3,3	0,6
11. I had the feeling that I was actually playing against a real team in the other city, although I did not meet them in person.	3,3	2

Tabelle A.1.: Ergebnisse der Fragebogenauswertung - Likert Skala von 1 (Strongly disagree) bis 5 (Strongly agree) - bei n = 3

A.2. Analyse der Raumbezogenen Datensammlung von CityExplorer

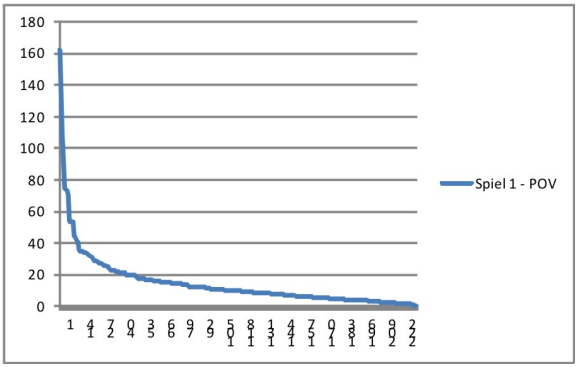


Abbildung A.1.: IQM des POV für die Datensammlung des ersten CityExplorer Spiels in Bamberg

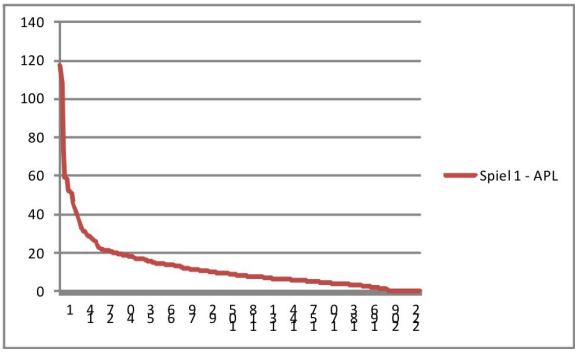


Abbildung A.2.: IQM des APL für die Datensammlung des ersten CityExplorer Spiels in Bamberg

X-Achse gibt den Index für eine durchgeführte Korrektur an - ein Spielstein kann dabei auch mehrmals korrigiert werden - an und die Y-Achse gibt das Ausmaß der von den Spielern durchgeführten Korrektur des POV und des APL in Metern an.

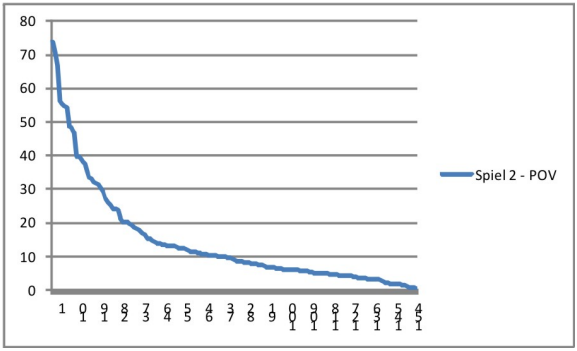


Abbildung A.3.: IQM des POV für die Datensammlung des zweiten CityExplorer Spiels in Bamberg

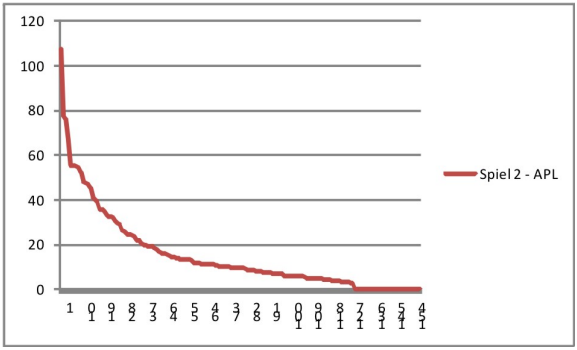


Abbildung A.4.: IQM des APL für die Datensammlung des zweiten CityExplorer Spiels in Bamberg

X-Achse gibt den Index für eine durchgeführte Korrektur an - ein Spielstein kann dabei auch mehrmals korrigiert werden - an und die Y-Achse gibt das Ausmaß der von den Spielern durchgeführten Korrektur des POV und des APL in Metern an.

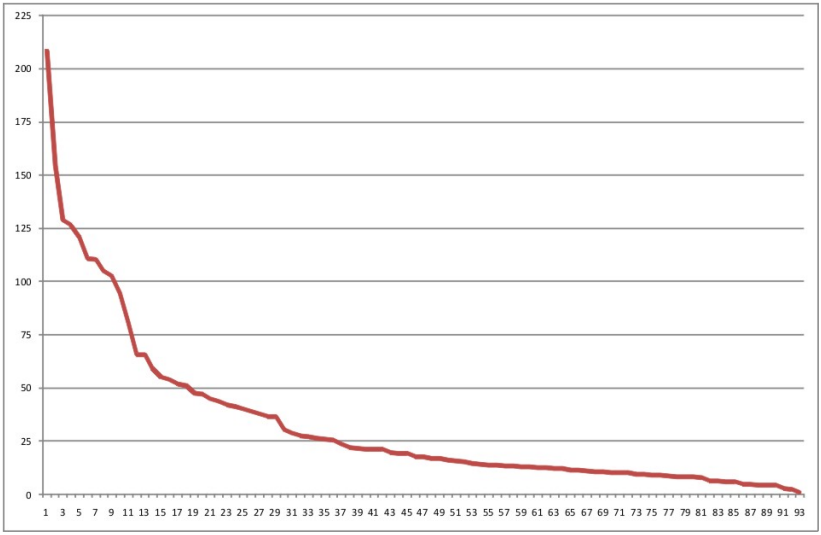


Abbildung A.5.: OQM Verteilung für die gesamte Stichprobe

A.3. GPS Tracks Analyse

Die Zeilen in den Tabellen A.2 und A.3 entsprechen den in den Spielevents aufgezeichneten Log-Dateien. Für einige Spieler/Teams sind aus technischen Gründen mehrere Log-Dateien pro Spielrunde aufgezeichnet worden, weshalb die Gesamtanzahl an Log-Dateien nicht mit der Anzahl der gespielten Runden übereinstimmen muss.

Spielfeld/Gute GPS Koordinaten	Weglänge (km)	Gesamt GPS Koordinaten	Spielevent
BAMBERG			
1122	2,6	1779	WELT
715	2,8	1591	WELT
1838	2,9	2505	WELT
690	3	2226	WELT
1181	3,9	2632	WELT
2080	4,9	3159	WELT
478	2,4	1781	WELT

1124	2,6	1532	WELT
366	2,8	2844	WELT
1183	2	1996	WELT
77	0,5	1091	WELT
508	2	1355	WELT
740	3,4	1819	WELT
439	0,9	873	WELT
859	1,4	1111	WELT
1432	7,2	2979	Bremen_1
191	0,8	1164	Andere
337	0,9	822	Andere
755	1,6	993	Andere
735	2,8	1600	Andere
417	2,1	1171	Andere
550	2,4	1576	Andere
789	3,2	1405	Andere
677	1,5	1227	Andere
137	0,6	2352	Andere
1025	1,9	1323	Andere
1467	0,5	3009	Andere
534	4,2	1791	MUT
2005	7,8	2266	MUT
715	6,6	2251	MUT
1670	8,4	2233	MUT
893	4,7	2182	EAI
1485	10,1	2080	EAI

Tabelle A.2.: Auflistung der gesammelten GPS Koordinaten in den 48 Geogames Partien - mit GeoTicTacToe und CityPoker - in Bamberg (Innenstadt und Feki), Coburg und Bremen; MUT = Mädchen und Technik [6]; EAI = Einführung in die Angewandte Informatik; WELT = Welterbetag [6]; BREMEN_1 = Verteiltes Spiel Bamberg vs. Bremen [4]

Spielfeld/Gute GPS Koordinaten	Weglänge (km)	Gesamt GPS Koordinaten	Spielevent
FEKI			
1154	2,9	1796	Andere
622	2,4	1070	Andere
278	0,2	676	Andere
1096	2,5	1426	Andere
837	2,5	1605	Andere
180	1,2	475	Andere
276	1,3	694	Andere
164	0,5	775	Andere
542	1,6	789	Andere
389	3,3	2053	Andere
214	0,8	408	Andere
393	3	2057	Andere
225	1	1989	Andere
273	1	648	Andere
544	2	1936	Andere
16	0,2	792	Andere
15	1	757	Andere
COBURG			
2094	3,1	2455	MUT
1862	3,9	2704	MUT
1931	3,3	2304	MUT
2091	4,8	2222	MUT
BREMEN			
1155	8,8	2725	Bremen_1

Tabelle A.3.: Auflistung der gesammelten GPS Koordinaten in den 48 Geogames Partien - mit GeoTicTacToe und CityPoker - in Bamberg (Innenstadt und Feki), Coburg und Bremen; MUT = Mädchen und Technik [6]; EAI = Einführung in die Angewandte Informatik; WELT = Welterbetag [6]; BREMEN_1 = Verteiltes Spiel Bamberg vs. Bremen [4]

A.4. GeoSnake Materialien

A.4.1. Vorstudie: Allgemeine Spielanweisungen

Regelvariante 1:

GeoSnake

In Geosnake sollen Sie die Positionen von geographischen Objekten (Gebäude, Straßen, Plätze) bestimmen.

Regeln:

- Wählen Sie eine Position, die das Objekt zweifelsfrei identifiziert.
- Vergeben Sie an jedem Objekt zwei Begriffe, die Ihrer Meinung nach mit diesem Objekt assoziiert werden können (Tags).
- Kreuzen Sie Ihre zurückgelegte Wegstrecke nicht und verwenden Sie keinen Straßenabschnitt mehr als ein Mal.

Gewinnbedingung:

- Die Anzahl an Objekten, die ein Spieler am Stück ohne Überschneidung besucht hat, ergibt seine Punktezahl.

Regelvariante 2:

Regeln:

- Wählen Sie die Position, die Ihrer Meinung nach von den meisten anderen Spielern gewählt wird.
- Vergeben Sie an jedem Objekt zwei Begriffe, die Ihrer Meinung nach von den meisten anderen Spielern mit diesem Objekt assoziiert werden (Tags).
- Kreuzen Sie Ihre zurückgelegte Wegstrecke nicht und verwenden Sie keinen Straßenabschnitt mehr als ein Mal.

Gewinnbedingung:

- Die Anzahl an Objekten, die ein Spieler am Stück ohne Überschneidung besucht hat, ergibt seine Punktezahl.

A.4.2. Vorstudie: Objekte für GPS-Teil

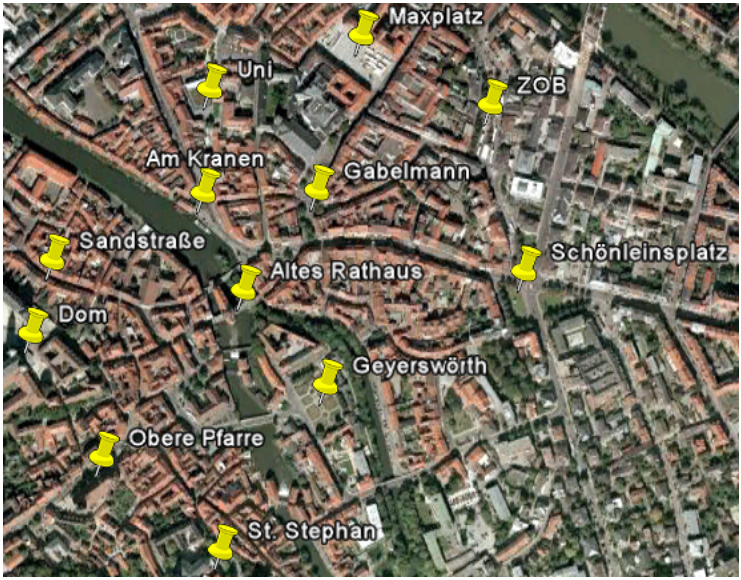


Abbildung A.6.: Geographische Verteilung der Objekte in der Vorstudie zu GeoSnake im GPS-Teil, © Karte: Google Earth (`http://earth.google.com/intl/de/`).

GeoSnake Objekte

Die Objekte können in beliebiger Reihenfolge abgelaufen werden. Alle Objekte sind zu besuchen. Als Beschreibung dürfen keine der hier verwendeten Namen verwendet werden.

Ab dem 7. Objekt benutzen Sie bitte die mitgegebene Regelvariante!!

- Otto-Friedrich Universität
- Obere Pfarre
- St. Stephans Kirche
- Sandstraße

- Altes Rathaus
- Zentraler Omnibusbahnhof (ZOB)
- Maxplatz
- Am Kranen
- Gabelmann
- Geyerswörth
- Bamberger Dom
- Schönleinsplatz

A.4.3. Vorstudie: Objekte für Bildschirm-Teil



Abbildung A.7.: Geographische Verteilung der Objekte in der Vorstudie zu GeoSnake im GPS-Teil, © Karte: Google Earth (`http://earth.google.com/intl/de/`).

GeoSnake Objekte

Die Objekte können in beliebiger Reihenfolge abgelaufen werden. Alle Objekte sind zu besuchen. Als Beschreibung dürfen keine der hier verwendeten Namen verwendet werden.

- Unterer Kaulberg
- Post in der Innenstadt
- St. Elisabeth (Sandstraße)
- Klein Venedig
- Karstadt
- Austraße
- Domberg
- Bolero
- Untere Rathausbrücke
- Heumarkt
- E.T.A. Hoffmann Theater
- Markusbrücke

A.4.4. Hauptstudie: Allgemeine Spielanweisungen

Regelvariante 1:

In Geosnake sollen Sie die Positionen von geographischen Objekten (Gebäude, Straßen, Plätze oder Regionen) bestimmen.

Regeln:

- Wählen Sie eine Position, die das Objekt zweifelsfrei identifiziert, also im Objekt oder möglichst nahe am Objekt liegt. In der unten stehenden Abbildung sehen Sie Beispiele von zulässigen, grün markierten und unzulässigen, rot markierten Positionen für die Konzerthalle.
- Kreuzen Sie Ihre zurückgelegte Wegstrecke nicht und verwenden Sie keinen Straßenabschnitt mehr als ein Mal. Die Abbildung zeigt zwei Arten von unzulässigen Spielzügen (Überkreuzen und Wegdoppelung), die mit Punktabzug bestraft werden.

Gewinnbedingung:

- Für jedes Objekt mit zulässiger Position nach Regel 1 erhält man einen Punkt.
- Bei einer Überschneidung wird das vorrangegangene identifizierte Objekt nicht gewertet.
- Wird eine Strecke zweimal in dieselbe oder entgegengesetzte Richtung wird das vorrangegangene und nachfolgende identifizierte Objekt nicht gewertet.
- Der Spieler mit den meisten Punkten gewinnt.



Regelvariante 2:

In Geosnake sollen Sie die Positionen von geographischen Objekten (Gebäude, Straßen, Plätze oder Regionen) bestimmen.

Regeln:

- Wählen Sie die Position, die Ihrer Meinung nach von den meisten anderen Spielern gewählt wird.
- Kreuzen Sie Ihre zurückgelegte Wegstrecke nicht und verwenden Sie keinen Straßenabschnitt mehr als ein Mal. Die Abbildung zeigt zwei Arten von unzulässigen Spielzügen (Überkreuzen und Wegdoppelung), die mit Punktabzug bestraft werden.

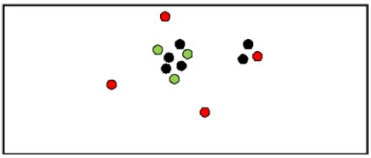
Gewinnbedingung:

- Für jedes Objekt dessen Position dort gewählt wurde, welche die meisten anderen Spieler auch gewählt haben, bekommt der Spieler einen Punkt.
- Bei einer Überschneidung wird das vorrangegangene identifizierte Objekt nicht gewertet.
- Wird eine Strecke zweimal in dieselbe oder entgegengesetzte Richtung wird das vorrangegangene und nachfolgende identifizierte Objekt nicht gewertet.

- Der Spieler mit den meisten Punkten gewinnt.







A.4.5. Hauptstudie: Objekte für GPS-Teil

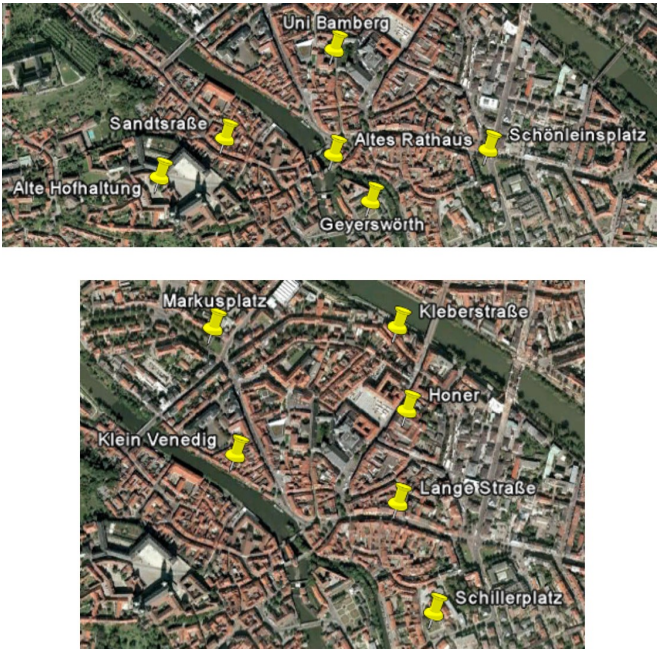


Abbildung A.8.: Geographische Verteilung der Objekte in der Hauptstudie zu GeoSnake im GPS-Teil, © Karte: Google Earth (`http://earth.google.com/intl/de/`).

GeoSnake Objekte V1

Die Objekte können in beliebiger Reihenfolge abgelaufen werden. Alle Objekte sind zu besuchen.

- Otto-Friedrich Universität
- Sandstraße
- Altes Rathaus
- Alt Hofhaltung
- Geyerswörth

- Schönleinsplatz

GeoSnake Objekte V2



Abbildung A.9.: Geographische Verteilung der Objekte in der Hauptstudie zu GeoSnake im PC-Teil, © Karte: Google Earth (`http://earth.google.com/intl/de/`).

Die Objekte können in beliebiger Reihenfolge abgelaufen werden. Alle Objekte sind zu besuchen.

- Klein Venedig
- Lange Straße
- Honer
- Schillerplatz
- Markusplatz
- Kleberstraße

A.4.6. Hauptstudie: Objekte für Bildschirm-Teil

Die Objekte können in beliebiger Reihenfolge abgelaufen werden. Alle Objekte sind zu besuchen.

- Klein Venedig
- Lange Straße
- Honer
- Schillerplatz
- Markusplatz
- Kleberstraße
- Otto-Friedrich Universität
- Sandstraße
- Altes Rathaus
- Alt Hofhaltung
- Geyerswörth
- Schönleinsplatz

A.4.7. Fragebogen Vor- und Hauptstudie

1. How old are you?

2. Please state your gender:

3. Where do you live? (select one option and fill in the details)

- Bamberg - since *x* years
- Bamberg region (Landkreis) - since *x* years
- Somewhere else, but I visit Bamberg on a regular basis.
- How many times per month?
- How many hours per visit?
- Somewhere else and I rarely or never visit Bamberg.
- How many times have you been to Bamberg in your life?
- Somewhere else, but I have once lived in Bamberg or its region for years.

4. I usually visit the city center of Bamberg (select none, one, or several options)

- 1. because I live there
- 2. because I go to work/school/university there
- 3. for shopping
- 4. for sightseeing
- 5. for going out in the evening
- 6. for visiting some friends or relatives
- 7. other:

5. I usually move around the city center of Bamberg (select one or several)

- 1. on foot
- 2. by bike

- 3. by bus
- 4. by car
- 5. other:

6. Have you ever attended a guided sightseeing tour in the city center of Bamberg?

- yes
- no

7. When moving around Bamberg (select one)

- 1. I never get lost because I exactly know my way.
- 2. I never get lost because of my navigation assistance (e.g. paper map, GPS).
- 3. I sometimes get lost.
- 4. I often get lost.

8. Have you ever played a location-based game before?

- Yes - Which one(s)?
- No
- No idea

9. I have played at least one game on my mobile phone in my live.

- Yes
- No
- No idea

10. I play mobile phone games on a regular basis.

- Strongly agree
- Agree
- Neutral

- Disagree
- Strongly disagree
- No idea

11. Please give a rough estimation of how much time per week you spend on mobile phone games:

12. Have you used a location-based service with your PC before - like, for example, a web-based routing planner or Google Maps/Earth?

- Yes
- No
- No idea

13. Have you ever used Google Maps/Earth to search and save a geographic location and share it with a friend?

- Yes
- No
- No idea

14. Are you familiar with GPS navigation systems (e.g. car navigation system or GPS handheld)?

- 1. Yes, I own such a system.
- 2. Yes, I am familiar with this kind of systems, but I do not own one.
- 3. I am not familiar with these systems.
- 4. No idea

15. Have you ever recorded GPS data while moving around outdoors and shared the resulting data (tracks or points of interest) on the internet?

- Yes
- How did you share the data? (select one or several)
- via e-mail

- by uploading it to an online platform or community
- Which one(s)?
- other
- No
- No idea

16. Have you ever used a location-based service with your mobile phone before (e.g. friend finder, you-are-here-map)?

- Yes
- No
- No idea

17. Do you know the ESP game (by Luis von Ahn)?

- Yes, and I have even played it
- I have heard about it, but never tried it out.
- I have never heard about this game.

A.4.8. Antworten Fragebogen Vorstudie

Question				
1.	25	25	22	32
2.	m	m	m	f
3.	Bamberg, 5	Landkreis Bamberg, 25	Bamberg, 3	Bamberg, 10
4.	3,5,6,Fußball spielen	1,2,3,5	2,5	1,3
5.	1,2,Kin“-der“-wagen	1,2,3,4	1,2,3	1
6.	yes	yes	no	yes
7.	1	1	1	1
8.	yes	yes	yes	yes
9.	yes	yes	yes	yes
10.	1	2	3	2
11.	0	0	1 h	0
12.	yes	yes	yes	yes
13.	yes	yes	yes	yes
14.	3	1	2	1
15.	no	yes, mail	no	yes, upload
16.	no	yes	no	no
17.	3	3	3	3

Tabelle A.4.: Antowrten zum Fragebogen der Vorstudie

Question				
1.	24	24	28	30
2.	m	m	m	f
3.	Bamberg, 2006	Bamberg, 1,5	Bamberg, 28	Bamberg, 2
4.	1,3,5	1,2,3,5	1,2,3,5	2,3,4,5
5.	1,2,3	2,4	1,2	1,2
6.	yes	no	yes	yes
7.	1	1,2	1	1

8.	yes	no	yes	yes
9.	yes	yes	yes	no
10.	3	1	4	1
11.	0,5 h	0	0	0
12.	yes	yes	yes	yes
13.	no	yes	no	yes
14.	1	1	2	3
15.	no	yes, upload	no	yes, mail
16.	no	yes	no	no
17.	3	3	3	3

Tabelle A.5.: Antowrten zum Fragebogen der Vorstudie

Question			
1.	23	23	34
2.	f	f	f
3.	Bamberg, 3	Landkreis Bamberg, 3	Bamberg, 2
4.	1	2,3,5,6	1,2,3,4,5,6
5.	1,2	1,2,4	1,2,3
6.	no	no	yes
7.	3	1	1
8.	no idea	no	yes
9.	yes	yes	no
10.	no idea	3	2
11.	0	0,5 h	0
12.	yes	yes	yes
13.	yes	yes	no
14.	2	1	3
15.	yes, mail	yes, mail	no
16.	no	no	no
17.	3	3	3

Tabelle A.6.: Antowrten zum Fragebogen der Vorstudie

A.4.9. Antworten Fragebogen Hauptstudie

Question				
1.	42	21	25	26
2.	f	m	f	f
3.	Landkreis Bamberg, 2	Bamberg, 0,2	Bamberg, 6	Bamberg, 2
4.	2,3,4,5	2,3,5	1,2,3,5,6	3,5
5.	1,3	1,2,3	1,2,3,4	1
6.	no	no	no	yes
7.	3	3	1	3
8.	no	no	no	no
9.	yes	yes	no	yes
10.	2	2	1	2
11.	0,5 h	0	0	0
12.	yes	yes	yes	no
13.	no	yes	yes	yes
14.	2	3	2	1
15.	no	no	no	no
16.	no	no	no	no
17.	3	3	3	3

Tabelle A.7.: Antworten zum Fragebogen der Hauptstudie

Question				
1.	27	23	28	23
2.	m	f	f	m
3.	Bamberg, 5,5	Bamberg, 4	Bamberg, 9	Bamberg, 4
4.	1,2,3,4,5,6	1,2,3,5,6	1,3,5,6	1,3,5,6
5.	1,2	1,2,4	2	2,3
6.	no	no	yes	no
7.	3	3	3	1
8.	no	no	no	yes
9.	yes	yes	yes	yes
10.	3	1	3	3

11.	0,5 h	0	0	2 h
12.	yes	yes	yes	yes
13.	yes	yes	yes	yes
14.	1	1	2	2
15.	no	no	no	no
16.	no	no	no	no
17.	3	3	3	3

Tabelle A.8.: Antworten zum Fragebogen der Hauptstudie

Question				
1.	25	21	27	26
2.	f	f	f	f
3.	Bamberg, 1	Bamberg, 1	(Landkreis) Bamberg, (2,5) 2,5	Bamberg, 7
4.	1,2,3,4,5,6	1,2,5,6	2,3,4,5,6	1,3,5,6
5.	1,2,3	1,3	1,2	1,2,4
6.	yes	no	yes	no
7.	3	1	3	1
8.	no	no	yes	yes
9.	yes	yes	yes	yes
10.	2	1	1	2
11.	0	0	0	0
12.	yes	yes	yes	yes
13.	no	yes	yes	no
14.	2	2	2	1
15.	no	no	no	no
16.	no	no	no	no
17.	3	3	3	3

Tabelle A.9.: Antowrten zum Fragebogen der Hauptstudie

A.5. Material zum semantischen Integrationsprozess

A.5.1. Evaluierungsstudie zum SVAC-Algorithmus

Kategorie	Index	Instanzen
restaurant	3	154
coffeehouse	41	34
tavern	42	26
bank	43	17
parking lot	44	40
convenience store	45	18
amusement arcade	46	5
monument	33	67
intersection	19	7
hotel	1	34
doctor	32	43
playground	34	4
street sign	36	17
fountain	25	6
hairstylist	29	13
shoe store	31	18
bakery	39	22
bridge	20	14
art	26	78
food	21	67
beer	35	39
church	28	30

Tabelle A.10.: Verwendete Kategorien in der Evaluierungsstudie für den SVAC-Algorithmus und die in der CityExplorer gezählten Instanzen

Kategorie	Kategorie	**Ähnlichkeit** nach [8]		Kategorie	Kategorie	**Ähnlichkeit** nach [8]
1	3	0.1504		3	19	0.0269
1	19	0.0269		3	20	0.0871
1	20	0.0871		3	21	0
1	21	0		3	25	0.0871
1	25	0.0871		3	26	0.055
1	26	0.055		3	28	0
1	28	0		3	29	0.0034
1	29	0.0034		3	31	0
1	31	0		3	32	0.0034
1	32	0.0034		3	33	0.0871
1	33	0.0871		3	34	0.0269
1	34	0.0269		3	35	0
1	35	0		3	36	0
1	36	0		3	39	0.0269
1	39	0.0269		3	41	0
1	41	0		3	42	0.1504
1	42	0.1504		3	43	0.055
1	43	0.055		3	44	0
1	44	0		3	45	0
1	45	0		3	46	0
1	46	0				

Tabelle A.11.: Nach [8] berechnete Ähnlichkeiten zwischen den Kategorien

Kategorie	Kategorie	**Ähnlichkeit** nach [8]		Kategorie	Kategorie	**Ähnlichkeit** nach [8]
19	20	0.2087		20	21	0
19	21	0.3502		20	25	0.204
19	25	0.204		20	26	0.2292
19	26	0.2416		20	28	0.1301
19	28	0.3978		20	29	0.0397
19	29	0.0034		20	31	0.055
19	31	0		20	32	0.2292
19	32	0.2087		20	33	0.0871
19	33	0.0269		20	34	0.0269
19	34	0.254		20	35	0.0034
19	35	0.1301		20	36	0.1301
19	36	0.1301		20	39	0.055
19	39	0.3109		20	41	0.055
19	41	0		20	42	0.055
19	42	0.0269		20	43	0.4816
19	43	0.3978		20	44	0.0269
19	44	0		20	45	0.055
19	45	0		20	46	0.055
19	46	0				

Tabelle A.12.: Nach [8] berechnete Ähnlichkeiten zwischen den Kategorien

Kategorie	Kategorie	**Ähnlichkeit** nach [8]		Kategorie	Kategorie	**Ähnlichkeit** nach [8]
21	25	0.0034		25	26	0.204
21	26	0.1301		25	28	0.1301
21	28	0.1301		25	29	0.0034
21	29	0		25	31	0
21	31	0		25	32	0.204
21	32	0		25	33	0.0871
21	33	0		25	34	0.0269
21	34	0		25	35	0
21	35	0		25	36	0
21	36	0.1301		25	39	0.0269
21	39	0		25	41	0
21	41	0		25	42	0.055
21	42	0		25	43	0.1301
21	43	0.1301		25	44 0	
21	44	0		25	45	0
21	45	0		25	46	0
21	46	0				

Tabelle A.13.: Nach [8] berechnete Ähnlichkeiten zwischen den Kategorien

Kategorie	Kategorie	**Ähnlichkeit** nach [8]		Kategorie	Kategorie	**Ähnlichkeit** nach [8]
26	28	0.1301		28	29	0.0034
26	29	0		28	31	0
26	31	0		28	32	0.2292
26	32	0.0034		28	33	0.0269
26	33	0.055		28	34	0
26	34	0.0269		28	35	0.0034
26	35	0		28	36	0.1301
26	36	0.3125		28	39	0.0269
26	39	0		28	41	0.1504
26	41	0		28	42	0.1504
26	42	0		28	43	0.4188
26	43	0.1301		28	44	0.0269
26	44	0		28	45	0
26	45	0		28	46	0
26	46 0					

Tabelle A.14.: Nach [8] berechnete Ähnlichkeiten zwischen den Kategorien

Kategorie	Kategorie	**Ähnlichkeit** nach [8]		Kategorie	Kategorie	**Ähnlichkeit** nach [8]
29	31	0.0397		31	32	0.0397
29	32	0.3923		31	33	0.0871
29	33	0.0397		31	34	0.0269
29	34	0.0269		31	35	0.0034
29	35	0.0034		31	36	0
29	36	0		31	39	0.387
29	39	0.0397		31	41	0
29	41	0		31	42	0
29	42	0.0034		31	43	0
29	43	0.0034		31	44	0
29	44	0		31	45	0.387
29	45	0.0397		31	46	0.055
29	46	0.0397				

Tabelle A.15.: Nach [8] berechnete Ähnlichkeiten zwischen den Kategorien

Kategorie	Kategorie	**Ähnlichkeit** nach [8]		Kategorie	Kategorie	**Ähnlichkeit** nach [8]
32	33	0.0397		33	34	0.0269
32	34	0.0269		33	35	0.0034
32	35	0.3125		33	36	0
32	36	0		33	39	0.3109
32	39	0.0397		33	41	0.0269
32	41	0.0397		33	42	0.0269
32	42	0.0034		33	43	0.0269
32	43	0.1301		33	44	0.289
32	44	0.0269		33	45	0.0871
32	45	0.0397		33	46	0.055
32	46	0.0397		35	36	0
34	35	0.0034		35	39	0.0034
34	36	0		35	41	0.0034
34	39	0		35	42	0.0034
34	41	0.0269		35	43	0.1301
34	42	0		35	44	0.0034
34	43	0.0269		35	45	0
34	44	0.289		35	46	0.0034
34	45	0				
34	46	0.0269				

Tabelle A.16.: Nach [8] berechnete Ähnlichkeiten zwischen den Kategorien

Kategorie	Kategorie	**Ähnlichkeit** nach [8]		Kategorie	Kategorie	**Ähnlichkeit** nach [8]
36	39	0		39	41	0
36	41	0		39	42	0.0269
36	42	0		39	43	0.0269
36	43	0		39	44	0
36	44	0		39	45	0.387
36	45	0		39	46	0.055
36	46	0		42	43	0.055
41	42	0.1504		42	44	0
41	43	0.055		42	45	0
41	44	0.0269		42	46	0
41	45	0		44	45	0
41	46	0		44	46	0
43	44	0		45	46	0.055
43	45	0				
43	46	0				

Tabelle A.17.: Nach [8] berechnete Ähnlichkeiten zwischen den Kategorien

A.6. GeoSnake - R Auswertung

```
1
2  ##Auswertung spiel1
3  #Datensatz spiel1 wählen
4
5  spiel1 <- read.table(file.choose(), header=T)
6  spiel1$Ort <- as.factor(spiel1$Ort)
7  spiel1$Regel <- as.factor(spiel1$Regel)
8  spiel1$Nutzer <- as.factor(spiel1$Nutzer)
9  spiel1$Distanz <- as.numeric(spiel1$Distanz)
10
11 library(lme4)
12 model1lm <- lmer(Distanz ~ factor(Regel)+time+(1+
       time|Nutzer), data=spiel1)
13 summary(model1lm)
14
15 #feste Effekte: Zeit ist nicht signifikant; Regel 2
       unterscheidet sich signifikant
16 #von Regel 1 (die mittleren Abstände sind geringer)
       ; Regel drei unterscheidet
17 #sich nicht signifikant von Regel 1, sind aber auch
       geringere Abstände
18 #zufällige (individuenspezifische Effekte): auch
       hier ist die Zeit nicht signifikant
19 # aber ein individuensprzifischer Interzept ist
       sinnvoll
20
21 model1lmz <- lmer(Distanz ~ factor(Regel)+(1|Nutzer
       ), data=spiel1)
22 summary(model1lmz)
23
24 # ohne Zeit als zufälligen Effekt sind Regel 2 und
       3 signifikant besser als Regel 1
25
26 #nun das ganze mit der Gamma-Verteilung
27
```

```
28 modell1lm <- glm(Distanz \~ Regel+time, data=spiel1,
       family="'Gamma"')
29 summary(modell1lm)
30
31 #damit man auch Vergleich von Regel 2 und 3 hat:
32 # Umkodierung der Variablen:
33
34 spiel1$regeln <- ifelse(spiel1$Regel==1, 3, 1)
35 spiel1$regelneu <- ifelse(spiel1$Regel==2, 2,
       spiel1$regeln)
36 spiel1$regelneu <- as.factor(spiel1$regelneu)
37
38 # das größere Modell rechnen
39
40 modell1lm <- lmer(Distanz ~ factor(regelneu)+time
       +(1+time|Nutzer), data=spiel1)
41 summary(modell1lm)
42
43 #Regel 2 ist nicht signifikant verschieden von
       Regel 3
44 # mit dem kleineren Modell
45
46 modell1lmz <- lmer(Distanz ~ factor(regelneu)+(1|
       Nutzer), data=spiel1)
47 summary(modell1lmz)
48
49 # auch hier Regel 2 nicht signifikant verschieden
       von Regel 3
50
51 modell1lm <- glm(Distanz ~ regelneu+time, data=
       spiel1, family="'Gamma"')
52 summary(modell1lm)
53
54 ##########################
55 ##Auswertung spiel2
56 #Datensatz spiel2 wählen
57
58 spiel2 <- read.table(file.choose(), header=T)
59 spiel2$Ort <- as.factor(spiel2$Ort)
```

```
spiel2$Regel <- as.factor(spiel2$Regel)
spiel2$Nutzer <- as.factor(spiel2$Nutzer)
spiel2$Distanz <- as.numeric(spiel2$Distanz)

#gemischtes lineares Modell mit Normalverteilung
#feste Effekte: Regel und Zeit
#Zufällige Effekte: Interzept und Zeit für die einzelnen Nutzer

model2lm4 <- lmer(Distanz~Regel+time+(1+time||Nutzer), data=spiel2)
summary(model2lm4)

#Regel 2 und 3 sind signifikant besser als Regel 1

#Generalisiertes lineares Modell mit Gamma-Verteilung
#Achtung: Parameter andersherum interpretieren

model2lm2 <- glm(Distanz ~ Regel+time, data=spiel2, family="'Gamma"')
summary(model2lm2)

#verwendet man die Gammeverteilung, ist nuur noch Regel 2 signifikant besser als
#Regel 3; aber hier keine zufälligen Effekte mehr!!

#zur Umkodierung

spiel2$regeln <- ifelse(spiel2$Regel==1, 3, 1)
spiel2$regelneu <- ifelse(spiel2$Regel==2, 2, spiel2$regeln)
spiel2$regelneu <- as.factor(spiel2$regelneu)

model2lm4 <- lmer(Distanz~regelneu+time+(1+time|Nutzer), data=spiel2)
summary(model2lm4)

```

```
91 #Regel 2 unterscheidt sich nicht signifiaknt von
       Regel 3
92 #nun nochmal mit Gamma-Verteilung
93
94 model2lm2 <- glm(Distanz \~ regelneu+time, data=
       spiel2, family="'Gamma"')
95 summary(model2lm2)
96
97 #hier ist Regel 2 signifikant (Niveau 0.1) besser
       als Regel 3
```

Literaturverzeichnis

[1] Schlossberg, M., Shuford, E.: Delineating 'public' and 'participation' in ppgis. URISA Journal **16**(2) (2005) 15–26

[2] Grant, L., Daanen, H., Benford, S., Hampshire, A., Drozd, A., Greenhalgh, C.: Mobimissions: the game of missions for mobile phones. In: SIGGRAPH '07: ACM SIGGRAPH 2007 educators program, New York, NY, USA, ACM (2007) 12

[3] Ashbrook, D., Starner, T.: Using gps to learn significant locations and predict movement across multiple users. Personal Ubiquitous Comput. **7**(5) (2003) 275–286

[4] Kiefer, P., Matyas, S., Schlieder, C.: Playing location-based games on geographically distributed game boards. In et al., M., ed.: 4th International Symposium on Pervasive Gaming Applications (PerGames 2007), Salzburg, Austria, Shaker Verlag Aachen (2007)

[5] Matyas, S.: Playful geospatial data acquisition by location-based gaming communities. The International Journal of Virtual Realities (IJVR) **6**(3) (2007) 1–10

[6] Kiefer, P., Matyas, S., Schlieder, C.: Learning about cultural heritage by playing geogames. In Harper, R., Rauterberg, M., Combetto, M., eds.: Proc. 5th Int. Conference on Entertainment Computing (ICEC 2006), Cambridge, UK, Springer (LNCS 4161) (2006)

[7] Morris, S., Morris, A., Barnard, K.: Digital trail libraries. In: 4th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '04, ACM Press (2004) 63–71

[8] Resnik, P.: Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. Journal of Artificial Intelligence Research **11** (1995) 95–130

[9] Goodchild, M.F.: Citizens as voluntary sensors: Spatial data infrastructure in the world of web 2.0. International Journal of Spatial Data Infrastructures Research **2** (2007) 24–32

[10] O'Reilly, T.: What is web 2.0? design patterns and business models for the next generation of software. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html (2005) accessed 30.01.2008.

[11] Miller, P.: Web 2.0: Building the new library. http://www.ariadne.ac.uk/issue45/miller/intro.html (2005) accessed 30.01.2008.

[12] Loehnert E., Wittmann E., P.J., F., S.: Paramount- public safety & commercial info-mobility applications & services in the mountains. In: 14th International Technical Meeting of the Satellite Division of The Institute of Navigation, Salt Lake City, Utah, USA, ACM Press (2001)

[13] Paolo Bellavista, A.K., Helal, S.: Location-based services: Back to the future. IEEE Pervasive Computing **7**(2) (2008) 85–89

[14] Kuhn, W.: Volunteered geographic information and giscience. In: NCGIA and Vespucci Workshop on Volunteered Geographic Information. (2007)

[15] Mason, W., Watts, D.J.: Financial incentives and the „performance of crowds". In: HCOMP '09: Proceedings of the ACM SIGKDD Workshop on Human Computation, New York, NY, USA, ACM (2009) 77–85

[16] Coleman, D., Georgiadou, Y., Labonte, J.: Volunteered geographic information: the nature and motivation of produsers. International Journal of Spatial Data Infrastructures Research, Special Issue GSDI-11, under review (2009)

[17] Toffler, A.: The Third Wave: The Classic Study of Tomorrow. Bantam (1984)

[18] Peltola, J., Karsten, H.: When play is not enough: Towards actually useful applications for digital entertainment. Mobility Roundtable 2006 (2006)

[19] von Ahn, L., Dabbish, L.: Labeling images with a computer game. In: ACM Conference on Human Factors in Computing Systems, ACM Press (2004) 319–326

[20] von Ahn, L., Dabbish, L.: Designing games with a purpose. Commun. ACM **51**(8) (2008) 58–67

[21] Björk, S., Holopainen, J.: Patterns in Game Design (Game Development Series). Charles River Media, Inc., Rockland, MA, USA (2004)

[22] Peitz, J., Björk, S.: Understanding pervasive games through gameplay design patterns. In: Digital Games Research Association (DiGRA) Conference. (2007)

[23] Surowiecki, J.: The Wisdom Of Crowds: Why The Many Are Smarter Than The Few And How Collective Wisdom Shapes Business, Economies, Societies And Nations. Little, Brown Book Group (2004)

[24] Hill, L.L.: Core elements of digital gazetteers: placenames, categories, and footprints. In: ECDL 2000: Research and Advanced Technology for Digital Libraries. (2000) 280–290

[25] Hill, L.L.: Georeferencing: the geographic associations of information. MIT Press (2006)

[26] Stork, D.G., Lam, C.P.: Open mind animals: Insuring the quality of data openly contributed over the world wide web. Technical Report of the AAAI'2000 Workshop on Learning from Imbalanced Data Sets (2000)

[27] Stork, D.G.: Character and document research in the open mind initiative. In: ICDAR '99: Proceedings of the Fifth International Conference on Document Analysis and Recognition, Washington, DC, USA, IEEE Computer Society (1999) 1

[28] Salen, K., Zimmerman, E.: Rules of Play: Game Design Fundamentals. The MIT Press (2003)

[29] Von Ahn, L.: Human computation. PhD thesis, Pittsburgh, PA, USA (2005) Adviser-Blum, Manuel.

[30] von Ahn, L.: Human computation. In: ICDE '08: Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, Washington, DC, USA, IEEE Computer Society (2008) 1–2

[31] Howe, J.: Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business. 1 edn. Crown Business (2008)

[32] Kittur, A., Chi, E.H., Suh, B.: Crowdsourcing user studies with mechanical turk. In: CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM (2008) 453–456

[33] Tversky, A., Kahneman, D.: Judgment under uncertainty: Heuristics and biases. Science **185** (1974) 1124–31

[34] Mrozinski, J., Whittaker, E., Furui, S.: Collecting a why-question corpus for development and evaluation of an automatic QA-system. In: Proceedings of ACL-08: HLT, Columbus, Ohio, Association for Computational Linguistics (June 2008) 443–451

[35] Yan, T., Marzilli, M., Holmes, R., Ganesan, D., Corner, M.: mcrowd: a platform for mobile crowdsourcing. In: SenSys '09: Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems, New York, NY, USA, ACM (2009) 347–348

[36] Rana, R.K., Chou, C.T., Kanhere, S.S., Bulusu, N., Hu, W.: Ear-phone: an end-to-end participatory urban noise mapping system. In: IPSN '10: Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks, New York, NY, USA, ACM (2010) 105–116

[37] von Ahn, L., Blum, M., Hopper, N.J., Langford, J.: Captcha: Using hard ai problems for security. In: In Proceedings of Eurocrypt, Springer-Verlag (2003) 294–311

[38] von Ahn, L., Kedia, M., Blum, M.: Verbosity: a game for collecting common-sense facts. In: CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems, New York, NY, USA, ACM (2006) 75–78

[39] von Ahn, L., Liu, R., Blum, M.: Peekaboom: a game for locating objects in images. In: CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems, New York, NY, USA, ACM (2006) 55–64

[40] von Ahn, L., Ginosar, S., Kedia, M., Liu, R., Blum, M.: Improving accessibility of the web with a computer game. In: CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems, New York, NY, USA, ACM (2006) 79–82

[41] Law, E., von Ahn, L.: Input-agreement: a new mechanism for collecting data using human computation games. In: CHI '09: Proceedings of the 27th international conference on Human factors in computing systems, New York, NY, USA, ACM (2009) 1197–1206

[42] Law, E., von Ahn, L., Dannenberg, R., Crawford, M.: Tagatune: a game for music and sound annotation. In: Proceedings of the 8th International Conference on Music Information Retrieval (ISMIR 2007). (2007)

[43] Yuen, M., Chen, L., King, I.: A survey of human computation systems. Computational Science and Engineering, IEEE International Conference on **4** (2009) 723–728

[44] Orkin, J., Roy, D.: The restaurant game: Learning social behavior and language from thousands of players online. Journal of Game Development **3** (2007) 39–60

[45] Siorpaes, K., Hepp, M.: Ontogame: weaving the semantic web by online games. In: ESWC'08: Proceedings of the 5th European semantic web conference on The semantic web, Berlin, Heidelberg, Springer-Verlag (2008) 751–766

[46] Siorpaes, K., Hepp, M.: Games with a purpose for the semantic web. IEEE Intelligent Systems **23**(3) (2008) 50–60

[47] Krause, M., Aras, H.: Playful tagging: folksonomy generation using online games. In: WWW '09: Proceedings of the 18th international conference on World wide web, New York, NY, USA, ACM (2009) 1207–1208

[48] Poesio, M.: Play your way to an annotated corpus: games with a purpose and anaphoric annotation. In: IWCS-8 '09: Proceedings of the Eighth International Conference on Computational Semantics, Morristown, NJ, USA, Association for Computational Linguistics (2009) 3–3

[49] Ma, H., Chandrasekar, R., Quirk, C., Gupta, A.: Improving search engines using human computation games. In: CIKM '09: Proceeding of the 18th ACM conference on Information and knowledge management, New York, NY, USA, ACM (2009) 275–284

[50] Edery, D., Mollick, E.: Changing the Game: How Video Games Are Transforming the Future of Business. FT Press (2009)

[51] Longley, P., Goodchild, M., Maguire, D., Rhind, D.: Geographic Information Systems and Science. John Wiley & Sons (2001)

[52] Tulloch, D.L.: Is vgi participation? from vernal pools to video games. GeoJournal **72**(3-4) (2008) 161–171

[53] Chrisman, N.R.: Design of geographic information systems based on social and cultural goals. Photogrammetric Engineering and Remote Sensing **53**(10) (1987) 1367–1370

[54] Sieber, R.: Public participation geographic information systems: A literature review and framework. Annals of the Association of American Geographers **96**(3) (2006) 491–507

[55] Dunn, C.E.: Participatory gis a people's gis? Progress in Human Geography **31**(5) (2007) 616–637

[56] Elwood, S.: Critical issues in participatory gis: Deconstructions, reconstructions, and new research directions. T. GIS **10**(5) (2006) 693–708

[57] Arnstein, S.R.: A ladder of citizen participation. Journal of the American Planning Association **35**(4) (1969) 216–224

[58] Hall, G.B., Chipeniuk, R., Feick, R.D., Leahy, M.G., Deparday, V.: Community-based production of geographic information using open source software and web 2.0. Int. J. Geogr. Inf. Sci. **24**(5) (2010) 761–781

[59] Elwood, S.: Volunteered geographic information: future research directions motivated by critical, participatory, and feminist gis. GeoJournal **72**(3) (August 2008) 173–183

[60] Elwood, S.: Volunteered geographic information: key questions, concepts and methods to guide emerging research and practice. GeoJournal **72**(3) (August 2008) 133–135

[61] Sui, D.Z.: The wikification of gis and its consequences: Or angelina jolie's new tattoo and the future of gis. Computers, Environment and Urban Systems **32**(1) (2008) 1–5

[62] Turner, A.: Introduction to neogeography. O'Reilly (2006)

[63] Goodchild, M.F., Fu, P., Rich, P.: Sharing geographic information: An assessment of the geospatial one-stop. Annals of the Association of American Geographers **97**(2) (2007) 250–266

[64] Goodchild, M.F.: Citizens as sensors: the world of volunteered geography. GeoJournal **69**(4) (2008) 211–221

[65] Craglia, M., Goodchild, M., Annoni, A., Camara, G., Gould, M., Kuhn, W., Mark, D., Masser, I., Maguire, D., Liang, S., Parsons, E.: Next-generation digital earth: A position paper from the vespucci initiative for the advancement of geographic information science. International Journal of Spatial Data Infrastructures Research **3** (2008) 146–167

[66] Flanagin, A., Metzger, M.: The credibility of volunteered geographic information. GeoJournal **72**(3) (August 2008) 137–148

[67] Bishr, M., Kuhn, W.: Geospatial information bottom-up: A matter of trust and semantics. (2007) 365–387

[68] Bishr, M., Mantelas, L.: A trust and reputation model for filtering and classifying knowledge about urban growth. GeoJournal **72** (July 2008) 229–237

[69] Li, S., Gong, J.: Mashup: a new way of providing web mapping and gis services. In: Proceedings of Commission IV. (2008) 639–649

[70] Haklay, M., Singleton, A., Parker, C.: Web mapping 2.0: The neogeography of the geoweb. Geography Compass **2** (2008) 2011–2039

[71] Espinoza, F., Persson, P., Sandin, A., Nyström, H., Cacciatore., E., Bylund, M.: Geonotes: Social and navigational aspects of location-based information systems. In: Ubicomp 2001: Ubiquitous Computing, Berlin: Springer (2001) 2–17

[72] Lane, G.: Urban tapestries: Wireless networking, public authoring and social knowledge. Personal Ubiquitous Comput. **7** (2003) 169–175

[73] Bamford, W., Coulton, P., Edwards, R.: Space-time travel blogging using a mobile phone. In: ACE '07: Proceedings of the international conference on Advances in computer entertainment technology, New York, NY, USA, ACM (2007) 1–8

[74] Caillois, R.: Man, play, and games. Translated edn. The Free Press of Glencoe, a devision of Simon and Schuster, Inc. (1961)

[75] Aarseth, E.: Computer game studies - year one. Game Studies - The International Journal of Computer Game Research **1**(1) (2001) http://www.gamestudies.org/0101/, letzter Zugriff 21.06.2010.

[76] Magerkurth, C., Cheok, A.D., Mandryk, R.L., Nilsen, T.: Pervasive games: bringing computer entertainment back to the real world. Comput. Entertain. **3**(3) (2005) 4–4

[77] Montola, M., Stenros, J., Waern, A.: Pervasive Games: Theory and Design. Morgan Kaufman (2009)

[78] Montola, M.: Exploring the edge of the magic circle: Defining pervasive games. In: Proc. Of Digital Experience: Design, Aestethics, Practice. (2005)

[79] Sotamaa, O.: All the world's a botfighters stage: Notes on location-based multi-user gaming. In: Computer Games and Digital Cultures, Tampere University Press (2002)

[80] Jackson, S.: Killer - the game of assassination. Steve Jackson Games (1982)

[81] Poremba, C.: Critical potential on the brink of the magic circle. In Akira, B., ed.: Situated Play: Proceedings of the 2007 Digital Games Research Association Conference, Tokyo, The University of Tokyo (September 2007) 772–778

[82] Niemi, J., Sawano, S., Waern, A.: Involving non-players in pervasive games. In: CC '05: Proceedings of the 4th decennial conference on Critical computing, New York, NY, USA, ACM (2005) 137–140

[83] Benford, S., Crabtree, A., Reeves, S., Sheridan, J., Dix, A., Flintham, M., Drozd, A.: The frame of the game: Blurring the boundary between fiction and reality in mobile experiences. In: CHI '06: Proceedings of

the SIGCHI conference on Human Factors in computing systems, New York, NY, USA, ACM (2006) 427–436

[84] Nieuwdorp, E.: The pervasive discourse: an analysis. Comput. Entertain. **5**(2) (2007) 13

[85] McGonigal, J.: his Might Be a Game: Ubiquitous Play and Performance at the Turn of the Twenty-First Century. Dissertation, University of California, Berkeley (2004)

[86] Piekarski, W., Thomas, B.: Arquake: the outdoor augmented reality gaming system. Commun. ACM **45**(1) (2002) 36–38

[87] Crabtree, A., Benford, S., Rodden, T., Greenhalgh, C., Flintham, M., Anastasi, R., Drozd, A., Adams, M., Row-Farr, J., Tandavanitj, N., Steed, A.: Orchestrating a mixed reality game 'on the ground'. In: CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM (2004) 391–398

[88] Grüter, B., Gröning, S., Milszus, W., Oks, M., Schneider, O.: Mobile gaming experience - working on an empirical prototype. workshop user experience - towards a unified view. In: NordiCHI 2006: The Fourth Conference on Human-Computer Interaction. (2006)

[89] McGonigal, J.: This is not a game: Immersive aesthetics and collective play. In: Digital Arts and Culture, DAC 2003, RMIT University (2003)

[90] Lindt, I., Ohlenburg, J., Pankoke-Babatz, U., Ghellal, S.: A report on the crossmedia game epidemic menace. Comput. Entertain. **5**(1) (2007) 8

[91] de Souza e Silva, A.: Alien-revolt: A case study of the first location-based mobile game in brazil. IEEE Technology and Society Magazine **27**(1) (2008) 18–28

[92] Falk, J., Ljungstrand, P., Björk, S., Hansson, R.: Pirates: proximity-triggered interaction in a multi-player game. In: CHI '01: CHI '01 extended abstracts on Human factors in computing systems, New York, NY, USA, ACM (2001) 119–120

[93] Björk, S., Holopainen, J., Ljungstrand, P., Akesson, K.P.: Designing ubiquitous computing games – a report from a workshop exploring ubiquitous computing entertainment. Personal Ubiquitous Comput. **6**(5-6) (2002) 443–458

[94] Walther, B.K.: Notes on the methodology of pervasive gaming. In: Entertainment Computing - ICEC 2005, Volume 3711/2005, Springer Berlin / Heidelberg (2005) 488–495

[95] Brown, D.R.: Pervasive games are not a genre! Technical report, School of Literature, Communication & Culture, Georgia Institute of Technology (2007)

[96] Kiefer, P., Matyas, S., Schlieder, C.: Systematically exploring the design space of location-based games. In: Pervasive 2006 Workshop. (2006) 183–190

[97] Benford, S., Magerkurth, C., Ljungstrand, P.: Bridging the physical and digital in pervasive gaming. Commun. ACM **48**(3) (March 2005) 54–57

[98] Matyas, S., Matyas, C., Mitarai, H., Kamata, M., Kiefer, P., Schlieder, C. In: Designing Location-based Mobile Games - The CityExplorer Use Case. Peter Lang Publishers (Digital Formations series, edited by Steve Jones) (2009) 187–203

[99] Kiefer, P., Matyas, S., Schlieder, C.: State space analysis as a tool in the design of a smart opponent for a location-based game. In Masuch, M., Hartmann, K., Beckhaus, S., Spierling, U., Sliwka, F., eds.: Proceedings of the Computer Science and Magic 2005, GC Developer Conference Science Track, Leipzig, Germany, Messe Leipzig (August 2005) isbn: 3-9804874-3-1.

[100] Rashid, O., Mullins, I., Coulton, P., Edwards, R.: Extending cyberspace: location based games using cellular phones. Comput. Entertain. **4**(1) (2006) 4

[101] Hofmann-Wellenhof, B., Lichtenegger, H., Collins, J.: Global Positioning System. Theory and practice. Springer (2001) 5th Edition.

[102] Benford, S., Crabtree, A., Flintham, M., Drozd, A., Anastasi, R., Paxton, M., Tandavanitj, N., Adams, M., Row-Farr, J.: Can you see me now? ACM Trans. Comput.-Hum. Interact. **13**(1) (2006) 100–133

[103] Drozd, A., Benford, S., Tandavanitj, N., Wright, M., Chamberlain, A.: Hitchers: Designing for cellular positioning. In: Proceedings of Ubicomp 06, Springer Lecture Notes in Computer Science, Vol. 4206, Springer (2006) 279–296

[104] Rudström, A., Höök, K., Svensson, M.: Social positioning: Designing the seams between social, physical and digital space. In: HCII 2005. (2005)

[105] Weiser, M.: The computer for the 21st century. SIGMOBILE Mob. Comput. Commun. Rev. **3**(3) (1999) 3–11

[106] Reid, J., Geelhoed, E., Hull, R., Cater, K., Clayton, B.: Parallel worlds: immersion in location-based experiences. In: CHI '05: CHI '05 extended abstracts on Human factors in computing systems, New York, NY, USA, ACM (2005) 1733–1736

[107] Brown, E., Cairns, P.: A grounded investigation of game immersion. In: CHI '04: CHI '04 extended abstracts on Human factors in computing systems, New York, NY, USA, ACM (2004) 1297–1300

[108] Jegers, K.: Pervasive game flow: understanding player enjoyment in pervasive gaming. Comput. Entertain. **5**(1) (2007) 9

[109] Oppermann, L., Broll, G., Capra, M., Benford, S.: Extending authoring tools for location-aware applications with an infrastructure visualization layer. In: 8th International Conference on Ubiquitous Computing (Ubicomp 2006), New York, NY, USA, Springer (2006)

[110] Nova, N., Girardin, F., Dillenbourg, P.: 'location is not enough!': an empirical study of location-awareness in mobile collaboration. In: WMTE '05: Proceedings of the IEEE International Workshop on Wireless and Mobile Technologies in Education, Washington, DC, USA, IEEE Computer Society (2005) 21–28

[111] Benford, S., Seager, W., Flintham, M., Anastasi, R., Rowl, D., Stanton, D., Bowers, J., T, N., Adams, M., Row, J., Oldroyd, A., Sutton, J.: The error of our ways: The experience of self-reported position in a location-based game. In: Proc. Ubicomp, Springer (2004) 70–87

[112] Jonsson, S., Waern, A.: The art of game-mastering pervasive games. In: ACE '08: Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology, New York, NY, USA, ACM (2008) 224–231

[113] Chalmers, M., Bell, M., Brown, B., Hall, M., Sherwood, S., Tennent, P.: Gaming on the edge: using seams in ubicomp games. In: ACE

'05: Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology, New York, NY, USA, ACM (2005) 306–309

[114] Chalmers, M., Dieberger, A., Höök, K., Rudström, A.: Social navigation and seamful design. Japanese Journal of Cognitive Studies **3**(11) (2004)

[115] Barkhuus, L., Chalmers, M., Tennent, P., Hall, M., Bell, M., Sherwood, S., Brown, B.: Picking pockets on the lawn: The development of tactics and strategies in a mobile game. In: UbiComp 2005: Ubiquitous Computing, Springer Berlin / Heidelberg - Volume 3660/2005 (2005) 358–374

[116] Broll, G., Benford, S., Oppermann, L.: Exploiting seams in mobile phone games. In: Workshop on PerGames at Pervasive Computing. (2006)

[117] Broll, G., Benford, S.: Seamful design for location-based mobile games. In: Entertainment Computing - ICEC 2005, Berlin: Springer (2005) 155–166

[118] Greenhalgh, C., Benford, S., Drozd, A., Flintham, M., Hampshire, A., Oppermann, L., Smith, K., Tycowicz, C.V.: Addressing mobile phone diversity in ubicomp experience development. In: UbiComp'07: Proceedings of the 9th international conference on Ubiquitous computing, Berlin, Heidelberg, Springer-Verlag (2007) 447–464

[119] Saarenpää, H., Korhonen, H., Paavilainen, J.: Asynchronous gameplay in pervasive multiplayer mobile games. In: CHI '09: Proceedings of the 27th international conference extended abstracts on Human factors in computing systems, New York, NY, USA, ACM (2009) 4213–4218

[120] Paelke, V., Oppermann, L., Reimann, C.: Mobile location-based gaming. In: Map-based Mobile Services-Design, Interaction and Usability. Shaker Verlag, Aachen, Germany (November 2008) 310–332 ISBN 978-3-540-37109-0.

[121] Cheok, A.D., Goh, K.H., Liu, W., Farbiz, F., Fong, S.W., Teo, S.L., Li, Y., Yang, X.: Human pacman: a mobile, wide-area entertainment system based on physical, social, and ubiquitous computing. Personal Ubiquitous Comput. **8**(2) (2004) 71–81

[122] Rashid, O., Bamford, W., Coulton, P., Edwards, R., Scheible, J.: Paclan: mixed-reality gaming with rfid-enabled mobile phones. Computers in Entertainment **4**(4) (October 2006)

[123] Steve, B., Martin, F., Adam, D., Rob, A., Duncan, R., Nick, T., Matt, A., Ju, R.F., Amanda, O., Jon, S.: Uncle roy all around you: Implicating the city in a location-based performance. In: ACE04: ACM conference on Advanced Computer Entertainment, ACM (2004)

[124] Struppek, M., Willis, K.S. In: Botfighters - A Game that Surrounds You. Birkhäuser Basel (2007) 226–227

[125] Bjerver, M.: Player behaviour in pervasive games - using the city as game board in botfighters (2006)

[126] Lankoski, P., Heliö, S., Nummela, J., Lahti, J., Mäyrä, F., Ermi, L.: A case study in pervasive game design: the songs of north. In: NordiCHI '04: Proceedings of the third Nordic conference on Human-computer interaction, New York, NY, USA, ACM (2004) 413–416

[127] Flintham, S.K., Benford, S., Capra, M., Green, J., Greenhalgh, C., Wright, M., Adams, M., Tandavanitj, N., Row-Farr, J., Lindt, I.: Day of the figurines: A slow narrative-driven game for mobile phones using text messaging. In: Pergames 2007. (2007)

[128] Bell, M., Chalmers, M., Barkhuus, L., Hall, M., Sherwood, S., Tennent, P., Brown, B., Rowland, D., Benford, S.: Interweaving mobile games with everyday life. In: SIGCHI Conference on Human Factors in Computing Systems, CHI '06, ACM Press (2006) 417–426

[129] Casey, S., Kirman, B., Rowland, D.: The gopher game: a social, mobile, locative game with user generated content and peer review. In: ACE '07: Proceedings of the international conference on Advances in computer entertainment technology, New York, NY, USA, ACM (2007) 9–16

[130] Peitz, J., Saarenpää, H., Björk, S.: Insectopia: exploring pervasive games through technology already pervasively available. In: ACE '07: Proceedings of the international conference on Advances in computer entertainment technology, New York, NY, USA, ACM (2007) 107–114

[131] Flintham, M.: Painting the town red: configuring location-based games by colouring maps. In: ACE '05: Proceedings of the 2005 ACM SIGCHI International Conference on Advances in computer entertainment technology, New York, NY, USA, ACM (2005) 9–18

[132] Oppermann, L., Koleva, B., Benford, S., Watkins, M., Jacobs, R.: Fighting with jelly: user-centered development of wireless infrastructure visualization tools for authoring location-based experiences. In: ACE '08: Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology, New York, NY, USA, ACM (2008) 322–329

[133] Kiefer, P., Matyas, S.: The geogames tool: Balancing spatio-temporal design parameters in location-based games. In Mehdi, Q., Gough, N., eds.: Proceedings of the 7th International Conference on Computer Games: Artificial Intelligence, Animation, Mobile, Educational and Serious Games (CGAMES 2005), Angouleme, France (November 2005) 216–222

[134] Schlieder, C., Kiefer, P., Matyas, S.: Geogames - designing location-based games from classic board games. IEEE Intelligent Systems **21**(5) (Sep/Okt 2006) 40–46

[135] Benford, S., Crabtree, A., Reeves, S., Sheridan, J., Dix, A., Flintham, M., Drozd, A.: Savannah: Designing a location-based game simulating lion behaviour. In: Advanced Computer Entertainment, ACE 2004, ACM Press (June 2004)

[136] Mueller, F., Agamanolis, S.: Sports over a distance. Computers in Entertainment **3**(3) (July 2005)

[137] Mueller, F., Cole, L., O'Brien, S., Walmink, W.: Airhockey over a distance: a networked physical game to support social interactions. In: ACM SIGCHI international Conference on Advances in Computer Entertainment Technology, ACE '06, ACM Press (June 2006) 70

[138] Rabin, S.: AI Game Programming Wisdom. Charles River Media, Inc., Rockland, MA, USA (2002)

[139] Jakl, A.: The workflow of c++ game development. Technical report, Symbian Developer Network, Nokia (2006) http://developer.symbian.com/main/documentation/symbian_cpp/development_tips/ (accessed February 2010).

[140] Kiefer, P., Matyas, S., Schlieder, C.: Playing on a line: Location-based games for linear trips. In Bernhaupt, R., Tscheligi, T., eds.: International Conference on Advances in Computer Entertainment Technology (ACE 2007), Salzburg, Austria, ACM (2007)

[141] Chen, L., Benford, S.: Your way your missions: from location-based to route-based pervasive gaming. In: ACE '07: Proceedings of the international conference on Advances in computer entertainment technology, New York, NY, USA, ACM (2007) 232–233

[142] Güting, R.H.: An introduction to spatial database systems. The VLDB Journal **3**(4) (1994) 357–399

[143] Dewdney, A.K.: The Planiverse-Computer contact with a two-dimensional world. Simon & Schuster (1984)

[144] Bichard, J., Brunnberg, L., Combetto, M., Gustafsson, A., Juhlin, O.: Backseat playgrounds: Pervasive storytelling in vast location based games. In: Proceedings of the 5th International Conference on Entertainment Computing - ICEC 2006. (2006) 117–122

[145] Gustafsson, A., Bichard, J., Brunnberg, L., Juhlin, O., Combetto, M.: Believable environments: generating interactive storytelling in vast location-based pervasive games. In: ACE '06: Proceedings of the 2006 ACM SIGCHI international conference on Advances in computer entertainment technology, New York, NY, USA, ACM (2006) 24

[146] Susi, T., Johannesson, M., Backlund, P.: Serious games - an overview. Technical report, School of Humanities and Informatics, University of Skövde, Sweden (2007)

[147] Schwabe, G., Göth, C.: Mobile learning with a mobile game: design and motivational effects. Journal of Computer Assisted Learning **21**(3) (June 2005) 204–216

[148] Meisenberger, M., Nischelwitzer, A.: The mobile learning engine (mle) - a mobile computer-aided, multimediabased learning application. In: Proc. Multimedia Applications in Education Conference (MApEC). (2004)

[149] Sharples, M., Taylor, J., Vavoula, G.: Towards a theory of mobile learning. Proceedings of MLearn 2005 conference (2005)

[150] Kiefer, P., Matyas, S., Schlieder, C.: Geogames - Integrating Edutainment Content in Location-Based Games. In: Concepts and Technologies for Pervasive Games - A Reader for Pervasive Gaming Research, Vol. 1. Shaker Verlag, Aachen, Germany (November 2007) 127–152 ISBN 978-3-8322-6223-5.

[151] Walz, S.P., Ballagas, R., Borchers, J., Mendoza, J., Kratz, S., Wartmann, C., Fuhr, C., Tann, M.J., Youn Shin, D., Hameed, B., Bardos, L., Hovestadt, L.: Cell spell-casting: Designing a locative and gesture recognition multiplayer smartphone game for tourists. In: PerGames, Third International Workshop on Pervasive Gaming Applications at Pervasive 2006. (May 2006)

[152] Ballagas, R., Walz, S.P., Kratz, S., Fuhr, C., Yu, E., Tann, M., Borchers, J., Ludger, H.: Rexplorer: A mobile, pervasive spell-casting game for tourists. In: CHI '07 extended abstracts on Human factors in computing systems. (2007)

[153] Anderson, I., Maitland, J., Sherwood, S., Barkhuus, L., Chalmers, M., Hall, M., Brown, B., Muller, H.: Shakra: tracking and sharing daily activity levels with unaugmented mobile phones. Mob. Netw. Appl. **12**(2-3) (2007) 185–199

[154] Capra, M., Radenkovic, M., Benford, S., Oppermann, L., Drozd, A., Flintham, M.: The multimedia challenges raised by pervasive games. In: Proceedings of the 13th Annual ACM international Conference on Multimedia – MULTIMEDIA '05, ACM Press (2005) 89–95

[155] Wolff, S., Grüter, B.: Context, emergent game play and the mobile gamer as producer. In: Informatik 2008, 38. Jahrestagung der Gesellschaft für Informatik e.V. (GI), GI-Edition - Lecture Notes in Informatics Vol -133 Bd.1 (September 2008) 495–500

[156] Bell, M., Reeves, S., Brown, B., Sherwood, S., MacMillan, D., Ferguson, J., Chalmers, M.: Eyespy: supporting navigation through play. In: CHI '09: Proceedings of the 27th international conference on Human factors in computing systems, New York, NY, USA, ACM (2009) 123–132

[157] Harzem, L., Hartge, T., Hubatsch, C., Kerney, K., Weber, E., Ziegenbalg, M., Grüter, B.: A serious mobile game for landmark production

- a work in progress report. In: 39. Jahrestagung der Gesellschaft für Informatik 2009 - 2nd Mobile Gaming Workshop, GI (2009)

[158] Schwering, A.: Approaches to semantic similarity measurement for geo-spatial data: A survey. Transactions in GIS **12**(1) (2008) 5–29

[159] Goldstone, R.L.: Similarity. In: MIT encylopedia of the cognitive sciences. MA: MIT Press (1999)

[160] Attneave, F.: Dimensions of similarity. The American Journal of Psychology **63**(4) (1950) 516–556

[161] Gärdenfors, P.: Conceptual Spaces: The Geometry of Thought. MIT Press, Cambridge, MA, USA (2000)

[162] Gärdenfors, P.: Conceptual spaces as a framework for knowledge representations. Mind and Matter **2**(2) (2004) 9–27

[163] Schwering, A., Raubal, M.: Measuring semantic similarity between geospatial conceptual regions. In: GeoSpatial Semantics, Springer (2005) 90–106

[164] Eckes, T. und Rossbach, H.: Clusteranalysen. Kohlhammer (1980)

[165] de Smith, M.J., Goodchild, M.F., Longley, P.A.: Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools. Troubador Publishing (2007)

[166] Kang, J.H., Welbourne, W., Stewart, B., Borriello, G.: Extracting places from traces of locations. SIGMOBILE Mob. Comput. Commun. Rev. **9**(3) (2005) 58–68

[167] Volz, S.: An iterative approach for matching multiple representations of street data. In: ISPRS Workshop on Multiple Representation and Interoperability of Spatial Data. (2006)

[168] Sayda, F.: Involving lbs users in data acquisition and update. In: AGILE 2005: Conference on geographic information science, Universidade Nova de Lisboa (2005)

[169] Walter, V., Fritsch, D.: Matching spatial data sets: a statistical approach. International Journal for Geographical Information Science **13**(5) (1999) 445–473

[170] Tversky, A.: Features of similarity. Psychological Review **84** (1977) 327–352

[171] Tversky, A.: Preference, Belief, and Similarity: Selected Writings. MIT Press (2004)

[172] Rodriguez, A., Egenhofer, M.: Comparing spatial entity classes: An asymmetric and context-dependent similarity measure. International Journal of Geographical Information Science **18**(3) (2004) 229–256

[173] Rodríguez, M.A., Egenhofer, M.J.: Determining semantic similarity among entity classes from different ontologies. IEEE Transactions on Knowledge and Data Engineering **15**(2) (2003) 442–456

[174] Miller, G.A.: Wordnet: a lexical database for english. Commun. ACM **38**(11) (1995) 39–41

[175] Miller, G.A., Hristea, F.: Wordnet nouns: Classes and instances. Comput. Linguist. **32**(1) (2006) 1–3

[176] D'Ulizia, A., Ferri, F., Formica, A., Grifoni, P., Rafanelli, M.: Structural similarity in geographical queries to improve query answering. In: SAC '07: Proceedings of the 2007 ACM symposium on Applied computing, New York, NY, USA, ACM (2007) 19–23

[177] Bernstein, A., Kaufmann, E., Bürki, C., Klein, M.: Object similarity in ontologies: A foundation for business intelligence systems and high-performance retrieval. In: Twenty-Fifth International Conference on Information Systems. (December 2004) 11–25

[178] Bernstein, A., Kaufmann, E., Bürki, C., Klein, M.: How similar is it? towards personalized similarity measures in ontologies. In: 7. Internationale Tagung Wirtschaftsinformatik. (February 2005) 1347–1366

[179] Bernstein, A., Kaufmann, E., Kiefer, C., Bürki, C.: Simpack: A generic java library for similarity measures in ontologies. Technical report, University of Zurich, Department of Informatics, Winterthurerstrasse 190, 8057 Zurich, Switzerland (August 2005)

[180] Rahm, E., Bernstein, P.A.: On matching schemas automatically. Technical report, Dept. of Comp. Science, Univ. of Leipzig (2001)

[181] Rahm, E., Bernstein, P.A.: A survey of approaches to automatic schema matching. The VLDB Journal The International Journal on Very Large Data Bases **10**(4) (2001) 334–350

[182] Markman, A.B., Gentner, D.: Structural alignment during similarity comparisons. Cognitive Psychology **25** (1993) 431–467

[183] Goldstone, R.L.: Similarity, interactive activation, and mapping. Journal of Experimental Psychology: Learning, Memory, and Cognition **20** (1994) 3–28

[184] Imai, S.: Pattern similarity and cognitive transformations. Acta Psychologica **41**(6) (1977) 433 – 447

[185] A, U.H., B, N.C., A, L.B.R.: Similarity as transformation. Cognition **87** (2003) 1–32

[186] Jones, C.B., Alani, H., Tudhope, D.: Geographical information retrieval with ontologies of place. In: COSIT 2001: Proceedings of the International Conference on Spatial Information Theory, London, UK, Springer-Verlag (2001) 322–335

[187] Tudhope, D., Taylor, C.: Navigation via similarity: Automatic linking based on semantic closeness. Information Processing & Management **33**(2) (1997) 233 – 242

[188] Hastings, J.T.: Automated conflation of digital gazetteer data. Int. J. Geogr. Inf. Sci. **22**(10) (2008) 1109–1127

[189] Gösseln, G., Sester, M.: Semantic and geometric integration of geoscientific data sets with atkis - applied to geo-objects from geology and soil science. In: ISPRS Commission IV Joint Workshop: Challenges in Spatial Analysis, Integration and Visualization II, ISPRS (2003) 111–116

[190] Gösseln, G.v., Sester, M.: Integration of geoscientific data sets and the german digital map using a matching approach. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences **35** (2004)

[191] Hess, C., Schlieder, C.: Trust-based recommendations for documents. AI Commun. **21**(2-3) (2008) 145–153

[192] O.K., F., E.J., S.: Grundlagen der Wirtschaftsinformatik - Konzepte, Modelle und Methoden. Oldenbourg, München, München, Germany (1993)

[193] Gillman, D.W.: Triangulations for rubber-sheeting. In: Proceedings of AUTOCARTO Bd. 7. (1985)

[194] Hartigan, J.A.: Consistency of single linkage for high-density clusters. Journal of the American Statistical Association **76**(374) (1981) 388–394

[195] Budanitsky, A., Hirst, G.: Semantic distance in wordnet: An experimental, application-oriented evaluation of five measures. In: Workshop on WordNet and Other Lexical Resources, in the North American Chapter of the Association for Computational Linguistics (NAACL-2001). (2001)

[196] Hirst, G., St-Onge, D.: Lexical chains as representations of context for the detection and correction of malapropisms. In Fellbaum, C., ed.: WordNet: An Electronic Lexical Database, MIT Press (1998) 305–332

[197] Jiang, J., Conrath, D.: Semantic similarity based on corpus statistics and lexical taxonomy. In: Proc. of the Int'l. Conf. on Research in Computational Linguistics. (1997) 19–33

[198] Lin, D.: An information-theoretic definition of similarity. In: Proceedings of the Fifteenth International Conference on Machine learning (ICML-98). (1998)

[199] Leacock, C., Chodorow, M. In: Combining local context and WordNet similarity for word sense identification. In C. Fellbaum (Ed.), MIT Press (1998) 305–332

[200] Ross, S.: First Course in Probability, A (8th Edition). 8 edn. Prentice Hall (2008)

[201] Pipino, L.L., Lee, Y.W., Wang, R.Y.: Data quality assessment. Commun. ACM **45**(4) (2002) 211–218

[202] Wand, Y., Wang, R.Y.: Anchoring data quality dimensions in ontological foundations. Commun. ACM **39**(11) (1996) 86–95

[203] Tayi, G.K., Ballou, D.P.: Examining data quality. Commun. ACM **41**(2) (1998) 54–57

[204] Strong, D.M., Lee, Y.W., Wang, R.Y.: Data quality in context. Commun. ACM **40**(5) (1997) 103–110

[205] Orr, K.: Data quality and systems theory. Commun. ACM **41**(2) (1998) 66–71

[206] Navratil, G.: Influences of Technology, Law, and Usability on Data Quality. GeoInfo Series Vienna (Ed. Andrew U. Frank), Institute for Geoinformation and Cartography (1995)

[207] Organization, I.S.: Geographic Information - Quality Principles. tc211. Number 19113 edn. (2002)

[208] Devillers, R., Gervais, M., Bedard, Y., Jeansoulin, R.: From metadata to quality indicators and contextual end-user manual. In: OEEPE/ISPRS Joint Workshop on Spatial Data Quality Management. (2002)

[209] Wang, F., Mäs, S., Reinhardt, W., Kandawasvika, A.: Ontology based quality assurance for mobile data acquisition. In: 19th international conference on Informatics for Environmental Protection: Networking Environmental Information - Proceedings. (2005)

[210] Wang, R.Y., Storey, V.C., Firth, C.P.: A framework for analysis of data quality research. IEEE Transactions on Knowledge and Data Engineering **7**(4) (1995) 623–640

[211] Wache, H., Vögele, T., Visser, U., Stuckenschmidt, H., Schuster, G., Neumann, H., Hübner, S.: Ontology-based integration of information - a survey of existing approaches. In: Proceedings of IJCAI-01 Workshop: Ontologies and Information Sharing. (2001) 108–117

[212] Francalanci, C., Pernici, B.: Data quality assessment from the user's perspective. In: IQIS '04: Proceedings of the 2004 international workshop on Information quality in information systems, New York, NY, USA, ACM (2004) 68–73

[213] Henriksson, R., Marjaana, K.T.: An ontology-driven approach for spatial data quality evaluation. In: Spatial Data Quality 2007: Proceedings of the 2007 5th International Symposium on Spatial Data Quality. (2007)

[214] Devillers, R., Jeansoulin, R.: Spatial Data Quality. ISTE, Geographical Information Systems Series, Newport Beach, USA (2006)

[215] Matyas, S.: Collaborative spatial data acquisition - a spatial and semantic data aggregation approach. In: Proceedings of the 10th AGILE International Conference on Geographic Information Science 2007. (2007)

[216] Egenhofer, J.M.: Spatial sql: A query and presentation language. IEEE Transaction and Data Engineering **6**(1) (1994) 86–95

[217] Matyas, S., Matyas, C., Schlieder, C., Kiefer, P., Mitarai, H., Kamata, M.: Designing location-based mobile games with a purpose: collecting geospatial data with cityexplorer. In: ACE '08: Proceedings of the 2008 International Conference on Advances in Computer Entertainment Technology, New York, NY, USA, ACM (2008) 244–247

[218] Fahrmeir, L., Tutz, G.: Multivariate Statistical Modelling Based on Generalized Linear Models. Springer, New York (1994)

[219] R Development Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. (2009) ISBN 3-900051-07-0.

[220] Geiss, J.: Creating a gold standard for sentence clustering in multi-document summarization. In: ACL-IJCNLP '09: Proceedings of the ACL-IJCNLP 2009 Student Research Workshop, Morristown, NJ, USA, Association for Computational Linguistics (2009) 96–104

[221] Zhao, Y., Karypis, G.: Evaluation of hierarchical clustering algorithms for document datasets. In: CIKM '02: International Conference on Information and Knowledge Management, New York, NY, USA, ACM (2002) 515–524

[222] Matyas, S., Wullinger, P., Matyas, C.: Building geospatial data collections with location-based games. In: Proceedings of the 32nd Annual German Conference on AI, LNAI 5803, Springer Verlag (2009) 703–710

[223] Küpper, A.: Location-based Services: Fundamentals and Operation. John Wiley & Sons (2005)

[224] Benford, S., Giannachi, G.: Temporal trajectories in shared interactive narratives. In: CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM (2008) 73–82

[225] Fischer, J.E., Benford, S.: Inferring player engagement in a pervasive experience. In: CHI '09: Proceedings of the 27th international conference on Human factors in computing systems, New York, NY, USA, ACM (2009) 1903–1906

[226] Anderson, E.F.: Playing smart - artificial intelligence in computer games. In: Proceedings of zfxCON03 Conference on Game Development. (2003)

[227] Doyle, P.: Virtual intelligence from artificial reality: Building stupid agents in smart environments. In: AAAI Spring Symposium: Artificial Intelligence and Computer Games. (1999)

UNIVERSITY OF BAMBERG PRESS

Lokalisierungstechnologie (z.B. GPS)stark gestiegen. Jeder, der ein solches Gerät besitzt, kann zum Datenlieferanten werden und in einem Netzwerk von Freiwilligen, die Rolle eines Datensensors übernehmen. In der vorliegenden Dissertation werde ich zwei Problemstellungen näher betrachten, die im Kontext von raumbezogenen Daten, erhoben von Gemeinschaften aus fachlich (hier: geowissenschaftlich) nicht geschulten Nutzern, noch ungelöst sind:
Wie können geowissenschaftliche Laien motiviert werden, freiwillig qualitätsgesicherte raumbezogene Daten zu erheben?
Der erste Beitrag meiner Dissertation liefert eine Antwort auf diese Frage. Er beinhaltet ein Framework zum Design von ortsbezogenen Spielen, die geowissenschaftliche Laien motivieren, raumbezogene Daten qualitätsgesichert in spielerischer Art und Weise zu sammeln. Um es mit den Worten von Peltola et al. (2006) auszudrücken: "game play is a natural motivator to participate in something that is not necessary or beneficial. [...] By controlling game events and perhaps game logics and rules, the agencies that ultimately use the gathered data, can steer players to do tasks supporting their needs".
Wie kann die Qualität von Sammlungen semantisch angereicherter raumbezogener Daten verbessert werden?
Eine wichtige Eigenschaft von ortsbezogenen Spielen zur Datenerhebung ist die wiederholte Spielbarkeit. Im Gegensatz zu anderen Domänen ist nämlich die Erstellung redundanter Daten im raumbezogenen Kontext sogar erwünscht. Das Zusammenführen von Daten mehrerer Nutzer kann man als Mehrfachmessung auffassen, die sich dazu nutzen lässt, die Qualität von Lokalisierung (Wo?) und Kategorisierung (Was?) zu verbessern. Der zweite Beitrag meiner Dissertation gibt eine Antwort auf die Fragestellung und besteht in einem Ansatz zur semantischen Integration der gesammelten raumbezogenen Daten.

ISBN 978-3-86309-002-9
ISSN 1867-7401
22,50 Euro