
Automatisierte Semantische Harmonisierung von Landnutzungsdaten

Mark M. HALL¹

Zusammenfassung

Spatial Data Infrastructures erleichtern den Zugriff auf Daten aus verteilten Datenquellen. Um diese Kombination von Datenquellen zu ermöglichen, müssen diese Datenquellen syntaktisch und semantisch harmonisiert werden. Für die semantische Harmonisierung ist es notwendig, dass ein semantisches Ähnlichkeitsmaß definiert wird, das den Vergleich zweier Konzepte möglich macht. Diese Arbeit beschreibt ein derartiges Ähnlichkeitsmaß und wie es auf den Bereich der Landnutzungsdaten angewandt werden kann.

1 Einleitung

Wiederverwendung und Verbreitung von Informationen sind wichtige Ziele beim Praxiseinsatz Geographischen Informationssystem. Eine Möglichkeit diese Ziele zu erreichen sind Spatial Data Infrastructures. Diese bieten leichten Zugriff auf verteilt gespeicherte Daten. Dazu ist es aber notwendig, dass die GI-Systeme aus denen die Spatial Data Infrastructure bestehen miteinander kommunizieren können. Dies wiederum setzt eine Harmonisierung der Datenformate, kartographischen Projektionen, Geometrien und der Inhalte der einzelnen GI-Systeme voraus.

Ein Projekt das diese Ziele verfolgt ist das ISAMAP Projekt², welches grenzüberschreitende Planung in den drei Ländern Friuli-Venezia Giulia (Italien), Slowenien und Kärnten (Österreich) ermöglichen und unterstützen soll. Im Rahmen dieses Projektes beschäftigt sich das an der Universität Klagenfurt durchgeführte Subprojekt HarmonISA (Mandl et al. 2006) mit der semantischen Harmonisierung von Landnutzungs- und Landbedeckungsdaten. Die Notwendigkeit der semantischen Harmonisierung ergibt sich, weil alle drei beteiligten Länder unterschiedliche Landnutzungskataloge entwickelt haben, die die vorhandenen Landnutzungen unterschiedlich abbilden. Die Gründe für diese Unterschiede sind das verschiedene Personen aufgrund unterschiedlicher Ausgangsdaten, mit unterschiedlichen Zielen die Landnutzungskataloge entworfen haben. Dies führt zu Verschiedenheiten in der Struktur, in der Schwerpunktsetzung und im Detaillierungsgrad der Modellierung. Um die Kataloge zu harmonisieren müssen diese semantischen Unterschiede erkannt und definiert werden.

Der Rest dieser Arbeit ist folgend aufgeteilt: Abschnitt 2 beschäftigt sich mit bestehenden Harmonisierungsverfahren, Abschnitt 3 beschreibt das im Rahmen des HarmonISA Projekts verwendete Ähnlichkeitsmodell, Abschnitt 4 beschreibt die Implementierung und wie

¹ mhall@edu.uni-klu.ac.at

² Ein Projekt im INTERREG IIIB CADSES framework, <http://www.isamap.info>

das Projekt in eine Spatial Data Infrastructure integriert werden kann und Abschnitt 5 enthält Schlussfolgerungen und Ausblicke.

2 Harmonisierungsverfahren

Harmonisierung bzw. Integration von Daten oder Schemata ist kein neues Problem. In jeder Situation in der heterogene Datenquellen zusammen zu führen sind, muss irgendein Harmonisierungsverfahren angewandt werden. Bisher wurden derartige Verfahren weitgehend manuell oder semi-automatisch durchgeführt, es gibt aber auch einige Verfahren für voll-automatische Harmonisierung. Diese Verfahren wurden mit leicht unterschiedlichen Zielsetzungen entwickelt und haben daher bestimmte Stärken und Schwächen. Primär werden diese Algorithmen danach klassifiziert, ob sie auf der Schema Ebene arbeiten oder auf der Instanz Ebene. In dieser Arbeit werden nur Schema Ebene Algorithmen betrachtet, da das die Ebene ist auf der innerhalb des HarmonISA Projekts die Harmonisierung durchgeführt wird.

Es gibt folgenden großen Gruppen an Algorithmen um Schemata zu matchen

- *Lexikalische Algorithmen* arbeiten auf der Buchstaben Ebene und errechnen die Ähnlichkeit zweier Konzepte aus der Anzahl an unterschiedlichen Buchstaben der Namen der Konzepte (LEVENSHEIN 1966).
- *Wörterbuch basierende Algorithmen* nutzen die in Wörterbüchern wie WordNet (MILLER 1995) vorhandenen Beziehungen um die Ähnlichkeit zweier Konzepte zu berechnen (CASTANO et al. 2005).
- *Strukturbasierte Algorithmen* verwenden Informationen über die Struktur der Schemata um die Ähnlichkeit der einzelnen Konzepte zu errechnen (MAEDCHE et al. 2002).
- *Description Logics Algorithmen* nutzen automatisierte Schlussfolgerungsalgorithmen um Konzepte eines Schemas in ein anderes einzufügen und aufgrund der Einfügeposition die Ähnlichkeit zu berechnen (LI et al. 2004).

Die generelle Problematik all dieser Ansätze ist, dass die Ähnlichkeitsberechnung nicht auf semantischen Prinzipien basiert. Die Algorithmen liefern zwar teilweise brauchbare bis sehr gute Prinzipien, doch es ist nicht gewährleistet, dass die Ergebnisse der Ähnlichkeitsberechnung dem entsprechen, was ein Mensch als Ähnlichkeit empfinden würde.

2.1 Semantische Modelle

Um das Problem der Semantik zu beheben ist es notwendig ein Modell als Basis zu haben, das kognitive sinnvoll ist und so eine gute Übereinstimmung mit der menschlichen Perzeption von Ähnlichkeit bietet. Drei dieser Modelle sind

- *Feature Modell*. Im Feature Modell werden Konzepte durch Listen von Attributen definiert (TVERSKY 1977 und RODRÍGUEZ et al. 2004)). Der Vorteil dieses Modells ist die Einfachheit sowohl in der Erstellung als auch im Vergleich von Konzepten, da hier das Verhältnis der Anzahl gemeinsamer Attribute zu der gesamten Anzahl

an Attributen berechnet wird. Der Nachteil ist das die Beschränkung auf eine Liste das Abbilden komplexerer Sachverhalte unmöglich ist.

- *Cognitive Spaces*. Das Cognitive Spaces Modell (GÄRDENFORS 2000) bildet Konzepte durch Hyperwürfel ab, wobei jedes Attribut als eigene Dimension modelliert wird. Der Vorteil dieses Modells ist, dass es sehr nahe der menschlichen Vorstellung von Konzepten kommt und dass aufgrund der geometrischen Form des Modells die Ähnlichkeit als Distanz errechnet werden kann. Der Nachteil ist, dass es nicht klar ist, wie die Dimensionen für die einzelnen Attribute aussehen.
- *Network Modell*. Das Network Modell (RADA 1989) ordnet alle Konzepte in einer Baum- oder Graphstruktur an. Der Vorteil hier ist wiederum die einfache Berechnung von Ähnlichkeiten durch Zählen der Kanten, die von einem Konzept zu einem anderen traversiert werden müssen. Der Nachteil dieses Modells ist, dass die einzelnen Konzepte nur als Knoten realisiert werden können. Etwaige komplexere innere Strukturen von Konzepten können nicht abgebildet werden.

3 Semantische Harmonisierung mittels Ontologien

Zur semantischen Harmonisierung wurde ein semantisches Modell entwickelt, auf das ein semantisches Ähnlichkeitsmaß definiert wurde. Damit dieses semantische Modell in einem automatisierten Harmonisierungssystem verwendet werden kann, muss es in eine elektronisch verwertbare Form gebracht werden. Dazu werden Ontologien eingesetzt. Eine Ontologie ist eine formale Spezifikation des Wissens einer Domäne und kann zur Abbildung und Verbreitung von Wissen genutzt werden (GRUBER 1993). Um derartige Ontologien zu implementieren gibt es viele verschiedene Ontologiesprachen. Im HarmonISA Projekt wurde aus folgenden Gründen die Web Ontology Language (OWL) (PATEL-SCHNEIDER et al. 2003) verwendet

- Eines der Ziele des HarmonISA Projekts ist eine möglichst standardisierte und offene Lösung. OWL ist die offizielle Empfehlung der W3C³ für eine Ontologiesprache und daher für diesen Zweck perfekt geeignet.
- OWL hat einen Teil seiner Wurzeln im Gebiet der Description Logics. Dies macht es möglich auf sehr einfache Art automatisierte Schlussfolgerungen auf OWL Ontologien durchzuführen (HORROCKS et al. 2003 und BAADER et al. 2003).
- Für OWL gibt es mit dem OWL Plugin⁴ für das Protégé⁵ Projekt eine sehr mächtige Ontologieentwicklungsumgebung. Dies ist besonders wichtig, da langfristig die Entwicklung und Wartung der Ontologien auch von Personen durchgeführt werden soll, deren Spezialisierung nicht im Gebiet des Knowledge Engineering liegt.

OWL bildet die Grundlage auf der die Verarbeitung der semantischen Informationen im HarmonISA System aufbaut.

³ <http://www.w3c.org> – Das World Wide Web Consortium (W3C) entwickelt interoperable Technologien (Spezifikationen, Richtlinien, Software und Werkzeuge) um das volle Potential des Web zu ermöglichen.

⁴ <http://protege.stanford.edu> – Protégé ist ein frei zugänglicher, open source Ontologieeditor.

⁵ <http://protege.stanford.edu/plugins/owl/> – Das OWL Plugin ist eine Erweiterung für den Protégé Editor welches das Editieren von OWL Ontologien ermöglicht.

3.1 Semantisches Modell

Um die Harmonisierung auf semantisch-kognitiv sinnvolle Beine zu stellen, wurde ein kognitiv sinnvolles Modell als Basis entwickelt, welches ein Hybridmodell aus Feature und Network Modell ist.

3.1.1 Hybrides Modell

Um die Schwächen der einzelnen Modelle auszugleichen wurde ein hybrides Modell entwickelt, welches auf den Feature und Network Modellen aufbaut. Angelehnt an das Network Modell, bildet die so genannte Skelett Ontologie die Grundlage auf die die Ontologien der Landnutzungskataloge aufbauen. In der Skelett Ontologie werden primitive Konzepte⁶ in mehreren unabhängigen Baumstrukturen angeordnet, wobei die einzelnen primitiven Konzepte primär aufgrund des Prinzips der Subsumption angeordnet werden. Dies bedeutet, dass wenn der Baum von den Blättern zur Wurzel durchlaufen wird, immer eine „ist-ein“ Beziehung zwischen zwei untereinander angeordneten Konzepten gilt.

Dadurch entsteht eine sehr saubere und einfache Struktur, die auch den Vorteil hat, dass das abgebildete Wissen eher als allgemeingültig akzeptiert wird und stabiler bleibt.

Wälder ≡
 hatOberflaeche.Ve
 getation ∧
 hatVegetation.Bau

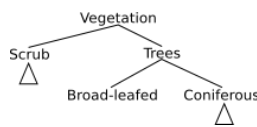


Abbildung 1 zeigt einen Ausschnitt des Baumes der die verschiedenen Vegetationsarten abbildet. Man sieht ganz klar, dass die Hierarchie auf ist-ein Beziehungen aufbaut und je tiefer man in den Baum eintaucht, desto spezieller werden die Konzepte.

Für jede der Baumstrukturen in der Skelett Ontologie werden mindestens eine oder mehrere Eigenschaften definiert, bei der Vegetation gibt es dann also die Eigenschaft *hatVegetation*. Der gültige Wertebereich der Eigenschaften wird auf das Wurzelement der Baumstruktur gesetzt, wodurch auch alle

Subklassen gültige Werteausprägungen sind. Diese Eigenschaften können dann verwendet werden, um die komplexeren Konzepte, welche die Landnutzungskategorien beschreiben, zu definieren.

Für jeden Landnutzungskatalog wird eine eigene Ontologie erstellt. Diese Ontologie importiert die Skelett Ontologie und erhält dadurch Zugriff auf alle dort definierten Eigenschaften und deren mögliche Werte. Um die Semantik der einzelnen Landnutzungskategorien einzufangen wird für jede Landnutzungskategorie ein Konzept in der Ontologie angelegt. Diese werden über eine Menge von hinreichend und notwendigen Eigenschaften definiert. Diese Eigenschaften sind quantifizierte Restriktionen auf die Wertebereiche der in der Skelett Ontologie definierten Eigenschaften.

Abbildung 2 zeigt wie die zwei Landnutzungskategorien „Wälder“ und „Nadelwälder“ aus dem CORINE Landnutzungskatalog abgebildet werden. Man sieht, dass die Definitionen des spezifischen Konzeptes „Nadelwälder“ eine Spezialisierung der Definitionen des abstrakteren Konzeptes „Wälder“ sind. Diese Eigenschaft kann durch ein automatisiertes Schlußfolgerungssystem ausgenutzt werden, um die Hierarchie der in der

⁶ Primitive Konzepte sind jene Konzepte in einer Ontologie, die nur durch notwendige Eigenschaften definiert werden.

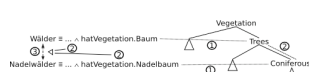
Landnutzungsontologie definierten Landnutzungskategorien zu errechnen. Der Vorteil hiervon ist, dass diese Hierarchie nicht extra gewartet werden muss und dadurch Änderungen an der Landnutzungsontologie leichter durchgeführt werden können.

Diese stark an die an RECTOR 2003 angelehnte Art das Wissen zu spezifizieren hat zwei große Vorteile

- Explizite Spezifikation des in den Landnutzungskatalogen vorhandenen Wissens. Durch die Art das Wissen in primitive Konzepte und definierte Konzepte zu trennen wird die Menge an explizitem Wissen maximiert.
- Automatische Erzeugung der Hierarchie der Landnutzungskategorien. Mit Hilfe von Description Logics Reasonern (RACER (HAARSLEV et al. 2001), FaCT (HORROCKS 1998), Pellet (SIRIN 2004)) kann aufgrund der notwendig und hinreichenden Eigenschaften die Hierarchie der Klassen berechnet werden. Der Vorteil davon ist, dass die Klassen in der Ontologie des Landnutzungskatalogs in einer flachen Liste verwaltet werden können und der Bearbeiter sich nicht um die Verwaltung der Hierarchie kümmern muss.
- Erleichterung der Wartung der einzelnen Ontologien, durch klarere Strukturen

3.2 Semantisches Ähnlichkeitsmaß

Aufbauend auf dieses Modell wurde ein semantisches Ähnlichkeitsmaß definiert. Das



Ähnlichkeitsmaß ermöglicht es zwei Konzepte zu vergleichen und anhand ihrer Definitionen die semantische Ähnlichkeit zu errechnen. Der Algorithmus ist ein hybrider Algorithmus in dem die Ergebnisse aus dem ersten Teilalgorithmus die Basis für die Berechnungen des zweiten Teilalgorithmus sind. Dabei läuft der Ähnlichkeitsvergleich immer nur in eine Richtung, also vom Quellkatalog zum Zielkatalog ab.

- (1) Wertebereiche der Eigenschaft im jeweiligen Baum der Skelett Ontologie finden

- (2) Ähnlichkeit der

Der Ablauf des Ähnlichkeitsalgorithmus ist folgendermaßen

1. Die Definitionen der zwei Landnutzungskategorien, deren Ähnlichkeit berechnet werden soll, werden aufgrund der Namen der Eigenschaften verglichen.
2. Definitionen, die nur in einer der zwei Landnutzungskategorien vorkommen, werden nicht weiterbearbeitet und die Ähnlichkeit wird einfach für jede fehlende Definition um einen konstanten Wert reduziert.
3. Bei Definitionen, deren Eigenschaften in beiden Landnutzungskategorien vorkommen, wird die Ähnlichkeit aus der Ähnlichkeit ihrer Wertebereiche errechnet.
4. Schlussendlich werden die Ähnlichkeiten der einzelnen Definitionen gewichtet, aggregiert und auf den Bereich 0 bis 1 normiert.

In den Fällen, wo eine Eigenschaft in beiden Landnutzungskategorien als Definition vorkommt wird die Ähnlichkeit aus dem Wertebereichen errechnet. Dabei werden vier Fälle unterschieden

- Beide Wertebereiche bestehen aus jeweils einem einzelnen primitiven Konzept. In diesem Fall ist die Ähnlichkeit abhängig von der Distanz der zwei primitiven Konzepte in der Skelett Ontologie wie in Abbildung 3 gezeigt.
- Der Wertebereich der Quellkategorie ist ein einzelnes primitives Konzept und der der Zielkategorie ist eine Vereinigung mehrerer primitiver Konzepte. Die Ähnlichkeit ist abhängig von der Distanz zwischen dem primitiven Konzept der Quellkategorie und dem nächsten primitiven Konzept der Zielkategorie

- Der Wertebereich der Quellkategorie ist eine Vereinigung mehrerer primitiver Konzepte und der der Zielkategorie ein einzelnes primitives Konzept. Die Ähnlichkeit wird gleich dem vorherigen Fall berechnet, aber niedriger gewichtet, da hier eine gewisse Asymmetrie gegeben ist.
- Beide Wertebereiche bestehen aus Vereinigungen primitiver Konzepte. Die Ähnlichkeit ist proportional zur Menge an überlappenden primitiven Konzepten.

Dieser Algorithmus wird dann für alle möglichen Kombinationen durchgeführt, die Quellkategorie wird mit allen Kategorien des Zielkatalogs verglichen. Aus den Vergleichsergebnissen wird jenes ausgewählt, das die höchste Ähnlichkeit hat. Das Quellkonzept wird der Liste an äquivalenten Konzepten des Zielkonzeptes hinzugefügt.

Dieser Vorgang wird für alle Konzepte des Quellkatalogs wiederholt und danach kann man im Zielkatalog nachsehen, welche der Konzepte aus dem Quellkatalog zu welchen im Zielkatalog dazupassen. Um ein Mapping in beide Richtungen zu erhalten, müssen Quell- und Zielkatalog vertauscht und der gesamte Prozess wiederholt werden.

Die Mappings, die so erzeugt wurden, können dann auf die eigentlichen Landnutzungsdaten angewandt werden, um so die Daten vom einen Landnutzungskatalog in den anderen überzuführen.

4 Umsetzung

Das HarmonISA Projekt involviert nicht nur die Definition des semantischen Ähnlichkeitsmaßes, sondern auch die Implementierung dieses in einer Anwendung, die es dem Endbenutzer möglich macht, einfach auf die harmonisierten Daten zuzugreifen (<http://harmonisa.uni-klu.ac.at>) und sie in eine Spatial Data Infrastructure zu integrieren. Zu diesem Zweck wurde eine Web Anwendung entwickelt, die einen Zugang zu den Mappings zwischen den Ontologien ermöglicht, aber auch Zugriff auf die harmonisierten Daten hat.

Der Grund für die Implementierung als Web Anwendung war, dass der Zugriff auf die Ergebnisse für den Endanwender möglichst einfach und ohne größeren Aufwand für den Endanwender möglich sein sollte. Die HarmonISA Anwendung bietet drei Schnittstellen

- *Web Anwendung.* Die Web Anwendung bildet die primäre Schnittstelle in das HarmonISA System. Dies ist die einzige Schnittstelle bei der der Benutzer sowohl auf die Landnutzungs- als auch auf die Ähnlichkeitsinformationen zugreifen kann.
- *OGC WFS Schnittstelle.* Die WFS Schnittstelle erlaubt den Zugriff auf die einzelnen Polygone, die die Datengrundlage bilden. Für jede Ontologie die innerhalb des Systems vorhanden ist, wird ein Layer angeboten über den man auf die Landnutzungsdaten, nach der jeweiligen Ontologie klassifiziert, zugreifen kann. Die Harmonisierung der Daten ist dabei nur auf der Klassifizierungsebene durchgeführt worden, die Polygone die die Datengrundlage bilden, sind nicht bereinigt worden. Es kann also sein, dass ein zusammenhängendes Gebiet aus mehreren Polygonen besteht, die alle dem gleichen Konzept zugeordnet sind.
- *OGC WMS Schnittstelle.* Die WMS Schnittstelle erlaubt den gleichen Zugriff wie die WFS Schnittstelle mit den gleichen Einschränkungen. Der Unterschied ist, dass als Ergebnis eine fertige Karte geliefert wird und nicht nur die Rohdaten.

Je nach Ziel des Endanwenders muss eine der Schnittstellen ausgewählt werden. Wenn die harmonisierten Daten in eine SDI eingebunden werden sollen, dann bieten sich die zwei OGC konformen Schnittstellen an. Der Vorteil der WFS Schnittstelle ist, dass mit den Daten leichter und flexibler weitergearbeitet werden kann. Das Problem ist, dass die Datenmenge der in GML kodierten Daten circa 100 MB beträgt, bei der Übertragung über das Internet also Performanceprobleme auftreten können. Die WMS Schnittstelle umgeht dieses Problem durch die Ausgabe in Kartenform, andererseits ist hier die Nachbearbeitung schwerer

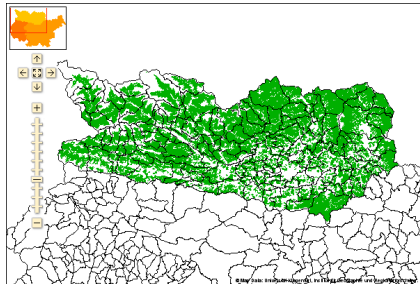


Abb. 5: CORINE Kategorie

Ein Nachteil beider OGC Schnittstellen ist, dass alles Wissen über die Herkunft und die Ähnlichkeitsberechnungen die zu den harmonisierten Daten geführt haben, bei der Transformation über die OGC Schnittstelle verloren gehen. Um auf diese Informationen zuzugreifen, muss die Web-Anwendung verwendet werden.

Die Web-Anwendung ermöglicht den Zugriff auf die harmonisierten Daten und ihre Ähnlichkeitsinformationen. Dieser Zugriff auf die Informationen der bereits harmonisierten Daten erfolgt über zwei Wege. Erstens kann der Endbenutzer über den Katalog Viewer (Abbildung 4) auf die Ergebnisse der Harmonisierung zugreifen. In drei Detailstufen kann untersucht werden, welche Landnutzungskategorien aus den Landnutzungskatalogen der Daten welchen Kategorien des aktuell ausgewählten Katalogs zugeordnet sind. Auch auf die detaillierten Ergebnisse der Ähnlichkeitsberechnung kann zugegriffen werden. Zusätzlich wird eine Karte des Gebiets generiert (Abbildung 5), in der die Ergebnisse der Harmonisierung angezeigt werden.

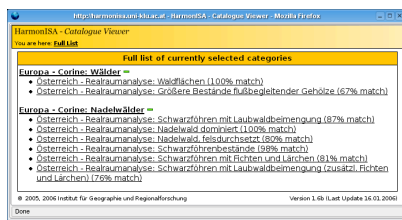


Abb. 4: CORINE Landnutzungskategorien

Für etwaige Analyse eines bestimmten Gebietes oder Landnutzung kann der Benutzer den gewünschten Landnutzungskatalog auswählen und festlegen welche der Landnutzungskategorien er angezeigt bekommen möchte. Außerdem kann in der Karte navigiert und gezoomt werden, um nur einen gewissen Bereich zu betrachten.

Zusätzlich wird an einer semantischen Abfragemöglichkeit gearbeitet, die es dem Endbenutzer erlaubt die semantischen Informationen zu spezifizieren, nach denen er suchen will. Diese Informationen werden mittels des Ähnlichkeitsmaßes mit den existierenden Landnutzungsontologien verglichen und die Ergebnisse im Katalog Viewer und in der Karte angezeigt.

Zusätzlich wird an einer semantischen Abfragemöglichkeit gearbeitet, die es dem Endbenutzer erlaubt die semantischen Informationen zu spezifizieren, nach denen er suchen will. Diese Informationen werden mittels des Ähnlichkeitsmaßes mit den existierenden Landnutzungsontologien verglichen und die Ergebnisse im Katalog Viewer und in der Karte angezeigt.

5 Schlussfolgerungen

Das HarmonISA System ermöglicht es, bestehende Landnutzungsdaten automatisch zu integrieren. Dadurch kann eine harmonisierte, homogene Sicht auf die bestehenden heterogenen Landnutzungsdaten bereitgestellt werden. Der Aufwand, um weitere Landnutzungskataloge dem System hinzuzufügen, ist konstant, da pro Landnutzungskatalog eine Landnutzungsontologie erstellt werden muss und diese dann automatisch mit allen bestehenden Landnutzungsontologien harmonisiert wird.

Das HarmonISA System wird zur Zeit nur auf Landnutzungsdaten angewandt, das Ähnlichkeitsmaß ist aber so ausgelegt, dass es auf beliebige Daten angewandt werden könnte, für die die semantischen Inhalte innerhalb des Modells abgebildet werden können. Eine in dieser Richtung besonders interessante Anwendung wäre der Kataster. Eine Abbildung der Kataster der drei beteiligten Regionen würde die grenzüberschreitende Planung wesentlich vereinfachen.

6 Literatur

- Mandl P. und Hall M. M. (2006): *HarmonISA – ein Geoinformationssystem zum Studium grenzüberschreitender Landnutzung*. In: Angewandte Geoinformatik 2006. Beiträge zum 18. AGIT-Symposium Salzburg. Herbert Wichman Verlag, Heidelberg
- Gruber T. R. (1993): *A translation approach to portable ontology specifications*. In: Knowledge Acquisition, Vol. 5(No. 2)
- Patel-Schneider P. F., Hayes P. and Horrocks I. (2003): *Owl web ontology language semantics and abstract syntax*. W3C Working Draft 31 März 2003
- Horrocks I., Patel-Schneider P. F., van Harmelen F. (2003): *From SHIQ and RDF to OWL: The Making of a Web Ontology Language*. Journal of Web Semantics, Vol. 1(1)
- Baader F., Calvanese D., McGuinness D., Nardi D. und Patel-Schneider P. F., editors (2003): *The Description Logic Handbook*. Cambridge University Press
- Tversky A. (1977): *Features of Similarity*. In: Psychological Review, Vol. 84(4): p. 327 – 352
- Rodríguez A. und Egenhofer M. (2004): *Comparing geospatial entity classes: An asymmetric and context-dependent similarity measure*. In: International Journal of Geographical Information Science, Vol. 18(3):229 – 256
- Gärdenfors P. (2000): *Conceptual Spaces: The Geometry of Thought*. MIT Press, Cambridge MA
- Rada R. (1989): *Development and application of a metric on semantic nets*. In: IEEE Transactions on systems, man and cybernetics. Vol. 19(1): p. 17 – 30
- Rector A. L. (2003): *Modularisation of domain ontologies implemented in description logics and related formalisms including owl*. In: Proceedings of Knowledge Capture 2003, p. 121 – 128, ACM
- Levenshtein I. V. (1966): *Binary Codes capable of correcting deletions, insertions and reversals*. Cybernetics and Control Theory, Vol. 10(8): p. 707 – 710
- Miller G. A. (1995): *WordNet: A Lexical Database for English*. Communications of the ACM, Vol. 38(11): p. 39 – 41
- Maedche A. und Staab S. (2002): *Measuring Similarity between Ontologies*. In: Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. Ontologies and the Semantic Web. p. 251 – 263

- Castano S., Ferrara A., Montanelli S. (2005): *Ontology-based Interoperability Services for Semantic Collaboration in Open Networked Systems*. In: Proceedings of the 1st International Conference on Interoperability of Enterprise Software and Applications
- Lei Li und Horrocks I. (2004): *A Software Framework for Matchmaking Based on Semantic Web Technology*. In: International Journal of Electronic Commerce. Vol. 8(4)
- Haarslev V., Möller R. (2001): *RACER System Description*. In: Lecture Notes in Computer Science, Vol. 2083, p. 701
- Horrocks I. (1998): *Using an expressive description logic: FaCT or fiction?* In: Proceedings of the 6th International Conference on Principles of Knowledge Representation and Reasoning (KR'98), p. 636 – 647
- Bijan Parsia Evren Sirin (2004): *Pellet: an owl dl reasoner*. In: Proceedings of the International Workshop on Description Logics (DL2004)