

HAW HAMBURG
INFORMATIK MASTER

HAUPTSEMINAR

**Datenintegration als
automatisierter Prozess**

Integration mittels semantischer Analyse

Bearbeiter:

Tom Schöner (2182801)

Betreuung:

Prof. Dr. Kai von Luck

Prof. Dr. Tim Tiedemann

31. August 2018

Inhaltsverzeichnis

1	Abstract	3
2	Aufbau	3
3	Einführung	4
3.1	Structural Integration	4
3.2	Semantic Integration	5
3.3	Data Reconciliation	6
4	Fragestellung und Problemeingrenzung	6
5	Allgemeine Methodik	7
5.1	Ontologie Modell	8
5.2	Semantische Analyse	8
5.3	Beispiel	9
6	Vergleichbare Arbeiten	11
6.1	A methodology to discover semantic features from textual resources	11
6.2	Semantic Mediator Querying	11
6.3	Ontology-Based Integration of Information — A Survey of Existing Approaches	12
7	Auswertung	12
8	Ausblick	13

Abbildungsverzeichnis

1	Source Integration Hierarchie [3]	4
2	Mediator Schema mit globaler Ontologie	7
3	Generierte Ontologie aus Beispieldatensatz A und B	10

Tabellenverzeichnis

1	Beispieldatensatz A	9
2	Beispieldatensatz B	9
3	Ergebnis aus Beispieldatensatz A und B als Tabelle	10

1 Abstract

Mit dem zunächst abstrakten Begriff „Daten“ können viele Dinge beschrieben werden. Unter anderem sind eine kurze Reihe von aufeinander folgenden Einsen oder eine Ansammlung von Namen ein Beispiel für Daten. Selbst die Abwesenheit von Werten fällt unter diese Kategorie. Daten erhalten erst durch eine kontextbezogene Sichtweise Bedeutung und lassen sich folgend interpretieren. Analog zu anderen Fachgebieten, wie beispielsweise der menschlichen oder maschinellen Sprache sollte zwischen Form (Syntax, Notation, Struktur; vgl. Abschnitt 3.1) und Semantik (Inhalt; vgl. Abschnitt 3.2) unterschieden werden [3].

Die Unterscheidung zwischen Form und Inhalt ist insbesondere beim Prozess der Datenintegration notwendig. Zwei Datensätze können trotz unterschiedlichen Speicherformats wie XML und JSON dieselbe Aussage oder mit anderen Worten, den selben Inhalt, haben. Beim Inhalt ist unbedingt der Kontext zu beachten. Der Kontext kann sich mitunter aus der Zusammenführung mehrerer, häufig heterogener, Datenquellen ergeben.

Die Datenintegration beschäftigt sich mit diesem nicht trivialen Thema der Datenzusammenführung, daher die Frage: Wie kann eine in sich geschlossene Projektion auf die heterogene Datenlandschaft erzeugt werden? Für Unterkategorien gibt es häufig Lösungen oder Lösungsansätze, welche Vorarbeit und manuelle repetitive Datenaufbereitung voraussetzen [14]. Inwiefern sich der Prozess besonders unter semantischen Aspekten automatisieren lässt, soll im Folgenden näher erörtert werden.

2 Aufbau

In den folgenden Abschnitten möchte ich zunächst, für ein besseres Verständnis, die Teilgebiete der Datenintegration nach [3] vorstellen. In Abschnitt 3 werden die strukturelle Integration, die semantische Integration und Data Reconciliation als wichtige Teilgebiete erläutert. Anschließend wird in Abschnitt 4 die Leitfrage und Problemeingrenzung der Ausarbeitung in Hinsicht

auf mögliche (Teil-)Automatisierungen vorgestellt. Abschnitt 5 beschreibt die allgemeine Methodik unter Verwendung von Ontologie Modellen. In Abschnitt 6 wird ein Blick auf vergleichbare Arbeiten geworfen. Abschließend werden in 7 und 8 die aktuelle Arbeit referiert und offene Fragestellungen und Schritte aufgezeigt.

3 Einführung

Christoph Koch [3] arbeitet in seiner Dissertation über die Datenintegration gegen mehrere autonome Schemata drei Teilgebiete der Datenintegration heraus (siehe Abbildung 1). Die Zielstellung ist, eine Interoperabilität der Informationssysteme zu schaffen. Die Teilgebiete sind zudem hierarchisch noch unter der Source Integration einzuordnen, zu welcher auch die Schema Integration gehört.

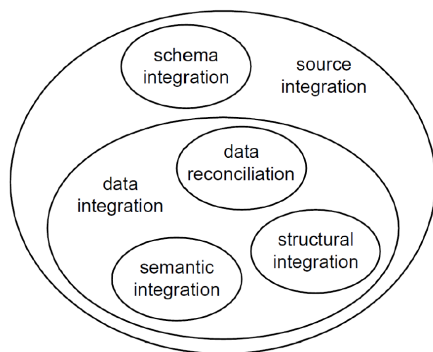


Abbildung 1: Source Integration Hierarchie [3]

3.1 Structural Integration

Bei der strukturellen Integration bedient man sich dem Prinzip der Kapselung [9]. Mittels *wrapping* wird versucht die Heterogenität verschiedener Modelle, Sprachen oder Protokolle aufzulösen. Dieser Ansatz ist besonders bei Datensätzen ohne Schreibzugriff oder der Kompatibilität halber bei Legacy-

Applikationen von Nutzen. *Global as View* und *Lokal as View* sind Beispiele der strukturellen Integration [11]. Sie ist Schema- und Applikationsabhängig.

3.2 Semantic Integration

Die semantische Integration ist für die Erkennung und Auflösung semantischer Unterschiede der Schemata zuständig [4]. Das Problem der unterschiedlichen Semantiken resultiert grundlegend auf der Tatsache, dass verschiedene oder unabhängige Quellen nicht demselben Schema folgen müssen. Beispielsweise können Temperaturdaten einer sich in Amerika befindenden Quelle Fahrenheit und aus Europa in Celsius dargestellt werden. Bei der semantischen Integration sollten solche Werte normiert werden; am Beispiel der Temperaturdaten könnten etwa alle Werte nach Kelvin umgewandelt werden.

Die semantische Integration setzt ein fundiertes Wissen über den Datensatz voraus. In *Comparative study of the systems of semantic integration of information: A survey (2015)* [1] werden diesbezüglich drei kritische Fragestellungen aufgestellt:

1. Was für Daten sind relevant?
2. Welche Systeme verfügen über welche relevanten Daten?
3. Wie kann durch Transformationen, Adaptionen und Paarungen für ein homogenes und verständliches Ergebnis gesorgt werden?

Weiterhin wird auf mögliche Ansätze der semantischen Integration verwiesen: *mapping based*, *query oriented* und *intermediary based*. Der dritte Ansatz *intermediary-based* ist besonders vielversprechend. Ähnlich der strukturellen Integration aus Abschnitt 3.1 wird hier eine Vermittler-, bzw. Mediatorschicht bereitgestellt, welche für die Interkommunikation und Interoperabilität mittels einer standardisierten Ontologie zuständig ist. Dieser Ansatz schafft es die Datenquellen weitgehend Schema- und Applikationsunabhängig zu modellieren. Guarino [5] schreibt über Ontologie passend:

„Ontologie [...] can be seen as the study of the organization and the nature of the world independently of the form of our knowledge about it“
(*Guarino, S. 628, 1995*)

3.3 Data Reconciliation

Betrachtet man sich einen Daten-Stream eines einfachen Sensors, welcher periodisch alle 150ms einen Empfänger anpingt, bewegt man sich schnell im Bereich der Data Reconciliation [6]. Hier werden Daten aufbereitet und bereinigt. Bei den Sensordaten können fehlerhafte Werte gefiltert oder auch mehrere Werte einer Zeitspanne aggregiert werden. Weitere Beispiele sind die Entfernung von Rechtschreibfehlern oder Duplikaten. Die Aufbereitung der Daten ist insbesondere als Vorarbeit der strukturellen und semantischen Integration hilfreich.

4 Fragestellung und Problemeingrenzung

Um aufzuzeigen, welche Prozesse der Integration sich (teil-)automatisieren lassen, sollte ein genauerer Blick auf die drei in Abschnitt 3 beschriebenen Unterkategorien geworfen werden. Überdies ist das Feld der Datenintegration weit gefächert und die Anzahl der Datenformate ist groß, entsprechend sollten Grenzen in Bezug auf die Fragestellung definiert werden. Die Informationsquellen werden in den folgenden Abschnitten wie folgt abstrahiert und vereinfacht.

- Die Daten liegen in einem einfachen Tabellenformat vor, ähnlich relativer Datenbanken.
- Die Spalten der Tabelle haben eine klare Bezeichnung.
- Die Daten liegen textuell vor.

Der Zugriff auf die Daten über existierende Datenbanksysteme wird hier nicht weiter verfolgt. Etwaige Datenbank Mappings werden in [7] erläutert.

Das Mediator Schema der semantischen Integration aus Abschnitt 3.2 ist durch seine im Hintergrund sich befindenden Ontologie ein guter Ansatzpunkt. Eine Ontologie zu definieren ist nicht trivial[1], sie wird aber für die erfolgreiche Kommunikation der Komponenten zwingend benötigt. Die Ontologie kann als Schicht über der Datenschicht dargestellt werden, wodurch die semantischen Unterschiede abstrahiert werden.

Das Kernproblem lässt sich folglich auf die automatische Ermittlung einer sinnvollen Ontologie unter Berücksichtigung der vereinfachten Informationsquellen (siehe Abbildung 2) und somit hauptsächlich auf Teilbereiche der semantischen Integration eingrenzen. Die Data Reconciliation kann als Vorarbeit der semantischen Integration zusätzlich hinzugezogen werden.

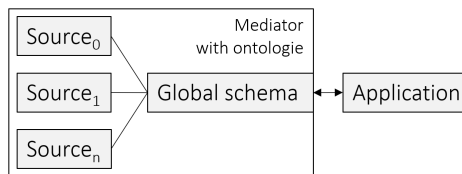


Abbildung 2: Mediator Schema mit globaler Ontologie

5 Allgemeine Methodik

Abschnitt 5.1 beschreibt brauchbare Ontologie Modelle, welche durch eine aus Abschnitt 5.2 beschriebene semantische Analyse definiert werden kann.

Das in Abschnitt 5.3 zunächst naiv gelöste Beispiel gibt daraufhin einen ersten Einblick in den Prozess und die Notwendigkeit einer Datenintegration. Die verwendeten Daten wurden für dieses Beispiel erstellt und stellen nicht zwangsweise reale Daten dar. Es ist mit nur zwei Tabellen bewusst klein gewählt.

5.1 Ontologie Modell

Die Ontologie¹ ist im gewissen Sinne das Benutzerhandbuch der Datenquellen und ein wichtiger Bestandteil des Systems. Sie sollte mit dem System inkrementell wachsen. Bei der Datenintegration muss man sich nicht auf eine globale Ontologie einschränken [2, 14], sondern kann durch einen hierarchischen Ansatz an Flexibilität gewinnen. Für einfache Fälle reicht jedoch der globale Ansatz aus. Auch beim hierarchischen Modell ist es wichtig, dass die oberste Schicht ein globales Verständnis des Systems besitzt.

Die *Web Ontology Language* (OWL2)² ist eine Möglichkeit Ontologien mittels einer Sprache zu beschreiben. OWL gehört zur \mathcal{SH} Familie der Beschreibungslogik. Eine genauere Erläuterung der Logik basierten Integration ist in [3] Seite 62 ff. zu finden.

5.2 Semantische Analyse

Um eine funktionierende Ontologie generieren zu können, bedarf es einer semantischen Analyse der Daten. Der Zusammenhang zwischen Daten gleicher oder divergenter Quellen ist für die Verarbeitung von inhaltlichen Anfragen/Queries bedeutsam. Erst durch einen semantischen Bezug kann ein globales Bild erstellt werden [14]. Das Resultat der semantischen Analyse kann folgend eine Ontologie sein.

In [14] werden drei Hauptursachen der Heterogenität genannt, namentlich *Confounding conflicts*, *Scaling conflicts* und *Naming conflicts*. Die Erkennung solcher semantischen Merkmale bedarf in der Regel eines mehrstufigen Verfahrens.

Einige der semantischen Analyse Verfahren sind im Bereich Maschine Learning mit rekurrentem neuronalen Netzes anzutreffen [10]. Das Verfahren aus [10] gibt beispielsweise Auskunft über die Ähnlichkeiten zweier Sätze und wäre demnach für die Erkennung von *Confounding conflicts* oder *Naming conflicts* geeignet. Weitere Teilbereich der semantischen Integration könnten

¹<http://tomgruber.org/writing/ontology-definition-2007.htm>

²<https://www.w3.org/OWL/>

ebenfalls von diesem oder ähnlichen Verfahren profitieren. Der in [2] (siehe auch Abschnitt 6.2) verwendete Algorithmus zeigt dies. Welche Teilbereiche der Analyse sich gut durch maschinelles Lernen abbilden lassen können, ist noch eine offene zu untersuchende Frage.

5.3 Beispiel

Es seien zwei Datenquellen A und B gegeben (siehe Tabelle 1 und 2). A beschreibt simple Orts- (*location*) bezogene Wetterdaten, wobei B Länder (*country*) ihrer Hauptstadt (*capital*) zuordnet. Die Aufgabe der zu generierenden Ontologie ist es, Zugriffe auf beide Quellen zu abstrahieren und kombinieren zu können. Dazu zählen, in diesem Fall nicht formal gestellt, Fragen wie: „Wie warm ist es etwa in Rom?“ Diese Fragen können nur durch eine Verbindung der Tabellen mit Hilfe einer semantischen Analyse gelöst werden.

location	temperature	humidty
Germany	28.992	43.23
France	20.119	70.17
Italy	32.216	33.32

Tabelle 1: Beispieldatensatz A

country	capital
Italy	Rome
Germany	Berlin
Sweden	Stockholm
Spain	Madrid

Tabelle 2: Beispieldatensatz B

Selbst bei diesem recht simplen Beispiel werden die Schwierigkeiten deutlich. Zunächst haben die Spalten keine übereinstimmenden Identifikatoren,

wodurch eine automatische Zusammenführung erschwert wird. Des Weiteren können die Daten selbst nach erfolgreicher Zusammenführung Lücken aufweisen. Beispielsweise ist Frankreich nur in A , aber nicht in B enthalten.

Um eine gemeinsame Sicht auf beide Tabellen zu ermöglichen, sollte auch der Inhalt der Spalten gegeneinander abgeglichen werden. Bei Texten sowie bei diesem Beispiel eignet sich häufig die Levenshtein Distanz, um kleinere Ungenauigkeiten der Wörter auszugleichen. Durch diese Spaltenanalyse lässt sich erkennen, dass die Spalten *location* und *country* äquivalent sind.³

Nach Abschluss der (semantischen) Analyse kann eine passende Ontologie mit Metadaten generiert werden (siehe Abbildung 3). Die während dieses Prozesses gesammelten Metadaten werden für die Auswertung der Ontologie verwendet. Zu den Metadaten gehört vor allem die Zuordnung des Eintrags in der Ontologie zur Tabelle. Ein weiteres Beispiel ist der Datentyp. Die Ontologie lässt schließlich auf Tabelle 3 schließen.

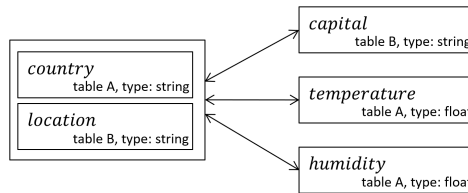


Abbildung 3: Generierte Ontologie aus Beispieldatensatz A und B

country	capital	temperature	humidty
Italy	Rome	32.216	33.32
Germany	Berlin	28.992	43.23
Sweden	Stockholm	-	-
Spain	Madrid	-	-
France	-	20.119	70.17

Tabelle 3: Ergebnis aus Beispieldatensatz A und B als Tabelle

³Die Paarung einzelner Zeilen kann durch fehlende Daten scheitern und ist folglich nur ein hinreichendes Kriterium. Ein genaueres Ergebnis bedarf einer semantischen Analyse.

6 Vergleichbare Arbeiten

6.1 A methodology to discover semantic features from textual resources

In „A methodology to discover semantic features from textual resources“ von 2011 [12] wird von den Autoren **Vicient et al.** das Extrahieren semantischer Merkmale aus textbasierten Quellen erläutert. Hauptsächlich wird Wikipedia aufgrund seiner weiterführenden Links in den zu analysierenden Texten verwendet. Die extrahierten Merkmale werden zusätzlich mit Hilfe einer Ontologie annotiert. Durch den Automatismus des Verfahrens lassen sich Parallelen zu der Zielsetzung dieser Ausarbeitung erkennen. Der genauere Algorithmus ist in der Arbeit als Pseudocode-Implementierung hinterlegt.

6.2 Semantic Mediator Querying

Eine weitere Analyse des Mediator Schemas ist im Jahr 2014 erschienenen „Semantic Mediator Querying“ von **Bouchou and Niang** [2] zu finden. Es werden unterliegende *agreement* und globale Ontologien in einer verteilten Architektur verwendet. Die untere Schicht ist der *Source Layer* und die obere globale Schicht der *Mediator Layer*. Alle Anfragen und Queries werden durch die Mediator Schicht beantwortet. Die globale Ontologie des Mediators muss das gesamte unterliegende System verstehen können. Der Aufbau der Ontologie geschieht inkrementell, die Registrierung neuer Quellen ist demnach möglich. Externe Anfragen werden über den Query Prozessor verarbeitet. Die Anfrage muss sich dabei an der globalen Ontologie orientieren. Intern wird sie dann entsprechend an die Ontologien der Quellen angepasst.

Die hierarchische Trennung der Quellen dient zum einen der Abstrahierung der Datenbanksysteme und der globalen Ontologie und zum anderen fördert sie die Skalierbarkeit, bzw. das inkrementelle Wachstum neuer und bestehender Quellen. Figur 2 aus [2] beschreibt diese Architektur sehr gut. Die semantische Analyse beruht auf Beschreibungslogik und bedient sich an vielen Stellen an einer modifizierten Version des effizienten Minicon Algorithmus [8], um

eingehende Queries mit mehreren Views weiter bearbeiten zu können.

6.3 Ontology-Based Integration of Information — A Survey of Existing Approaches

Wache et al. rezensieren in „Ontology-Based Integration of Information — A Survey of Existing Approaches“ von 2001 [14] etwa 25 Integrationsansätze und kategorisieren diese. So werden mehrere Kategorien wie einfache globale Ontologien, multiple lokale Ontologien oder hybride Ansätze aufgezeigt. Interessant ist hier die Aufstellung der Vor- und Nachteile genannter Verfahren. Grundlegend liegt hier der Implementationsaufwand dem Nutzen und der Erweiterbarkeit gegenüber. Die Implementierung einer globalen Ontologien wird, im Gegensatz zum hybriden Ansatz, als *straight forward* eingeschätzt. Änderungen der Datenquellen führen wiederum zu einer Anpassung der globalen Ontologie.

Der hybride Ansatz wurde entwickelt, um den Mängeln der beiden vorherigen Ansätze entgegenzuwirken. Einzig die Wiederverwendung von lokalen Ontologien wird durch das an die globale Ontologie angepasste Vokabular als schwierig eingeschätzt. Beim beschriebenen Ansatz MECOTA [13], einem hybriden Verfahren, werden lokale Ontologien mit Hilfe von Annotationen mit der globalen Ontologie/ dem globalen Vokabular verbunden. Viele der untersuchten Systeme (unter Anderem OBSERVER oder SIMS) beruhten für die Beschreibung der Ontologien auf beschreibungslogischen Sprachen.

Obwohl die Zusammenstellung 2001 und somit 17 Jahre vor dieser Ausarbeitung entstanden ist, thematisiert sie viele heute noch relevante Themen. Die enthaltenen Referenzen zu weiteren Ressourcen und konkreten Implementationen bieten zusätzlich einen guten Überblick über das Thema der semantischen Datenintegration.

7 Auswertung

Die Wahl eines sinnvollen Verfahrens für die semantische Analyse ist auf Grund der hohen Heterogenität der Daten ein elementares Ziel. Textlastige

Datensätze können damit sinnvoll interpretiert und semantisch untereinander in der Ontologie verknüpft werden. Bei computergenerierten Datensätzen, deren Inhalt nicht textuell ist, sondern aus Zahlen besteht und womöglich keine beschreibenden Metadaten vorhanden sind, ist die Analyse kaum aussagekräftig. Beispielsweise können Sensordaten, die automatisch in einer Tabelle mit den Bezeichnern t_1 bis t_{22} hinterlegt werden, ohne Kontext nur schwer Interpretiert werden.

Die Darstellung der Ontologie ist im Vergleich trivial. Etablierte Sprachen wie OWL bieten hier bereits ein solides Fundament, auf dem aufgebaut werden kann.

8 Ausblick

Zum Zeitpunkt der Fertigstellung dieser Ausarbeitung befindet sich das Verfahren und die Analyse für eine automatisierte Datenintegration in einem sehr frühen Stadium. Eine prototypische Implementation liegt noch nicht vor. Diese Ausarbeitung dient mehr der Erläuterung der Problemstellung, Sammlung von relevanten Ressourcen und somit einen Einstieg in das Thema der Datenintegration sowie möglichen Automatisierungsansätzen im Bereich der semantischen Integration. Im nächsten Schritt sollte entsprechend eine konkrete Vorgehensweise für die Generierung von Ontologien unter spezifizierten Rahmenbedingungen definiert werden. Etablierte Verfahren hierfür wurden in Abschnitt 6.3 erläutert. Ein Abgleich mit heutigen Verfahren muss natürlich weiterhin vorgenommen werden.

Der leicht provokante Titel „Datenintegration als automatisierter Prozess“ bleibt hierbei als Leitfaden bestehen. Vergleichbare Arbeiten aus Abschnitt 6 zeigen die wiederholt auftretenden Probleme eines generellen Ansatzes, aber zugleich auch die Relevanz von Datenintegration in der heutigen Zeit.

Literatur

- [1] O. Banouar and S. Raghay. Comparative study of the systems of semantic integration of information: A survey. In *2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA)*, pages 1–8, Nov 2015. doi: 10.1109/AICCSA.2015.7507235.
- [2] Béatrice Bouchou and Cheikh Niang. Semantic mediator querying. In *Proceedings of the 18th International Database Engineering & Applications Symposium, IDEAS '14*, pages 29–38, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2627-8. doi: 10.1145/2628194.2628218. URL <http://doi.acm.org/10.1145/2628194.2628218>.
- [3] Christoph Koch. *Data Integration against Multiple Evolving Autonomous Schemata*. PhD thesis, Technische Universität Wien; Fakultät für Technische Naturwissenschaften und Informatik, 2001.
- [4] Hector Garcia-Molina, Yannis Papakonstantinou, Dallan Quass, Anand Rajaraman, Yehoshua Sagiv, Jeffrey D. Ullman, Vasilis Vassalos, and Jennifer Widom. The tsimmis approach to mediation: Data models and languages. *Journal of Intelligent Information Systems*, pages 117–132, 1997.
- [5] Nicola Guarino. Formal ontology, conceptual analysis and knowledge representation. *Int. J. Hum.-Comput. Stud.*, 43(5-6):625–640, December 1995. ISSN 1071-5819. doi: 10.1006/ijhc.1995.1066. URL <https://doi.org/10.1006/ijhc.1995.1066>.
- [6] Matthias Jarke, Maurizio Lenzerini, Yannis Vassiliou, and Panos Vassiliadis. *Fundamentals of Data Warehouses*. Springer Verlag, 2000.
- [7] Kamran Munir and M. Sheraz Anjum. The use of ontologies for effective knowledge modelling and information retrieval. *Applied Computing and Informatics*, 14(2):116 – 126, 2018. ISSN 2210-8327. doi: <https://doi.org/10.1016/j.aci.2017.07.003>. URL <http://www.sciencedirect.com/science/article/pii/S2210832717300649>.

- [8] Rachel Pottinger and Alon Halevy. Minicon: A scalable algorithm for answering queries using views. *The VLDB Journal*, 10(2):182–198, Sep 2001. ISSN 0949-877X. doi: 10.1007/s007780100048. URL <https://doi.org/10.1007/s007780100048>.
- [9] Mary Tork Roth and Peter Schwarz. Don't scrap it, wrap it! a wrapper architecture for legacy data sources. In *Proceedings of the 1997 International Conference on Very Large Data Bases (VLDB'97)*, 1997.
- [10] Adrian L. Sanborn and Jacek Skryzalin. Deep learning for semantic similarity. 2015. URL <https://www.semanticscholar.org/paper/Deep-Learning-for-Semantic-Similarity-Sanborn/e1138936eb4bd0c81f73f3e25aa76e6ee285d3cc>.
- [11] Jeffrey Ullmann. Information integration using logical views. *Proceedings of the 6th International Conference on Database Theory*, pages 19–40, 1997.
- [12] C. Vicient, D. Sánchez, and A. Moreno. A methodology to discover semantic features from textual resources. In *2011 Sixth International Workshop on Semantic Media Adaptation and Personalization*, pages 39–44, Dec 2011. doi: 10.1109/SMAP.2011.13.
- [13] H. Wache, T. Scholz, H. Stieghahn, and B. König-Ries. An integration method for the specification of rule-oriented mediators. In *Proceedings 1999 International Symposium on Database Applications in Non-Traditional Environments (DANTE'99) (Cat. No.PR00496)*, pages 109–112, Nov 1999. doi: 10.1109/DANTE.1999.844947.
- [14] H. Wache, T. Voegelé, T. Visser, H. Stuckenschmidt, H. Schuster, G. Neumann, and S. Huebner. Ontology-based integration of information - a survey of existing approaches. In H. Stuckenschmidt, editor, *IJCAI-01 Workshop: Ontologies and Information*, pages 108–117, 2001. URL <http://www.cs.vu.nl/~heiner/public/ois-2001.pdf>.