

Starke Kopplung: Interaktion als Schlüssel für das Semantic Web

Philipp Heim, Thomas Schlegel, Thomas Ertl

Institut für Visualisierung und Interaktive Systeme (VIS), Universität Stuttgart

Abstract

In diesem Beitrag stellen wir den Ansatz der starken Kopplung vor. Sowohl bei der Erstellung von, als auch beim Zugriff auf Informationen im Semantic Web informieren sich dabei Mensch und Computer ständig gegenseitig darüber, in welcher Weise übermittelte Informationen interpretiert wurden. Dadurch können Fehler bei der Übersetzung zwischen Informationen mit impliziter und expliziter Bedeutung schneller erkannt, korrigiert und dadurch möglicherweise daraus resultierende Probleme vermieden werden. Basierend darauf stellen wir Kriterien auf, anhand derer bestehende und neue interaktive Ansätze im Semantic Web evaluiert und verbessert werden können.

1 Einleitung

Vor 10 Jahren skizzierte Tim Berners-Lee die Vision eines Webs, in dem es Computern möglich sein wird, alle darin enthaltenen Informationen zu verstehen und daher Aufgaben zu übernehmen, die bisher nur von Menschen ausgeführt werden konnten (Berners-Lee & Fischetti 1999). Hierfür muss sowohl die Bedeutung der einzelnen Informationen im Web, als auch deren Beziehungen untereinander mit Hilfe sogenannter Ontologien eindeutig beschrieben werden. Intelligenten Agenten wäre es damit möglich, Informationen automatisch auf Grund ihrer Bedeutung ausfindig zu machen, auszutauschen und durch deren Verknüpfung neues Wissen zu generieren.

Mit seiner Vision legte Tim Berners-Lee den Grundstein für das Semantic Web in seiner derzeitigen Form, in dem die Bedeutung von Daten durch semantische Strukturen repräsentiert und in formalen Sprachen wie *RDF*⁵⁴ und *OWL*⁵⁵ abgespeichert ist. Die Strukturen enthalten Aussagen über reale Objekte wie „Berlin“ oder „Bundesrepublik Deutschland“, jedes mit einer *URI* (Uniform Resource Identifier), einer zugewiesenen ontologischen Klasse, wie „Stadt“ oder „Land“, und einer beliebigen Anzahl an Eigenschaften die Verbindungen zu anderen Objekten definieren, wie „ist Hauptstadt von“.

⁵⁴ RDF (Resource Description Framework): <http://www.w3.org/RDF/>.

⁵⁵ OWL (Web Ontology Language): <http://www.w3.org/TR/owl-features/>.

Der bis heute fehlende Durchbruch des Semantic Web, als wichtige Ergänzung und Weiterentwicklung des World Wide Web, resultiert in erster Linie aus seinem bis jetzt noch nicht offensichtlichen und zu großen Teilen auch noch nicht vorhandenen breiten Nutzen für den durchschnittlichen Nutzer. Dies liegt zum einen daran, dass nur ein Bruchteil der im bestehenden Web verfügbaren Information im Semantic Web repräsentiert ist, und zum anderen daran, dass auf die im Semantic Web bereits vorhandenen Informationen oftmals nicht geeignet durch die Menschen zugegriffen werden kann. Beide Probleme resultieren letztlich aus der bereits durch Aristoteles erkannten Tatsache, dass reale Objekte, die sie repräsentierenden Symbole und deren Bedeutungen nicht eindeutig zusammengehören (vgl. semiotisches Dreieck) und somit z.B. natürlich-sprachliche Texte, Bilder oder Gesten, nicht eindeutig interpretiert werden können. So können Informationen aus dem bestehenden Web oftmals überhaupt nicht oder nur in geringer Qualität automatisch in semantische Strukturen mit expliziter Bedeutung übersetzt werden, und eine manuelle Überführung ist auf Grund der großen Menge an Informationen im Web zu aufwendig. Auf der anderen Seite führen Anfragen von Nutzern mit impliziter Bedeutung, z.B. natürlich-sprachliche Anfragen, selten zu den gewünschten Ergebnissen, da auf Seiten des Computers nicht klar wird, wonach gesucht werden sollte.

Um im Semantic Web die beschriebenen Probleme bei der automatischen Interpretation von Symbolen ohne eindeutig zugewiesene Bedeutung zu vermeiden, schlagen wir in diesem Beitrag das Konzept der *starken Kopplung* zwischen Mensch und Computer vor. Die Idee dabei ist, dass sich Mensch und Computer iterativ und in kurzen Abständen gegenseitig mitteilen, auf welche Weise sie die Äußerungen des anderen interpretiert haben. Dadurch findet eine ständige Kontrolle der Interpretation statt, und Missverständnisse oder Fehler können frühzeitig erkannt werden. Dies ermöglicht eine rechtzeitige Korrektur und somit die iterative Annäherung an das erwünschte Ergebnis. Darüber hinaus können in jedem Dialogzyklus neue Erkenntnisse gewonnen werden und eine entsprechende Änderung der Zielvorstellung noch während des Dialogs erfolgen. Die starke Kopplung kombiniert somit automatische und interaktive Verfahren in einer Weise, die die jeweiligen Schwächen weitestmöglich auszugleichen versucht, um ein optimales Zusammenspiel zu erreichen.

Insbesondere aus dem Umfeld der *Semantic Web User Interaction Community (SWUI)* existieren bereits diverse Beiträge zum Thema Mensch-Computer-Interaktion im Semantic Web. Wie in (Heath 2005) und (Battle 2006) wird dabei jedoch meist aus der Nutzerperspektive analysiert und argumentiert. Nutzergruppen werden charakterisiert, die jeweiligen Tätigkeitsfelder identifiziert und Ansätze zur Unterstützung dieser Tätigkeiten entwickelt. Die Fragestellung lautet stets „Welchen Nutzen kann das Semantic Web für den Menschen bringen“ und nicht „In welcher Weise kann der Mensch mithelfen, um die Probleme des Semantic Web zu lösen“. Da die Probleme des Semantic Web aber dessen Nützlichkeit massiv einschränken, führt kein Weg an der Frage nach geeigneten Ansätzen für die Lösung dieser Probleme vorbei.

Im Folgenden motivieren wir, ausgehend von einer Analyse der bestehenden Probleme im Semantic Web, unseren Ansatz der starken Kopplung und erklären anhand eines iterativen Prozessmodells die generelle Lösungsstrategie. Daraufhin beschreiben wir die verschiedenen Phasen, die während des iterativen Prozesses, sowohl bei der Erstellung, als auch beim Zu-

griff zwischen Mensch und Computer durchlaufen werden und leiten entsprechende Kriterien ab, die erfüllt sein sollten, um diese Phasen optimal zu unterstützen. Diese können sowohl zur Bewertung bestehender Ansätze, als auch zur Verbesserung neuer Ansätze eingesetzt werden.

2 Problemanalyse

Die grundsätzliche Idee des Semantic Web ist es, die Bedeutung von Daten explizit in Form von semantischen Strukturen vorzuhalten und somit auch für den Computer eine eindeutige Interpretation der Daten zu ermöglichen. Damit soll es dem Nutzer ermöglicht werden, mit Hilfe automatischer Verfahren schneller und gezielter auf Informationen zugreifen zu können. Dies erfordert allerdings, dass möglichst alle vorhandenen Daten, bzw. deren Bedeutungen, durch semantische Strukturen repräsentiert sind. Im Folgenden beschreiben wir die bei der Nutzung, d.h. bei der Erstellung von und beim Zugriff auf semantische Strukturen, auftretenden Probleme.

2.1 Probleme bei der Erstellung semantischer Strukturen

Bei der automatischen Übersetzung von Daten mit implizierter Bedeutung, wie beispielsweise Textdokumente, Bilder oder Videos, in semantische Strukturen mit expliziter Bedeutung lassen sich die folgenden Probleme feststellen:

- *Erstellungsproblem 1 (EPI): Die benötigten ontologischen Klassen und deren Verbindungen existieren noch nicht.* Um die Bedeutung von Daten aus einer bestimmten Domäne, z.B. Finanzdaten, in Objekte und Eigenschaften des Semantic Web übersetzen zu können, müssen die ontologischen Klassen und Verbindungen hierfür bereits im Vorfeld modelliert worden sein (z.B. Bank, Konto, Geld, Kontoinhaber etc.). Leider existieren ausgefeilte und standardisierte Klassenhierarchien (Ontologien) im Moment nur zu sehr populären Domänen, wie z.B. zu Dokumenten (*Dublin Core*⁵⁶), sozialen Netzwerken (*FOAF*⁵⁷) oder Geo-Daten (*GeoNames*⁵⁸). Eine automatische Generierung von neuen Klassen und Verbindungen aus textuellen Informationen, wenn noch keine entsprechenden ontologischen Strukturen vorhanden sind, ist schwierig.
- *Erstellungsproblem 2 (EP2): Unstrukturierte Daten, wie natürlich-sprachliche Texte, können meist nur in unzureichender Qualität automatisch übersetzt werden.* Die Ergebnisse automatischer Verfahren bei der Übersetzung unstrukturierter Daten sind in vielen Fällen unvollständig oder fehlerhaft, da die Semantik natürlicher Sprache zumeist nicht in ausreichendem Maße erkannt wird. Eine rein manuelle Extraktion ist auf Grund der großen, sich ständig ändernden und wachsenden Menge an unstrukturierten Daten im Web jedoch zu aufwendig.

⁵⁶ Dublin Core: <http://dublincore.org/>.

⁵⁷ FOAF (Friend Of A Friend): <http://www.foaf-project.org/>.

⁵⁸ GeoNames: <http://www.geonames.org/>.

- *Erstellungsproblem 3 (EP3): Strukturierte Daten liegen nur in beschränktem Umfang vor und führen auch nicht immer zu fehlerfreien semantischen Strukturen.* Je nach dem Grad der Strukturiertheit können die Daten in mehr oder weniger guter Qualität automatisch in semantische Strukturen überführt werden. Beispiele hierfür sind *Triplify*⁵⁹, das die Strukturen in Datenbanken ausnutzt, und *DBpedia*⁶⁰, das die Tabellenstruktur der Infokästen in Wikipedia-Artikeln verwendet, um automatisch entsprechende semantische Strukturen zu generieren. Leider liegen die meisten Daten unstrukturiert vor und auch bei strukturierten Daten führt die automatische Erstellung immer wieder zu fehlerhaften semantischen Strukturen.
- *Erstellungsproblem 4 (EP4): Bestehende Strukturen werden nicht verwendet, sondern immer wieder neu erstellt.* Oftmals werden die gleichen Objekte, Klassen oder Verbindungen mehrmals an unterschiedlichen Stellen erstellt, was zu redundanten, widersprüchlichen oder unvollständigen semantischen Strukturen führen kann.

2.2 Probleme beim Zugriff auf semantische Strukturen

Beim Zugriff auf semantische Strukturen ist die Vision, dass der Mensch sein Informationsbedürfnis in einer möglichst menschengerechten Form kundtut, z.B. in Form einer natürlich-sprachlichen Anfrage wie „Wie groß war Essen vor dem Krieg?“. Daraufhin soll der Computer die Anfrage geeignet interpretieren, die gewünschten Informationen im Semantic Web automatisch auffinden und als Ergebnis zurückliefern. Die bei diesem Prozess auftretenden Probleme sind in Abb. 1 dargestellt.

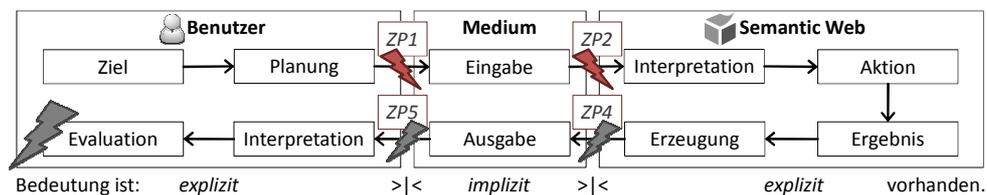


Abbildung 1: (Linearer Prozess) Um ein bestimmtes Ziel zu erreichen, tätigt der Benutzer eine Eingabe, die automatisch mit Hilfe des Semantic Web interpretiert und ausgewertet wird. Daraufhin wird eine entsprechende Ausgabe erzeugt, welche wiederum vom Benutzer interpretiert und im Abgleich mit dem Ziel evaluiert wird.

Das allgemeine Modell der Mensch-Rechner-Interaktion (Fährnich & Ziegler 1984) wird dabei für die Darstellung der Probleme beim Zugriff auf semantische Strukturen entsprechend adaptiert. Auf Grund der mehrfachen Übergänge zwischen Daten mit expliziter Bedeutung und Daten mit impliziter Bedeutung (vgl. Abb. 1) lassen sich die folgenden Probleme beim Zugriff auf semantische Strukturen identifizieren:

⁵⁹ Triplify: <http://triplify.org>.

⁶⁰ DBpedia: <http://dbpedia.org/>.

- *Zugriffsproblem 1 (ZP1): Der Mensch formuliert die Anfrage unpräzise oder falsch.* Schon bei der Übersetzung des Informationsbedürfnisses in eine konkrete Anfrage können Fehler auftreten (Abb.1, Übergang: Planung→Eingabe).
- *Zugriffsproblem 2 (ZP2): Die automatische Interpretation der Anfrage schlägt fehl.* Die implizite Bedeutung der Anfrage muss erst automatisch interpretiert werden, bevor entsprechende Aktionen durchgeführt werden können. Natürliche Sprache, Bilder und Gesten sind jedoch oft nicht eindeutig automatisch einer Bedeutung zuordenbar und daher schwierig zu verarbeiten (Abb.1, Übergang: Eingabe→Interpretation).
- *Zugriffsproblem 3 (ZP3): Die angefragten Daten sind nicht, nur teilweise oder fehlerhaft im Semantic Web vorhanden.* Auf Grund der beschriebenen Probleme bei der Erstellung (siehe Abschnitt 2.1) sind viele Inhalte nicht, oder nur unzureichend im Semantic Web repräsentiert. Somit reicht die Qualität auf Grund redundanter, fehlender, falscher oder widersprüchlicher Klassen, Objekte oder Verbindungen nicht aus, um immer verlässliche Ergebnisse zu ermöglichen. Eine automatische Kompensation durch Auswahl geeigneter Ersatzdaten ist schwierig.
- *Zugriffsproblem 4 (ZP4): Die gefundenen semantischen Strukturen werden falsch oder unvollständig dargestellt.* Die gefundenen Daten mit expliziter Bedeutung müssen in geeigneter Weise ausgegeben und dargestellt werden. Der Übersetzung in Daten mit implizierter Bedeutung liegen bestimmte Annahmen und Modellen zu Grunde, die von Menschen erdacht wurden. Je nach Alter, Kultur und Erfahrung können die Annahmen und Modelle jedoch stark variieren (Abb.1, Übergang: Erzeugung→Ausgabe).
- *Zugriffsproblem 5 (ZP5): Die dargestellten Inhalte werden falsch vom Menschen interpretiert.* Genau wie bei ZP4 liegen der Interpretation bestimmte Annahmen und Modelle zu Grunde die von Mensch zu Mensch stark unterschiedlich sein können. Damit kann ein und dieselbe Darstellung auf sehr unterschiedliche Weise interpretiert werden (Abb.1, Übergang: Ausgabe→Interpretation).

Jedes der angesprochenen Probleme kann dazu führen, dass das Ergebnis nicht mit den Zielvorstellungen, dem Informationsbedürfnis des Menschen, übereinstimmt. Auf Grund der Linearität des Prozesses (vgl. Abb.1) können sich die Probleme unkontrolliert akkumulieren und machen somit ein zufriedenstellendes Ergebnis immer unwahrscheinlicher. Auch kann nicht nachvollzogen werden, welche konkreten Probleme zum falschen Ergebnis beigetragen haben, um entsprechend gegensteuern zu können.

3 Lösungsansatz „starke Kopplung“

Alle analysierten Probleme, sowohl bei der Erstellung von, als auch beim Zugriff auf semantische Strukturen, resultieren direkt oder indirekt aus einer fehlerhaften Überführung von Daten mit impliziter Bedeutung in Daten mit expliziter Bedeutung oder umgekehrt. Dies ist vergleichbar mit den Verständnisproblemen zwischen zwei Menschen. Hier kann es auch vorkommen, dass der Sender einer Nachricht die Bedeutung so unglücklich in Symbole, z.B.

Wörter, übersetzt, dass die ursprünglich intendierte Bedeutung verloren geht, oder die Aussage sogar eine andere, naheliegendere Bedeutung bekommt. Auf der anderen Seite können aber auch vom Sender klar und eigentlich eindeutig formulierte Sätze beim Empfänger falsch verstanden werden.

Da sich die Menschen dieser Problematik bewusst sind, existieren entsprechende Lösungsstrategien, um Verständnisprobleme so weit wie möglich zu vermeiden. Die entscheidende Strategie hierbei ist die des Dialogs. In einem Dialog werden Inhalte in mehreren Schritten in einem iterativen Prozess ausgetauscht (im Gegensatz zum linearen Prozess in Abb. 1). Für jeden Schritt bekommt der übermittelnde Dialogpartner eine Rückmeldung darüber, ob und auf welche Weise die übermittelten Daten beim Gegenüber interpretiert wurden. Je kürzer die Zyklen sind, desto schneller können Missverständnisse erkannt und aufgeklärt werden. Wir sprechen hierbei von einer „starken Kopplung“.

In diesem Beitrag schlagen wir das Prinzip der starken Kopplung als allgemeine Lösungsstrategie für die beschriebenen Probleme im Semantic Web vor. Aufbauend auf (Fährnich & Ziegler 1984) wird in Abb. 2 der dazu passende Prozess dargestellt. Dieser beschreibt, wie das Prinzip der starken Kopplung, sowohl bei der Erstellung von, als auch beim Zugriff auf semantische Strukturen, eingesetzt werden kann, um Probleme zu vermeiden. In kurzen Iterationen werden Daten mit impliziter Bedeutung gesendet und deren Interpretation beim Empfänger anhand der Rückmeldung kontrolliert (Dialogebene). Die Möglichkeit der Iteration auf konzeptueller Ebene erlaubt einen schrittweisen Zugriff, bzw. Veränderung der Daten und die Anpassung der Zielvorstellung noch während des Prozesses.

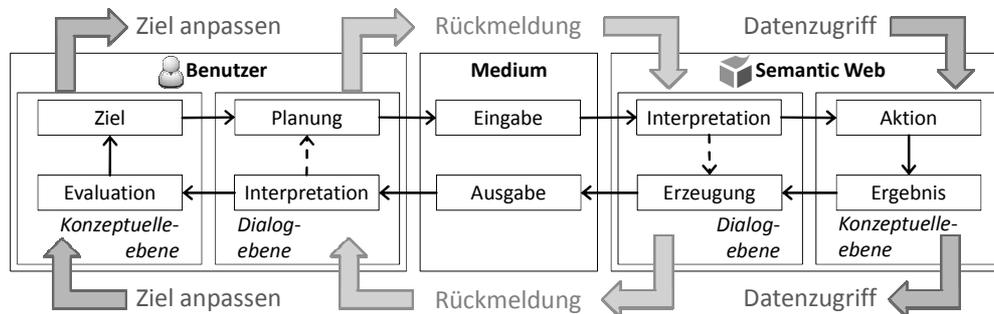


Abbildung 2: (Iterativer Prozess) Durch kurze Zyklen werden Probleme bei der Interpretation oder Erzeugung von Daten mit implizierter Bedeutung schnell erkannt und können somit noch während des Prozesses behoben werden.

4 Iterative Erstellung semantischer Strukturen

Allgemein lässt sich die Erstellung von Information in die folgenden Phasen unterteilen (Turk 2006): Gewinnung, Überarbeitung, Speicherung und Verbreitung. Wir verwenden diese Einteilung, um die Tätigkeiten bei der iterativen Erstellung semantischer Strukturen geeignet zu gliedern. Im Folgenden beschreiben wir die Tätigkeiten in den einzelnen Phasen

und leiten Kriterien (*Erstellungskriterien, EK*) für deren optimale Unterstützung ab. Indem Applikationen diese Kriterien erfüllen, soll es Benutzern ermöglicht werden, die in Abschnitt 2.1 analysierten Probleme zu vermeiden und somit die Erstellung semantischer Strukturen effektiv und effizient gestalten zu können.

In der Phase der *Gewinnung* werden aus Eingabedaten, wie z.B. Textdokumenten, semantische Strukturen extrahiert. Hierfür müssen die Eingabedaten analysiert, gefundene Informationen aggregiert, entsprechende Klassen gebildet und die Informationen zu Objekten und Verbindungen zwischen diesen überführt werden. *Kriterien sind*: Interaktives Data Mining, wie z.B. Clusterverfahren, statistische oder linguistische Analysen (*EK1.1*), eine geeignete Präsentation der Ergebnisse durch Übersichts- und Detailvisualisierungen (*EK1.2*), eine Unterstützung bei der Erstellung passender Klassenstrukturen durch kollaborative Ansätze wie im *Social Semantic Web* (Blumauer & Pellegrini 2009) (*EK1.3*) und eine schnelle und einfache Extraktion und Klassifikation von relevanten Informationen zusammen mit einer Rückverfolgbarkeit bis zu den Fundorten (*EK1.4*). Durch die Zusammenarbeit von Mensch und Computer, aber auch von Menschen untereinander, lassen sich semantische Strukturen schneller und in höherer Qualität erstellen (vgl. *EP1, EP2, EP3*).

Die nächste Phase, die *Überarbeitung*, beinhaltet die Aufgaben der Qualitätskontrolle und das Einhalten von Standards. Gewinnung und Überarbeitung können eine innere Schleife bilden. Für die gewonnen semantischen Strukturen muss überprüft werden, ob entsprechende Klassen, Objekte oder Verbindungen im Semantic Web schon existieren. *Kriterium ist*: Geeignete Möglichkeiten zum Abgleich mit bereits im Semantic Web existierenden semantischen Strukturen (*EK2.1*). Hierdurch werden redundante und sich widersprechende Daten vermieden (vgl. *EP4*).

Die Phase der *Speicherung* macht die neuen, noch nicht im Semantic Web vorhandenen semantischen Strukturen persistent. *Kriterien sind*: Unterstützung der gängigen Formate, wie RDF und OWL (*EK3.1*), sowie freier und performanter Zugriff auf die neu erstellten Daten über das Internet, z.B. per SPARQL⁶¹ (*EK3.2*).

Die letzte Phase, die *Verbreitung*, sorgt für die Verteilung der Information an die entsprechenden Stellen. Neu erstellte semantische Strukturen müssen mit bestehenden Strukturen geeignet verknüpft werden. *Kriterien sind*: Geeignete Unterstützung für die Suche nach passenden Anknüpfungspunkten (*EK4.1*).

5 Iterativer Zugriff auf semantische Strukturen

Wie bei der Erstellung lassen sich auch beim iterativen Zugriff auf semantische Strukturen verschiedene Phasen unterscheiden. Eine mögliche Unterteilung beinhaltet die folgenden sechs Phasen (Kuhlthau 1988): Initiierung, Selektion, Exploration, Fokussierung, Zusammenstellung und Präsentation. Wieder nutzen wir diese Einteilung, um die Tätigkeiten wäh-

⁶¹ SPARQL (SPARQL Protocol and RDF Query Language): <http://www.w3.org/TR/rdf-sparql-query/>

rend des iterativen Zugriffs auf semantische Strukturen geeignet zu gliedern (vgl. Abb. 3) und leiten aus den Tätigkeiten *Zugriffskriterien (ZK)* ab, um die in Abschnitt 2.2 analysierten Probleme zu vermeiden.

Die erste Phase, die *Initiierung*, beginnt mit der Wahrnehmung eines Informationsbedürfnisses und führt zu einer konkreten Definition des Problems im Bezug zu bereits vorhandenen Informationen und früheren Erfahrungen. *Kriterien sind*: Eine kontinuierliche Unterstützung schon bei der Problemdefinition (*ZK1.1*) und, ausgehend davon, Vorschläge, wie das Informationsbedürfnis am besten gestillt werden kann anhand von bereits früher durchgeführten Zugriffen (*ZK1.2*). Durch die Auflistung bereits früher erfolgreich gestellter Informationszugriffe, bzw. der entsprechend gestellten Anfragen, können sich Nutzer daran orientieren und vermeiden somit möglicherweise Probleme bei der Anfrageformulierung (vgl. *ZP1*).

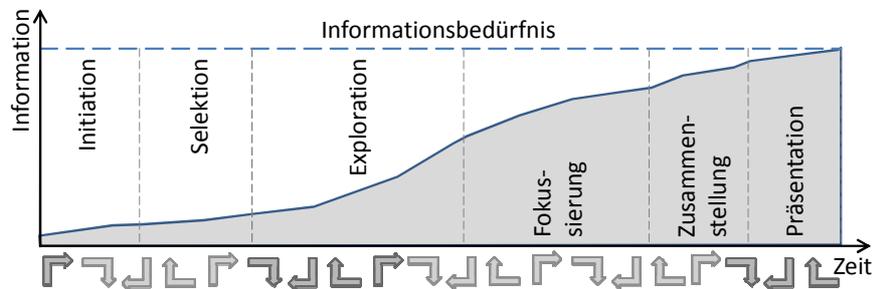


Abbildung 3: Die starke Kopplung beim iterativen Zugriff auf semantische Strukturen erzeugt viele Dialogzyklen (rotierende Pfeile auf Zeit-Achse; vgl. Pfeile in Abb. 2). Diese lassen sich in sechs Phasen untergliedern.

Die nächste Phase, die *Selektion*, beinhaltet die Aufgaben, das zu untersuchende Thema zu finden und auszuwählen. *Kriterien sind*: Ein Überblick über alle im Semantic Web verfügbaren Themen (*ZK2.1*), zum Beispiel in Form einer Karte in die hinein und wieder heraus gezoomt werden kann, zusammen mit automatisch generierten Vorschlägen für geeignete Themen anhand der eingegebenen Problembeschreibung, von Schlüsselwörtern oder sonstigen Anhaltspunkten, die vom Nutzer zur Verfügung gestellt werden (*ZK2.2*). Durch einen ständigen Abgleich von Nutzereingaben und entsprechenden Themenvorschlägen von Seiten des Computers, wird eine falsche Interpretation der Anfrage schnell bemerkt und kann daher frühzeitig korrigiert werden (vgl. *ZP2*).

Die Selektion wird gefolgt von der *Explorationsphase*. Diese beinhaltet die Untersuchung des gesamten Themenspektrums, die Begutachtung der vorhandenen Informationen und das in Bezug setzen zu bereits Bekanntem. *Kriterien sind*: Eine grafische Repräsentation der Information, die vom Nutzer verstanden werden kann (*ZK3.1*), intuitive und einfache Interaktionsmöglichkeiten (*ZK3.2*), die Möglichkeit Details bei Bedarf ein- und ausblenden zu können (*ZK3.3*), Sortier- und Blätterfunktionen für die Handhabung großer Datenmengen (*ZK3.4*) und Zoomfunktionen in Kombination mit Fokus- und Kontext-Techniken (*ZK3.5*). Durch eine interaktive Exploration können die, im Semantic Web falschen und fehlenden Inhalte möglicherweise bemerkt und durch andere Informationen ersetzt werden (vgl. *ZP3*).

Die *Fokussierung* beinhaltet Aufgaben wie die Formulierung von Hypothesen, die zur Anwendung bestimmter Filter führen und den Zielbereich damit immer weiter einschränken kann. Dies ist ein iterativer Prozess und benötigt daher mehrere Interaktionsschritte. *Kriterien sind*: Die Möglichkeit Filter interaktiv und intuitiv formulieren und ändern zu können (ZK4.1), die direkte Sichtbarkeit der Filterwirkung (ZK4.2), die Kombinationsmöglichkeit unterschiedlicher Filter (ZK4.3), die Unterstützung hierarchischer Filter (ZK4.4), die Rückverfolgbarkeit von Effekten auf die sie hervorrufenden Filter (ZK4.5) und die Möglichkeit, zu jeder Zeit den Zielbereich zu ändern (ZK4.6). Hierfür eignet sich insbesondere das Konzept der *facettierten Suche* (Hearst et al. 2002), bei der der Nutzer immer alle noch im Semantic Web verbleibenden Optionen zur Erstellung einer Suchanfrage angezeigt bekommt und diese nur noch auswählen muss, um sie seiner Anfrage hinzuzufügen. Somit werden automatisch ausschließlich bereits semantisch eindeutig definierte Klassen, Objekte und Eigenschaften für die Anfrage verwendet und so Mehrdeutigkeit vermieden (vgl. ZP1, ZP2 und ZP3).

Nach der Fokussierung findet die *Zusammenstellung* statt. Aufgaben sind das Selektieren und Zusammensetzen von für das fokussierte Thema relevanten Informationen. *Kriterien sind*: Einfache Mechanismen zur interaktiven Selektion relevanter Funde und deren Export in standardisierte Formate, sodass sie für eine weitere Nutzung, z.B. in anderen Systemen, zur Verfügung stehen (ZK5.1).

Die letzte Phase, die *Präsentation*, beinhaltet die Aufgabe, die gefundenen Informationen darzustellen. *Kriterium ist*: Eine breite Palette an Darstellungsmöglichkeiten (ZK6.1). Durch unterschiedliche Formen der Präsentation können Missverständnisse durch Fehlinterpretationen oder Fehler bei der Darstellung mit einzelnen Visualisierungstechniken verhindert werden (vgl. ZP4 und ZP5).

Anhand der Tools *mSpace* (Hearst et al. 2002) und *gFacet* (Heim et al. 2010) zeigen wir in Tabelle 1 exemplarisch, wie die beschriebenen Zugriffskriterien (ZK) bei der Bewertung bestehender Ansätze verwendet werden können. Beide Tools verwenden das Konzept der facettierten Suche und unterstützen damit die Formulierung eindeutiger Anfragen. Wie in Tabelle 1 zu sehen, findet jedoch insbesondere in den ersten und letzten beiden Phasen keine oder nur eine sehr unzureichende Unterstützung beim Zugriff auf semantische Strukturen statt.

Phasen	Initiierung		Selektion		Exploration					Fokussierung						Zusammenstellung		Präsentation
	ZK1.1	ZK1.2	ZK2.1	ZK2.2	ZK3.1	ZK3.2	ZK3.3	ZK3.4	ZK3.5	ZK4.1	ZK4.2	ZK4.3	ZK4.4	ZK4.5	ZK4.6	ZK5.1	ZK5.2	
mSpace	-	-	-	-	+	+	-	-	-	+	+	+	-	-	-	-	-	-
gFacet	-	-	-	+	+	+	+	+	-	+	+	+	+	+	+	-	-	-

Tabelle 1: Kriterien für eine optimale Unterstützung der in sechs Phasen ablaufenden Aktivitäten beim Zugriff auf semantische Strukturen und in wie weit diese von den Tools *mSpace* und *gFacet* erfüllt werden.

6 Zusammenfassung

In diesem Beitrag nehmen wir die bestehenden Probleme im Semantic Web als Ausgangspunkt und zeigen wie diese durch die Zusammenarbeit von Mensch und Computer vermieden werden können. Durch den Ansatz einer starken Kopplung zwischen Nutzer und Semantic Web werden Informationen nicht mehr in einem linearen, sondern in einem iterativen Prozess in kleinen Schritten ausgetauscht. Dabei gibt der Empfänger nach jedem Schritt eine Rückmeldung darüber, auf welche Weise er die Daten mit impliziter Bedeutung, in Daten mit expliziter Bedeutung übersetzt hat. Missverständnisse können dadurch frühzeitig erkannt, entsprechend korrigiert und so dem gewünschten Ergebnis mit jeder Iteration näher gekommen werden. Auch können die Zielvorstellungen noch während des Prozesses angepasst werden, um somit flexibel auf neue Erkenntnisse reagieren zu können. Ausgehend von einer detaillierten Analyse der Probleme, sowohl bei der Erstellung von, als auch beim Zugriff auf semantische Strukturen, schlagen wir verschiedene Kriterien für eine optimale Unterstützung dieser Tätigkeiten vor, die zur Bewertung aber auch zur Verbesserung von bestehenden und neuen Ansätzen eingesetzt werden können.

Literaturverzeichnis

- Battle, L. (2006). *Preliminary Inventory of Users and Tasks for the Semantic Web*. In: 3rd. Intl. Semantic Web User Interaction Workshop (SWUI 2006).
- Berners-Lee, T. & Fischetti, M. (1999). *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*. Harper, USA.
- Blumauer, A. & Pellegrini, T. (2009): *Social Semantic Web*. Springer, Berlin.
- Fähnrich, K.P. & Ziegler, J. (1984): *Workstations using Direct Manipulation as interaction mode - aspects of design, application and evaluation*. Proc. Interact 1984, Vol. II, S. 203-208.
- Hearst, M., English, J., Sinha, R., Swearingen, K. & Yee, P. (2002): *Finding the Flow in Web Site Search*. Communications of the ACM, 45 (9), S. 42-49.
- Heath, T, Dzbor, M. & Motta, E. (2005): *Supporting User Tasks and Context: Challenges for Semantic Web Research*. In: Proc. Workshop on End-User Aspects of the Semantic Web (UserSWeb).
- Heim, P., Ertl, T. & Ziegler, J. (2010): *Facet Graphs: Complex Semantic Querying Made Easy*. In: Proc. of the 7th Extended Semantic Web Conference (ESWC 2010), Springer.
- Kuhlthau, C.C. (1988). *Developing a model of the library search process: cognitive and affective aspects*. Reference Quarterly, S. 232-242.
- Schraefel, m.c., Smith, D., Owens, A., Russell, A., Harris, C. & Wilson, M. (2005). *The evolving mSpace platform: leveraging the semantic web on the trail of the memex*. In: Proc. of Hypertext 2005, ACM Press, S. 174-183.
- Turk, Z. (2006). *Construction informatics: Definition and ontology*. Advanced Engineering Informatics, Volume 20, Issue 2, S. 187-199.

Kontaktinformationen

Dipl.-Inf. Philipp Heim, Email: philipp.heim@vis.uni-stuttgart.de