

Memex360 – Persönliches Wissensmanagement mit Theseus ORDO

Björn Decker
Ralph Traphöner

Attensity Europe GmbH

Europaallee 10, 67657 Kaiserslautern, Germany

bjoern.decker@attensity.com ralph.traphoener@attensity.com

Abstract

Die Herausforderung, sich in der beständig zunehmenden Information zurechtzufinden, ist nicht neu. Durch verschiedene Applikationen - insbesondere des Web 2.0 - sind Lösungen entstanden, die einen Benutzer dabei unterstützen, einen Teil der für ihn relevanten Information zu verwalten und somit Wissen zu generieren. Allerdings bleiben die jeweiligen Informationen auf die jeweilige Applikation beschränkt. Mit Memex360 - einem Ergebnis des Use Cases ORDO im Forschungsprogramm THESEUS - wird eine integrative Schicht auf die Informationen implementiert, wobei die jeweiligen Werkzeuge weiter genutzt werden können. In dem Beitrag wird die Idee und Architektur von Memex360 vorgestellt.

1. Einleitung

Die Herausforderung, sich in der beständig zunehmenden Menge an Information zurechtzufinden, ist nicht neu. Bereits 1945 veröffentlichte Vannevar Bush die Idee des MEMEX [Bush, 1945]. Dieser MEMEX (memory extender) unterstützt einen Wissensarbeiter dabei, die für ihn relevante Information zu erfassen, zu annotieren und miteinander in Bezug zu setzen. Grundidee dabei war es, den Menschen bei der Verwaltung und Erschließung seines Wissens zu unterstützen – nicht den Menschen zu ersetzen.

Eine Vielzahl von Web 2.0 Applikationen ermöglicht es heute, Informationen zu erfassen, zu annotieren und mit einander in Bezug zu setzen. Blogs erlauben es, persönliche Notizen zu erfassen. Soziale Netzwerke helfen dabei, passende Ansprechpartner zu finden. Und nicht zuletzt Suchmaschinen – sei es auf dem Desktop oder im Web – sind in der Lage, umfangreiche Datenbestände schnell zu erschließen.

Allerdings stehen diese Lösungen in der Regel für sich allein. Erschließt sich ein Nutzer ein neues Wissensgebiet, kann er zum Beispiel eine Notiz zu einem Dokument in einem Blogbeitrag anlegen, oder in einer Annotation in dem jeweiligen Dokument oder der Notiz-Funktion des Bookmark-Dienstes. Welche Form der Annotation genutzt wird, hängt von den Vorlieben des Benutzers und seiner Situation ab – die Notiz bei seinem bevorzugtem Bookmark-Dienst wird z.B. unterwegs benutzt, während die Notiz in einem Dokument genutzt wird, wenn er keine Internet-Verbindung hat. Diese Annotationen stehen also nur in der jeweiligen Applikation zur Verfügung – die jeweilige Information bleibt also isoliert.

Doch nicht nur die Informationen in den Dokumenten bleiben isoliert. Auch das Wissensmodell, mit denen sich ein Benutzer Dokumente erschließt und klassifiziert, liegen in der Regel nur in der jeweiligen Applikation vor. Die in einem Personal Information-Management (PIM)-Werkzeug wie Outlook angelegten Kategorien liegen nur hier vor – um damit Blog-Einträge oder Bookmarks zu kategorisieren, müssten die Kategorien in der jeweiligen Applikation manuell gepflegt werden. Aufgrund des damit verbundenen hohen Aufwandes unterbleibt dabei meistens ein Abgleich der Kategorien.

Zusätzlich hat ein Nutzer in den verschiedenen Applikationen Informationen, welche die Erschließung von Dokumenten weiter unterstützen können. Beispielsweise können die Kontaktdaten im PIM oder in sozialen Netzwerken genutzt werden, um Firmennamen und Personen in relevanten Dokumenten zu identifizieren.

Die technische Grundlage, um diese Informationen mittels einer integrativen Schicht miteinander in Bezug zu setzen, liegt in den technischen Standards, mit denen die Inhalte in den jeweiligen Applikationen angesprochen werden können. Im Bereich der Blogs haben sich zum Beispiel die Blogger API [Winer, 2002] und die Meta-Weblog API [Anonymous] etabliert, um sowohl Inhalte und Kategorien abzufragen als auch zu erstellen. Durch RSS [Winer, 2003] Benachrichtigungen existiert ein Mechanismus, um Änderungen in einer Applikation an andere Applikationen oder den Benutzer zu kommunizieren. So können aktuelle Änderungen zeitnah von Suchdiensten in einer maschinell verarbeitbaren Form erfasst werden.

Memex360 implementiert diese integrative Schicht, welche die von einem Benutzer verwendeten Quellen integriert, miteinander in Bezug setzt und die Ergebnisse an den Benutzer meldet.

2. Memex360 – Idee und Architektur

Der Ausgangspunkt von Memex360 ist, dass der Benutzer seine bisher verwendeten Werkzeuge zum persönlichen Wissensmanagement weiterhin verwenden kann. Memex360 sorgt für die entsprechende Integration der unterschiedlichen Applikationen. Die entsprechende Architektur ist in der folgenden Abbildung dargestellt. Die obere Schicht (Mashups und Browser Integration) deckt mögliche Werkzeuge zum persönlichen Wissensmanagement ab. Von links nach rechts sind dies:

- PIM-Werkzeuge und RSS-Feed-Aggregatoren, um dem Benutzer beim Selbstmanagement zu unterstützen und über Neuigkeiten auf dem laufenden zu halten.

- Bookmarks-Verwaltungs-Werkzeuge, um Referenzen auf relevante Dokumente zu erfassen.
- Blogging oder Wikis Tools, um persönliche Notizen zu erfassen
- Visualisierungs-Tools, z. B. Mindmaps, um Informationen miteinander in Bezug zu setzen.

Mittels standardisierter Zugriffs-Mechanismen werden zum Einen die Inhalte dieser Applikationen durch die Service-Plattform erfasst und aufbereitet. Zum Anderen werden die Inhalte – wie ähnliche Dokumente oder Klassifikationen - in der Service-Plattform den verschiedenen Applikationen zur Verfügung gestellt. Die Service-Plattform hat dabei die folgenden vier grundlegenden Dienste:

- Der *Such-Service* indiziert zum Einen die relevanten Dokumente und reichert sie mit Metadaten an. Zum Anderen stellt dieser Service eine föderierte Suche zur Verfügung, um weitere Informationsquellen – durch Wolken angedeutet - zu integrieren. Neu hinzugekommene Informationen werden mittels konfigurierbarer RSS Feeds an den Nutzer weitergeleitet.
- Der *Referenz-Dienst* verwaltet die Referenzen des Benutzers auf relevante Dokumente. Weiterhin leitet der Referenz-Dienst neue Referenzen an den Such-Service zur weiteren Bearbeitung weiter.
- Der *Zettelkasten-Dienst* erfasst persönliche Notizen und setzt sie mit den Dokumenten aus dem Referenz-Dienst in Bezug.

- Den *Ontologie-Dienst*, um das persönliche Wissensmodell des Benutzers zu pflegen und den anderen Diensten zur Verfügung zu stellen. Die Ontologie wird entweder durch Hinzunahme von Metadaten in den Applikationen (z.B. neue Kategorien) oder dem Visualisierungs-Werkzeug gepflegt. Vorschläge für Erweiterungen der Ontologie können dabei durch die Analyse des Such-Services erstellt werden.

3. Die Erstellung einer Marktanalyse – Ein Beispiel für die Nutzung von Memex360

Herr X ist mit der Erstellung einer Marktanalyse zum Thema „Capsaicin“ beauftragt – einem Bestandteil des Chillis, welcher z.B. in der Schmerztherapie genutzt wird. Dazu nutzt er die Mindmap, welche er zu einem ähnlichen Thema bereits angelegt hatte. Ausgangspunkt sind die Fragen der bereits durchgeführten Marktanalyse zu Definitionen, ähnlichen Wirkstoffen, Studien und Anbietern. Die entsprechenden bereits angelegten Kategorien und Klassen in der Ontologie kann er wiederverwenden. Die neuen Einträge werden darüber hinaus mit der aktuellen Aufgabenstellung als Arbeitskontext verschlagwortet. Herr X sucht nun nach entsprechenden Begriffen – und richtet gleichzeitig entsprechende Benachrichtigungen auf dem Such-Service ein. So wird er später über Änderungen informiert. Relevante Treffer merkt er sich im Referenzdienst und fertigt eigene Notizen an. Via RSS stehen diese Notizen und Referenzen auch im Mindmap Tool zur Verfügung und können hier mit anderen Ergebnissen seiner

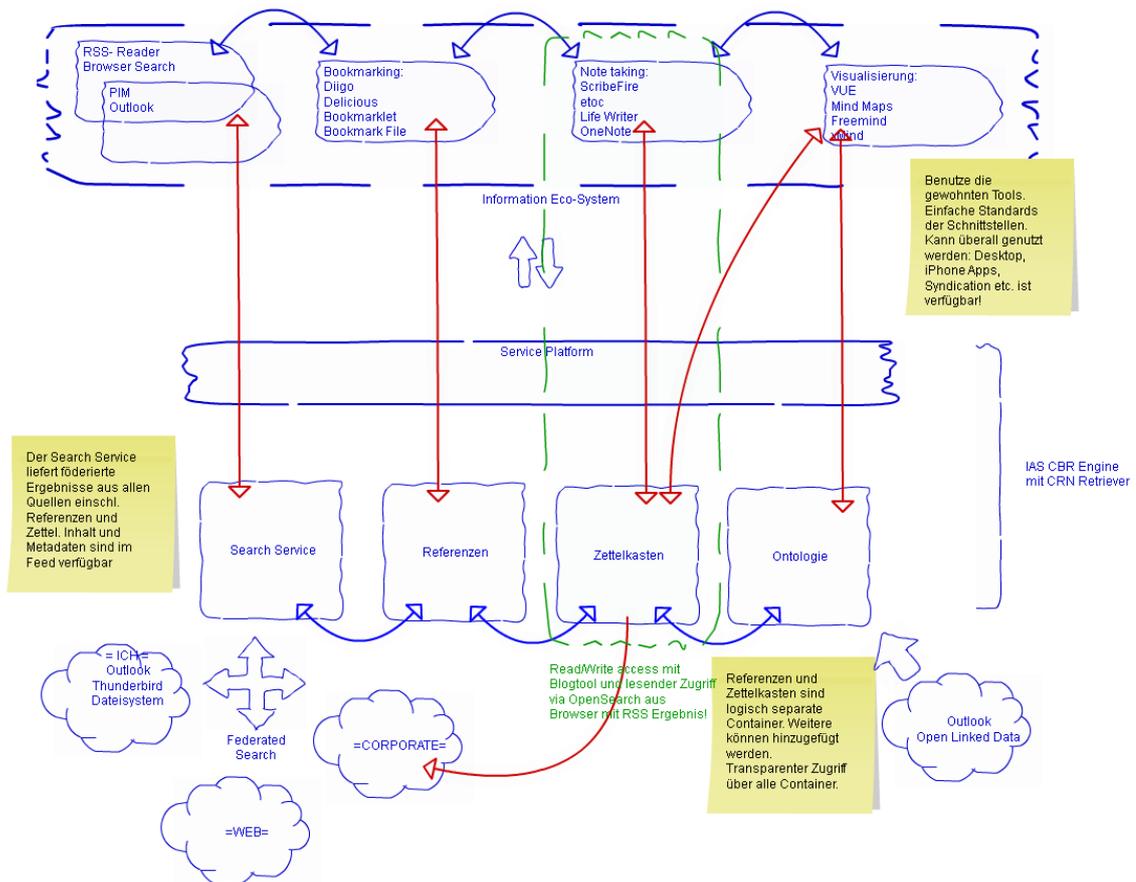


Abbildung 1: Memex360 Architekturskizze

Recherche in Bezug gesetzt werden.

Bei der Marktübersicht erstellt er unter anderem einen Überblick über Forschungsinstitute und Herstellerfirmen – jeweils von Herrn X angelegten Unterklassen von Anbieter. Vorschläge hierzu werden aufgrund der linguistischen Analyse der gefundenen Objekte – oder bereits vorhandenen Metadaten – unterbreitet.

Bestandteil der o.g. Klassen ist dabei eine Referenz auf eine Instanz der Klasse Ort. Auch hierfür können für den Anbieter entsprechende Vorschläge generiert werden. Die Ortsangabe wiederum kann genutzt werden, um über Open Linked Data (z. B. geonames.org) weitere Informationen zu ermitteln. So können die Geokoordinaten hinzugefügt werden, um die räumliche Verteilung der Anbieter zu analysieren.

Während der Recherche verfeinert Herr X auch an anderen Stellen sein Wissensmodell. Er findet z. B. immer wiederkehrende Personen, welche er als wichtige Autoren erkennt. Diese kann er z.B. in seinem Adressbuch ablegen und so in auch in die Ontologie einfügen. Auch richtet er neue Unterkategorien ein. Diese Erweiterungen können dann genutzt werden, um mit den bestehenden Anfragen weitere relevante Dokumente zu suchen und zu identifizieren.

Durch die zunehmende Anreicherung der Mindmap können darüber hinaus weitere, ähnliche Einträge in anderen Mindmaps identifiziert und zur Integration in die aktuelle Mindmap vorgeschlagen werden.

4. Prototypische Implementierung

Die Implementierung des Memex360 Prototyps hat den Status eines Proof of Concept. Dabei wird der Suchdienst durch eine CBR¹ Komponente zur Verfügung gestellt, d. h. es handelt sich im Kern um eine strukturierte fallbasierte Anwendung.

Alle Informationsobjekte, seien es Webseiten oder eigene Notizen und Dokumente, werden als strukturierte Fälle repräsentiert. Die erforderlichen Metadaten werden durch die Verschlagwortung des Anwenders und Ontologie basierte Informationsextraktion aus dem Inhalt des Informationsobjekts gewonnen. Letzteres erfolgt mit einer auf linguistischen Regeln basierenden Text Mining Engine (TME), die auch im Rahmen von THESEUS ORDO entwickelt wird.

Das Schema – die Ontologie – der strukturierten CBR Engine kann domänenspezifisch modelliert werden. Dabei ist es weitgehend unerheblich welcher Formalismus verwendet wird. Weitgehend in dem Sinne, das die genutzte IAS objektorientierte Modelle nutzt, die eine extensionale Klassendefinition voraussetzen. Eine bijektive Abbildung in OWL ist nicht möglich, da nicht alle OWL Konstrukte unterstützt werden. SKOS und RDFS Modelle sind abbildbar und Modelle können importiert werden.

Modellierung durch den Anwender bezieht sich dabei immer auf Instanzen. Schema, d. h. Klassen, können nicht zur Laufzeit bearbeitet werden.

Bewusst wurde Memex360 so ausgelegt, dass die Benutzung nicht an dedizierte Clients gebunden ist. Vielmehr bettet sich Memex360 nahezu unsichtbar in die bestehende Arbeitsumgebung ein:

- Referenzen und Bookmarks können mit allen gängigen Tools verwaltet werden, soweit diese ihre

Inhalte als RSS Feeds zur Verfügung stellen. Die Autoren nutzen Delicious² und Zotero³.

- Neue Bookmarks werden als solche erkannt und automatisch durch Crawler oder RSS Abonnement zur Fallbasis hinzugefügt.
- Die direkte Erstellung eigener Inhalte und Notizen kann mit Blogwerkzeugen erfolgen. Hierzu wurde Memex360 um die gängigen Blog API Protokolle erweitert. Damit wird z. B. Windows Live Writer zu einem komfortablen Editor für die eigene Fallbasis.
- Die Verschlagwortung von Inhalten erfolgt jeweils durch die nativen Tagging-Mechanismen der vorgenannten Werkzeuge. Wird dabei in z.B. Delicious ein neues Tag vergeben, so wird es in die Ontologie aufgenommen, steht dann im Live Writer als Schlagwort zur Verfügung und kann auch dort in den Kategoriebaum eingeordnet werden, der wiederum von der CBR-Engine als Ähnlichkeitsmaß interpretiert wird.
- Die Suche nach Inhalten erfolgt mit Hilfe des Open Search Interface direkt aus der Suchbox des Browsers. Dort wird Memex360 genauso als Suchmaschine zur Verfügung gestellt wie Bing oder Google. Die Treffer zu einer Suche werden als RSS Feed⁴ zurückgeliefert, der auch die strukturierten Falldaten enthält. Dies ermöglicht es insbesondere auch, eine Suche als Abonnement zu einem Standard RSS Feed Reader hinzuzufügen und damit bei neuen Treffern zur gespeicherten Suche auf dem Laufenden zu bleiben.
- Für die im Marktanalyse Szenario genannte Mindmap Funktionalität wurde die Visual Understanding Environment (VUE)⁵ genutzt. VUE verarbeitet RSS Feeds und erlaubt das strukturieren und verbinden von Inhalten und Metadaten per Drag und Drop aus einem RSS Objekt. Beinhaltet ein Fall im Trefferfeed z. B. eine neue erkannte Person, so kann diese als Konzept auf die Arbeitsfläche gelegt und anschließend mit der Klasse Personen verbunden werden. Noch zu realisieren ist ein Rückkanal von VUE in die Ontologie des Memex360, um modellrelevante Änderungen automatisch zu übernehmen. Alternativ wird zurzeit der Einsatz von Wise Mapping⁶ als Web basiertes Modellierungs- und grafisches Recherchewerkzeug geprüft.

Eine weitere wünschenswerte Erweiterung ist das Abonnement von Quellen für Instanzen aus der sogenannten Open Linked Data Cloud. Aktuell erfordert dies noch einen expliziten Import aus RDF Dateien oder einem Triplestore.

² <http://www.delicious.com>

³ <http://www.zotero.org>

⁴ RSS als Protokoll stellt kein Sicherheitsrisiko dar, solange das Memex360 Backend im Unternehmensnetzwerk betrieben und per VPN zugegriffen wird

⁵ <http://vue.tufts.edu/>

⁶ <http://www.wisemapping.org/>

The screenshot shows the Memex360 search interface for 'capsaicin'. The search results are displayed under the heading 'Ordo Ergebnisse' and 'Trefferkarten'. The results are categorized into 'Drug Intervention' and 'Medical Conditions'. A bar chart shows the frequency of various medical conditions. The table below lists the conditions and their corresponding number of results (Treffer).

Rubrik	Treffer
<input type="checkbox"/> Pain	82
<input type="checkbox"/> Burns	23
<input type="checkbox"/> Hypersensitivity	23
<input type="checkbox"/> Hyperalgesia	20
<input type="checkbox"/> Neuralgia	17
<input type="checkbox"/> Diabetes Mellitus	16
<input type="checkbox"/> Inflammation	16
<input type="checkbox"/> Smoking	14
<input type="checkbox"/> Herpes Zoster	13
<input type="checkbox"/> Arthritis	13
<input type="checkbox"/> Peripheral Nervous System Diseases	12

On the right side, there is a detailed view of a study titled 'Capsaicin Nanoparticle in Patient With Painful Diabetic Neuropathy'. The study objectives are to determine efficacy and safety of 0.75% topical capsaicin nanoparticle preparation versus placebo in patient with painful diabetic neuropathy. The study is a randomized, double-blind, placebo-controlled trial.

Abbildung 2: Memex360 - Suche nach Capsaicin mit JavaScript Interface

Neben der Benutzung über Werkzeuge mit Schnittstellen mit Web 2.0 Standards verfügt Memex360 über eine REST API im JSON Datenformat. Auf diese Schnittstelle können – in Ergänzung zur Benutzung durch die Standardwerkzeuge dedizierte Benutzerschnittstellen aufgesetzt werden. Mit dem JSON/REST API wurde z. B. eine dedizierte Suchschnittstelle für einen Demonstrator in der Life Science Domäne implementiert. Abbildung 2 zeigt das Interface, welches mit JQuery realisiert wurde.

5. Vergleichbare Ansätze

Mit diesen Funktionalitäten hat Memex360 das persönliche Wissensmanagement im Fokus. Aus Platzgründen werden im Folgenden zwei ausgewählte ähnlich gelagerte Ansätze präsentiert, welche aus Sicht der Autoren die stärkste Ähnlichkeit zum Memex360 Konzept aufweisen: Semantische Wikis und Semantische Desktops.

Durch semantische Wikis (z.B. [Oren et al 2006]) können persönliche Notizen als Wiki-Seiten – und damit mit einem eigenen GUI - mit Metadaten angereichert und durch Links miteinander in Bezug gesetzt werden. Externe Informationsobjekte können verlinkt werden. Damit kann ein semantischer Wiki Funktionen der zuvor genannten Standard-Applikationen übernehmen (z.B. Bookmarks in einer Liste strukturieren). Semantische Desktops (z.B. [Sauer mann, 2005]) stellen wiederum eine Infrastruktur und entsprechende GUI-Komponenten zur Verfügung, um Informationsobjekte mit semantischen Annotationen (Metadaten und Relationen) miteinander in Bezug zu setzen. Im Fokus dieser Auszeichnung sind dabei die auf dem Desktop verfügbaren Informationen. Beide Applikationsklassen enthalten dabei sowohl relevante Daten als auch Informationen über das Wissensmodell des Benutzers. Aus Sicht von Memex360 können diese Applikationsklassen als weitere Werkzeuge fungieren, um den Wissensraum des Benutzers zu erschließen. Sie stellen damit weitere Komponenten im Informationsökosystem dar, die aber nicht Memex360 als Monitor des Systems ersetzen.

6. Zusammenfassung und Ausblick

Memex360 integriert die unterschiedlichsten Informationsquellen und baut dabei auch Brücken zwischen diesen Quellen. Damit wird durch Memex360 der Dispersion der persönlichen Information begegnet.

Die vorgestellten Ideen und Konzepte werden im Rahmen des THESEUS ORDO Projektes weiterentwickelt.

Danksagung

Das diesem Bericht zugrundeliegende Vorhaben wurde mit Mitteln des Bundesministeriums für Wirtschaft und Technologie unter dem Förderkennzeichen 01MQ07005 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Literatur

- [Anonymous] Anonymous. Blogger APIs - Google Code. <http://code.google.com/apis/blogger/>.
- [Bush, 1945] Bush, Vanavar. 1945. As We May Think - Magazine - The Atlantic. The Atlantic (Juli): 47-61. <http://www.theatlantic.com/magazine/archive/1969/12/as-we-may-think/3881/>.
- [LOD] Linked Data Org. Linked Data | Linked Data - Connect Distributed Data across the Web. <http://linkeddata.org/>.
- [Winer, 2002] Winer, Dave. 2002. RFC: MetaWeblog API. März 14. <http://www.xmlrpc.com/metaWeblogApi>.
- [Winer, 2003] Winer, Dave. 2003. RSS 2.0 Specification (RSS 2.0 at Harvard Law). Juli 15. <http://cyber.law.harvard.edu/rss/rss.html>.
- [Oren et al 2006] Oren, Eyal et al. 2006. Semantic Wikis for Personal Knowledge Management (Database and Expert Systems Applications, LNCS). Springer Berlin
- [Sauer mann, 2005] Sauer mann, Leo. 2005. The semantic desktop-a basis for personal knowledge management (Proceedings of the I-KNOW).