

Wissensmanagement mit DYONIPOS: Konzepte, Algorithmen und Technologien

Michael Granitzer, Know-Center Graz, Österreich, mgrani@know-center.at

Josef Makolm, Bundesministerium für Finanzen, Österreich, Josef.Makolm@bmf.gv.at

Doris Ipsmiller, m2n, Österreich, ipsmiller@m2n.at

1. Einleitung

Workflow Management Systeme leisten einen wesentlichen Beitrag zur Produktivitätssteigerung und Unterstützung von Arbeitsprozessen in Organisationen. Nichtsdestotrotz stellen hochkomplexe, dynamische und vor allem sehr wissensintensive Aufgabenstellungen neue Herausforderung an solche Systeme. Vor allem die Vorgabe starrer Abläufe und die damit verbundenen geringen Freiheitsgrade, sowie die meist unvollständige Integration semi-strukturierter oder unstrukturierter Wissensressourcen führen zu Bruchstellen in der Wissensarbeit.

Zielsetzung des Forschungsprojekts DYONIPOS – DYNAMIC ONtology based Integrated Process OptimiSation – ist die Entwicklung eines kontextsensitiven, intelligenten und flexiblen Systems, welches zu einer signifikanten Produktivitätssteigerung in wissensintensiven Unternehmen beiträgt. Basierend auf semantischen Technologien und Methoden des Knowledge Discovery zielt DYONIPOS auf die Analyse und Integration von dynamischen, undokumentierten, ad-hoc Wissensflüssen ab. Semantische Technologien ermöglichen dabei eine flexible und einheitliche Behandlung von in der Organisation strukturiert und unstrukturiert vorliegenden Daten und deren Verwendung in wissensintensiven Geschäftsprozessen. DYONIPOS adressiert zwei entscheidende Arbeitsbereiche der Wissensarbeit in agilen Organisationen: den einzelnen Prozessausführenden bei der effizienten, standardisierten aber dennoch flexiblen Abarbeitung von Geschäftsprozessen gemäß seiner individuellen Arbeitsweise einerseits und den Prozessingenieur bei der Entwicklung standardisierter Prozesse, basierend auf tatsächlich von den unterschiedlichen Wissensarbeitern ausgeführten Geschäftsprozessinstanzen andererseits. Konkret lassen sich daraus folgende technisch/wissenschaftlichen Fragestellungen ableiten:

1. die semantische Anreicherung/Harmonisierung von Informationsbeständen
2. die semantische Erfassung von Abläufen und Tätigkeiten in Unternehmen
3. die Entwicklung pro-aktiver, kontextsensitiver Information Delivery Methoden
4. die Entwicklung von Werkzeugen zur Analyse von Wissensflüssen und ad-hoc Tätigkeiten in Unternehmen

Der vorliegende Beitrag beschreibt das DYONIPOS System, dessen zentrale Komponenten sowie die Ergebnisse der durchgeführten Evaluierung. Die Evaluierung der Komponenten erfolgte in einem Pilotbetrieb mit ausgewählten Benutzern beim Use-Case Partner, dem österreichischen Bundesministerium für Finanzen, Sektion V. In diesem hochkomplexen, wissensintensiven Umfeld konnten Evaluierungsergebnisse gesammelt und die Anwendbarkeit von DYONIPOS überprüft werden. DYONIPOS wird nun bis Ende 2009 in der Sektion V ausgerollt und steht für den weiteren Einsatz zur Verfügung.

2. DYONIPOS Komponenten

Um einen Überblick über die verschiedenen entwickelten DYONIPOS Komponenten geben zu können, wird der in DYONIPOS umgesetzte Prozess in Abbildung 1 dargestellt.

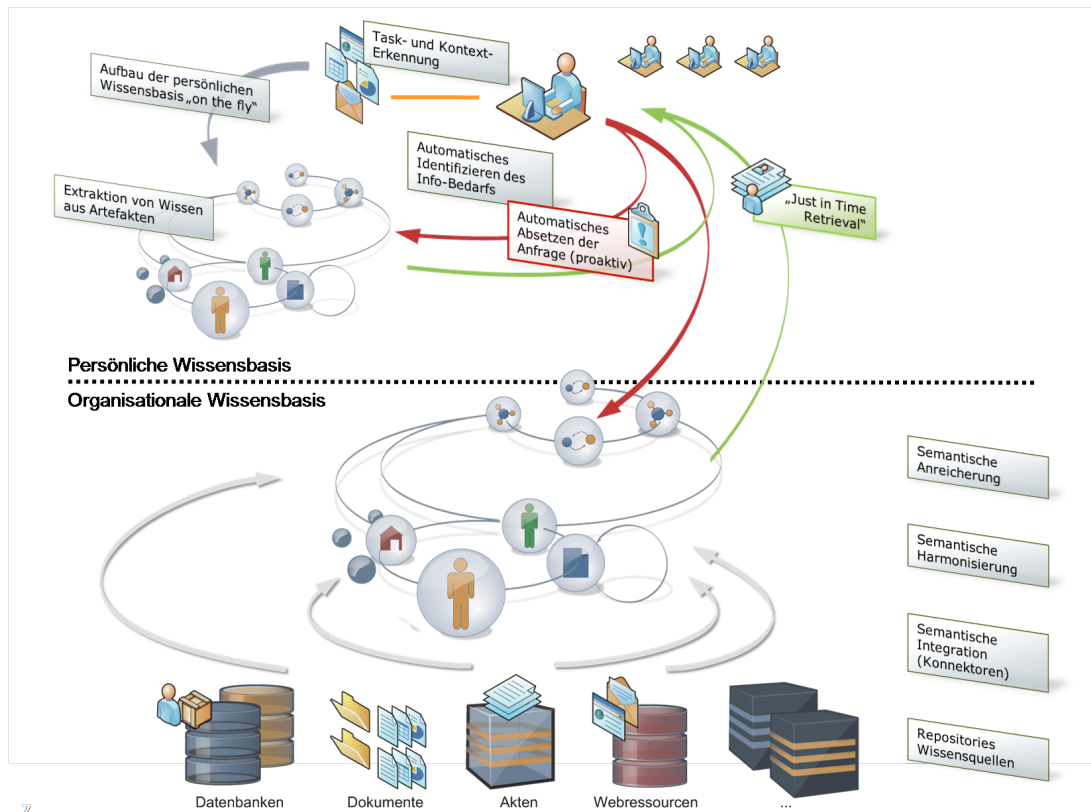


Abbildung 1: Überblick DYONIPOS Workflows.

DYONIPOS analysiert die aktuelle Tätigkeit des Wissensarbeiters, erkennt mit welchen Tasks sich dieser gerade beschäftigt und welche Informationsressourcen dafür herangezogen werden. Daraus wird der aktuelle Arbeitskontext abgeleitet. Es erfolgt die Überführung in ein einheitliches, RDF-basiertes Format, die on-the-fly-Indizierung der erkannten Informationsressourcen sowie die Überführung der semantisch angereicherten Ressourcen in die persönliche Wissensbasis des Benutzers. Diese ist aus datenschutzrechtlichen Gründen ausschließlich für den jeweiligen Wissensarbeiter einsichtig und enthält neben für den Wissensarbeiter relevanten Informationsressourcen auch die von ihm durchgeführten Tätigkeiten (Tasks).

DYONIPOS erkennt die Arbeitsschritte des Benutzers und liefert ihm pro-aktiv relevante Ressourcen aus der persönlichen und der organisationsweiten Wissensbasis sowie zum Kontext passende Konzepte wie etwa Mitarbeiter oder andere Organisationen, die sich mit ähnlichen Themenstellungen beschäftigen. Die organisationsweite Wissensbasis enthält dabei alle unternehmensrelevanten Informationen. Sie harmonisiert die organisationsweit verfügbaren Informationsquellen mittels semantischer Technologien äquivalent zur persönlichen Wissensbasis. Das so aufgebaute „organisationale Gedächtnis“ bietet somit einen konsolidierten Zugriff auf diverse Artefakte wie etwa Dokumentenbestände, Personeninformation, Firmeninformation und standardisierte Prozessabläufe.

Der Wissensarbeiter kann spezielle Informationen, wie etwa interessante Webseiten, für die organisationale Wissensbasis freigeben. Durch Integration der aus der persönlichen Wissensbasis freigegeben Informationen erfolgt eine ständige Erweiterung der Unternehmens-Wissensbasis. Die gewonnenen Erkenntnisse und neuen Ressourcen stehen wiederum allen Mitarbeitern zur Verfügung – der Kreislauf schließt sich.

Neben der Rolle des Wissensarbeiters, ist der Prozess-Ingenieur (bzw. weiter gefasst der Knowledge Engineer), für die Optimierung der organisationalen Wissensflüsse und Prozesse verantwortlich. Für den Wissensarbeiter besteht die Möglichkeit, seine, mittels DYONIPOS erfassten Tasks aus seiner persönlichen in die organisationale Wissensbasis freizugeben. Die sich so ständig erweiternde Wissensbasis erlaubt dem Knowledge Engineer die Analyse der

durchgeführten Tätigkeiten und der Ableitung von möglichem Optimierungspotential hinsichtlich Prozessabarbeitung und Informationsbereitstellung.

2.1. Datenaufbereitung und Vorverarbeitung

Semantische Harmonisierung von heterogenen Datenquellen: Die semantische Harmonisierung heterogener Datenbestände ist Voraussetzung für effiziente Single-Point-of-Access-Zugriffsmöglichkeiten auf Informationsquellen [1]. Unter Anwendung von RDFS bzw. OWL DL erfolgt die Abstraktion unterschiedlicher Datenquellen wie dem ELAK-System, verschiedenen Dokument-Repositories, Web-Ressourcen und Asset-Management-Systemen. Die Informationsressourcen werden mittels eines flexiblen Konnektoren-Prinzips im Rahmen des semantischen DataNetworks in ein im Projekt entwickeltes RDF/OWL Schema überführt und unter Anwendung spezieller Indizierungsmethoden suchbar gemacht. Die zugrunde liegenden Mechanismen sind dabei flexibel und konfigurierbar gehalten, so dass eine Erweiterung sehr einfach möglich ist.

Semantische Anreicherung: Neben der oben beschriebenen Harmonisierung von Datenquellen erfolgte die Anreicherung der Datenbestände unter Anwendung von Text Mining Methoden. Aus den vorliegenden unstrukturierten Texten werden Instanzen von Konzepten wie Personen, Orte, Organisationen/Abteilungen und Begriffsnetze extrahiert sowie deren Semantik disambiguiert. Vor allem die Personenerkennung weist aus Gründen des Datenschutzes die Notwendigkeit einer genauen Identifikation von im System bekannten („Whitelist“), d.h. eindeutig bestimmbar Personen auf. Um Verknüpfungen zwischen solchen benannten Entitäten sowie zwischen Entitäten und allgemeinen Konzepten abbilden zu können, erfolgte die Entwicklung/Verfeinerung von statistischen, co-occurrence basierten Verfahren. Diese Verfahren dienen als Grundlage zur Berechnung eines assoziierten Index, welcher Relationen zwischen Entitäten und wichtigen Konzepten abbildet.

2.2. Unterstützung des Wissensarbeiters

Für eine möglichst effiziente und punktgenaue Informationsbereitstellung für den Wissensarbeiter stellt DYONIPOS relevante Dokumente und mit dem aktuellen Arbeitskontext assoziierte Personen, Organisationseinheiten etc. pro-aktiv bereit. Aus dieser Aufgabenstellung ergibt sich der zentrale Aspekt der Erfassung des Arbeitskontextes respektive der aktuellen Aufgabe eines Wissensarbeiters und deren Überführung in eine formale, RDF-basierte Struktur. Die Formalisierung geschieht einerseits vor dem Hintergrund der Weiterverarbeitung bei der Taskanalyse und des Prozess-Minings andererseits zur Identifikation von relevanten, im aktuellen Kontext notwendigen Ressourcen. Die Bereitstellung der entsprechenden Ressourcen erfolgt über Just-In-Time Retrieval Techniken [7]. Eine explizite Formulierung von Suchanfragen durch den Benutzer ist somit nicht mehr notwendig, im Bedarfsfall jedoch möglich. Die Suche selbst erfolgt auf dem semantisch harmonisierten Datenbestand unter Berücksichtigung der unterschiedlichen semantischen Metadaten [4,5].

Kontexterfassung: Um auf die abstrakte Ebene des Benutzerkontexts gelangen zu können, werden alle Benutzerinteraktionen mit dem System sowie die daraus resultierende Reaktion des Systems erfasst. Diese Interaktionen werden im Folgenden als 'Events' bezeichnet. Maus Clicks, Maus Bewegungen oder Tastatureingaben repräsentieren Beispiele für solche Events. Eigens für unterschiedliche Applikationen entwickelte Sensoren erweitern die den Events zugrunde liegende Information um z.B. Inhalte die gerade mit der Maus markiert oder Inhalte die in die Zwischenablage kopiert wurden. Diese Sensoren ermöglichen nicht nur das Abgreifen der inhaltlichen Komponenten sondern auch struktureller Metadaten (z.B. Autor, Seitenanzahl eines Dokuments, Speicherort eines Dokuments oder die Webadresse).

Die Menge an Events, die auf dieser untersten Ebene aufgenommen wird, ist für eine weitere Verarbeitung nicht mehr handhabbar. Daher kommen Filter- und Aggregationsmechanismen zur Überführung der Events in aussagekräftigere Strukturen zum Einsatz. Neben dem Verwerfen nicht aussagekräftiger Events, wie beispielsweise Mausbewegungen, erfolgt die Aggregation der Sequenzen von Einzelevents zu so genannten Eventblöcken auf Basis von expliziten Regeln. Ein Indikator für das Ende eines Eventblocks stellt beispielsweise ein Wechsel der Applikation sowie das Wechseln der Domäne beim Gebrauch eines Internet Browsers dar.

Die Zuweisung von Eventblöcken zu Tasks/Aufgaben resultiert aus einem inhaltlichen Vergleich der Menge an Eventblöcken. Die Annahme, dass Eventblöcke, die sich inhaltlich sehr ähnlich sind, zu ein und demselben Taskcluster gehören, hat sich im Laufe der Untersuchungen bestätigt. Erst durch das aktive Eingreifen des Benutzers, sprich durch das Benennen eines vorgeschlagenen Taskclusters, entstehen Benutzertasks (siehe [3,6] für Details). Das Resultat dieses automatischen Prozesses wird als erster Vorschlag für den Benutzer verstanden. Dieser kann die resultierenden Tasks überprüfen, gegebenenfalls abändern und benennen. Das Benutzerfeedback ermöglicht es dem System einerseits, den Abstraktionsprozess zu evaluieren und kontinuierlich zu verbessern, andererseits werden auf diese Weise Trainingsdaten erzeugt.

Kontextverarbeitung: Der erfasste Kontext inkl. dem gerade bearbeiteten Inhalt des Benutzers wird wie oben beschrieben semantisch harmonisiert, anonymisiert und in die persönliche Wissensbasis überführt. Als Speichermedium für die aufgezeichneten Informationen und extrahierte Metadaten fungiert clientseitig ein Index, der auf dem Open Source Project Lucene basiert. Die Verwendung eines Index erleichtert dabei sowohl Suche, als auch Verwaltung größerer Informationsmengen.

Kontextbasiertes Retrieval: Die Suchfunktionalität erstreckt sich nicht nur auf die persönliche Wissensbasis – also jene Ressourcen die am Client durch den Benutzer bearbeitet und durch DYONIPOS indiziert wurden – sondern schließt auch die organisationsweite, semantisch harmonisierte Wissensbasis ein. Während das aktive Absetzen von Suchen ebenfalls zu den entwickelten Retrievalmöglichkeiten zählt, fokussiert DYONIPOS jedoch auf das pro-aktive anliefern von relevanten Ressourcen. Dazu wird der aufgezeichnete Benutzerkontext für Retrievaloperationen aufbereitet, indem die in den Event-Blöcken ermittelten Inhalte und Metadaten als Ähnlichkeitssuche Anwendung finden [5].

Kritisch hierbei ist vor allem die Identifikation des Zeitpunktes, an dem Information vom Benutzer benötigt wird [2], also ein Informationsbedürfnis vorliegt. In diesem Projekt wurden mehrere Ansätze bzw. mehrere Indikatoren erprobt, die eine solche Suchoperation auslösen, wobei der Fokus auf der Identifikation geeigneter Regeln basierte. Identifiziert DYONIPOS beispielsweise eine Suche im WWW, so durchsucht es ebenfalls die persönliche und globale Wissensbasis und bietet dementsprechend Ressourcen an.

4. Modellbasiertes Applikationsdesign

Semantische Technologien kommen im Rahmen von DYONIPOS neben dem Aufbau der Datenmodelle sowie der Datenharmonisierung für die Definition der gesamten Applikationslogik zum Einsatz. Welche (Benutzer-)Interaktionen welche internen Programmflüsse anstoßen bzw. wie diese interne Ablauflogik gestaltet ist, wird durch semantische Modelle festgelegt, die on-the-fly – daher ohne Überführung in Code bzw. Kompilierung – mittels der „ExecutionEngine“ ausgeführt werden. Die ExecutionEngine steuert somit alle Systemkomponenten bzw. Services semantisch: von den Indizierungs-Abläufen bis hin zur Interaktion mit dem Benutzer, wobei das Interface selbst wiederum rein modellbasiert aufgebaut ist. DYONIPOS kann sich somit enorm flexibel und rasch an diverse Anforderungen und Anwendungsgebiete anpassen und erfüllt, gemeinsam mit dem flexiblen

DataNetwork, die Anforderungen hinsichtlich der Übertragbarkeit des Systems für einen weiteren Produktiveinsatz.

5. Zusammenfassung

Im Beitrag wurde das DYONIPOS System zur pro-aktiven Bereitstellung von Informationsressourcen und zur dynamischen Erfassung wissensintensiver Task vorgestellt. Algorithmische und Use-Case Evaluierungen im Bereich des E-Governments zeigen das Potential der entwickelten Techniken zur Verbesserung der Wissensarbeit in hochkomplexen Anwendungsdomänen.

Literaturangaben

- [1] Bergamaschi, S., Castano, S., Vincini, M., and Beneventano, D. 2001. Semantic integration of heterogeneous information sources. *Data Knowl. Eng.* 36, 3 (Mar. 2001), 215-249. DOI= [http://dx.doi.org/10.1016/S0169-023X\(00\)00047-1](http://dx.doi.org/10.1016/S0169-023X(00)00047-1)
- [2] Granitzer, M., Kienreich, W., Sabol, V., Andrews, K., and Klieber, W. Evaluating a system for interactive exploration of large, hierarchically structured document repositories. In Proc. IEEE Symposium on Information Visualization INFOVIS 2004, pages 127–134, 2004.
- [3] Iqbal, S. T. & Bailey, B. P. (2007), Understanding and developing models for detecting and differentiating breakpoints during interactive tasks, in 'CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems', ACM Press, New York, NY, USA, pp. 697—706.
- [4] Kröll, M., Rath, A. S., Granitzer, M., Lindstaedt, S., and Tochtermann, K. Contextual retrieval in knowledge intensive business environments. In Workshop Information Retrieval of the Special Interest Group Information Retrieval (FGIR), 2007.
- [5] Lux, M.; Tochtermann, K. & Granitzer, M. (2006), Retrieval basierend auf Semantischen Metadaten, in 'Tagungsband der Multikonferenz Wirtschaftsinformatik 2006, Passau (D)'.
[6] Oliver, N.; Smith, G.; Thakkar, C. & Surendran, A. C. (2006), SWISH: semantic analysis of window titles and switching history, in 'IUI '06: Proceedings of the 11th international conference on Intelligent user interfaces', ACM, New York, NY, USA, pp. 194--201.
- [7] Rath, A. S.; Kröll, M.; Lindstaedt, S. & Granitzer, M. Gronau, P. D. N., ed. (2007), *Low-Level Event Relationship Discovery for Knowledge Work Support*, Vol. 2, GITO GmbH, Berlin.