

### **3 Weitere Projekte aus dem Umfeld**

#### **3.1 DAFFODIL - Nutzerorientiertes Zugangssystem für heterogene Digitale Bibliotheken**

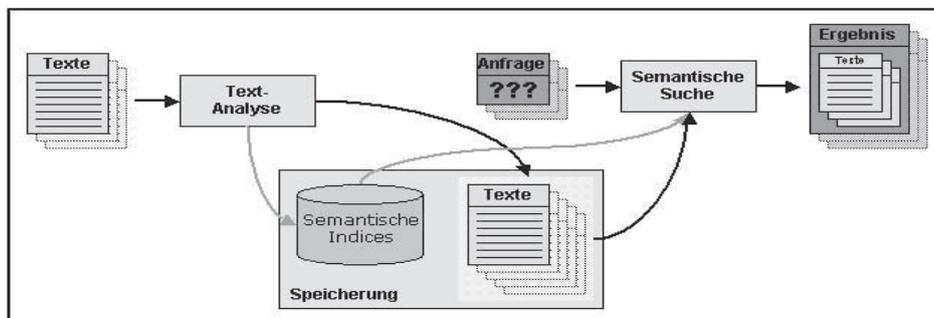
Daffodil ist ein virtuelles digitales Bibliotheks-System, das die strategische Unterstützung des Suchenden während des gesamten Information-Retrieval-Prozesses zum Ziel hat. Für einen Benutzer äußert sich diese in Form von höheren Suchfunktionen, sogenannten Strategemen, die Funktionalitäten über das allgemein übliche Maß hinaus bereitstellen. Der aktuelle Prototyp kann über [www.daffodil.de](http://www.daffodil.de) genutzt werden. Dazu wird via Java Webstart Technologie die grafische, intuitiv bedienbare Benutzeroberfläche gestartet. Zur Zeit existieren eine Vielzahl von Diensten zum Suchen und Browsen in Digitalen Bibliotheken, sowie Dienste zur Verwaltung von gefundenen Objekten. An zentraler Stelle steht das Suchwerkzeug, das über eine formularbasierte Schnittstelle eine einheitliche Anfrageformulierung über aktuell 15 verteilte Datenquellen bereitstellt. Dabei wird die Anfrageformulierung durch proaktive Dienste unterstützt, um schon vor der eigentlichen Anfrageprozessierung Fehler zu erkennen, bzw. dem Benutzer Vorschläge zur aktuellen Anfrage zu unterbreiten. Das Ergebnis einer Anfrage wird anschließend zusammengefasst und einheitlich dargestellt.

Weitere Dienste zum Suchen und Browsen sind der Referenz- und Zitationsbrowser, der Journal- und Konferenzbrowser, das Coautorenwerkzeug, neben einer Reihe von unterstützenden Werkzeugen wie einem Thesaurus und dem Klassifikationsbrowser. Zur nachhaltigen Speicherung und Verwaltung von gefundener Information von Benutzern und Gruppen dient die persönliche Handbibliothek. Dort können alle bereitgestellten Digitalen Bibliotheksobjekte, darunter Metadaten, Volltexte, Autoren, Begriffe, Webseiten, etc. in strukturierter Weise abgelegt werden. Das Konzept der Awareness ermöglicht Gruppenarbeit und Langzeitverfolgung von Anfragen. Aktuell wird Daffodil im Rahmen des EU-Projekts DELOS zur Evaluation von Digitalen Bibliotheken als Basisframework eingesetzt werden, um grafische Werkzeuge, Dienste oder ganze Digitale Bibliotheken mit- und untereinander zu vergleichen. Zudem werden zur Zeit weitere Werkzeuge zur Kollaboration von Suchenden untereinander oder mit Experten (z.B. Bibliothekaren) in das System integriert, um den Benutzern effektivere Möglichkeiten zu bieten, ein bestehendes Informationsbedürfnis zu bearbeiten. Das Projekt Daffodil wurde durch die (DFG) innerhalb des Schwerpunktprogramms "Distributed Processing and Delivery of Digital Documents"(V 3D2) gefördert.

#### **3.2 Semantik-basiertes Retrieval Biosystematischer Dokumente**

Im Projekt "Collaborative Research: Development of New Digital Library Applications in the Context of a basic Ontology for Biosystematics Information Using the Literature of Entomology (Ants)" werden derzeit biosystematische Dokumente in großem Umfang digitalisiert. Um die sukzessive entstehende Kollektion sinnvoll nutzen zu können, müssen die Texte in XML annotiert und für den Zugriff über ein Retrieval-System bereitgestellt werden. Ohne die Annotationen würde sich das Retrieval als extrem schwierig

gestalten, da sich die Dokumente in Form, Sprache und Wortwahl sehr ähnlich sind, bestehen sie doch zum größten Teil aus anatomischen Beschreibungen von Insekten, die den einzelnen Körpermerkmalen Beschreibungen zuordnen, aus ihren Taxonomischen Namen und ihren Fundorten. Auf dem unmarkierten Text wären klassische Techniken für Volltext-Retrieval (etwa Boolesches oder Vectorspace-Retrieval) gerade für die Suche nach Insekten mit bestimmten Körpermerkmalen (bzw. Dokumenten, die Insekten mit bestimmten Merkmalen beschreiben) nicht geeignet, da sich hier die Zusammengehörigkeit von Merkmal und Beschreibung nicht hinreichend abbilden lässt. Zudem würden die Taxonomischen Namen aufgrund ihrer sehr geringen Dokumentfrequenz (im Idealfall ist diese eins, da die Namen ein Insekt eindeutig identifizieren) aus jedem klassischen Volltextindex herausfallen, wodurch zum Auffinden der Beschreibung einer ganz bestimmten Spezies eine aufwändige Volltextsuche ohne die Unterstützung durch einen Index ausgeführt werden müsste.



Auf der anderen Seite ist die manuelle Annotation sehr zeitaufwändig und daher kostenintensiv. Daher wird an der Universität Karlsruhe ein System entwickelt, das sowohl ein automatisches Markup leistet, als auch die Dokumente speichert und über eine Retrieval-Engine zugreifbar macht, wobei die markierten Teile zur Indizierung genutzt werden. Die Annotation stützt sich auf mehrere dynamisch aktivier- und entfernbar Analysekomponenten, die unter anderem Pattern-Matching und NLP-Techniken einsetzen, um semantische Einheiten in den Dokumenten zu identifizieren und zu markieren. Derzeit umfasst dies Körpermerkmale und die ihnen zugeordnete Beschreibung, Ortsangaben und die aus mehreren Wörtern bestehenden Taxonomischen Namen.

## Ausblick

Mit *io-port.net* steht am Projektende ein Kompetenz- und Dienstleistungsnetz für die Informatik zur Verfügung, das weltweit publiziertes Informatikwissen mit ca. 2 Millionen Einträgen strukturiert und standardisiert nachweist, Zugang zu Volltexten direkt (u.a. LNI-Reihe der GI) oder per Dokumentlieferdienste anbietet sowie weitere thematisch fokussierte Angebote der Informatik (Fachgruppen, Kompetenznetze) und personalisierte Dienste (Erfassung persönlicher Publikationslisten) integriert. *io-port.net* verwendet dabei eine tiefe Datenintegration, die über eine für viele Portale typische lose Kopplung verschiedener Anbieter weit hinausgeht und ermöglicht den Nutzern einheitli-

che Anfragen über einen Datenbestand in definierter Qualität. Semantische Werkzeuge unterstützen dabei sowohl die Nutzer bei Suche und Navigation in den Inhalten als auch den Portalbetreiber bei Aufbau und Betrieb des Portals.

Das Angebot des Informationsportals *io-port.net* ist derzeit fokussiert auf Endnutzer in Hochschulen und Forschungseinrichtungen. Um diese Nutzer langfristig zu halten und in das Portal aktiv einzubeziehen sowie neue Nutzer bzw. Nutzergruppen ansprechen zu können, ist die Realisierung weiterer, innovativer Mehrwertdienste notwendig. Im Vordergrund sollten dabei Mehrwertdienste stehen, die den Nutzer bei allen Schritten der Informationsbeschaffung und vor allem auch der Informationsorganisation am Arbeitsplatz effizient unterstützen: v.a. bei Recherche, Analyse und Bewertung von Ergebnismengen, beim direkten Zugriff auf relevante Quellen, bei der individuellen Informationsorganisation sowie der Kollaboration mit Fachkollegen.

## Literaturverzeichnis

- [Ap05] Apps, A: Guidelines for Encoding Bibliographic Citation Information in Dublin Core Metadata, Dublin Core Metadata Initiative, June 2005, <http://dublincore.org/documents/dc-citation-guidelines/>
- [Cu05] Cusomano, M.A: Google: what it is and what it is not. Communication of the ACM 48(2): 15-17 (2005)
- [DCES] Dublin Core Metadata Element Set, Version 1.1: Reference Description, Dublin Core Metadata Initiative, December 2004, <http://dublincore.org/documents/dces/>
- [DCMI] Dublin Core Metadata Terms, Dublin Core Metadata Initiative, June 2005, <http://www.dublincore.org/documents/dcmi-terms/>
- [KLRWW04] Klink, S.; Ley, M.; Rabbidge, E.; Reuther, P.; Walter, B. Weber, A.: Browsing and Visualizing Digital Bibliographic Data. Symposium on Visualization, Konstanz, Germany, May 19-21, Eurographics Association, 2004; S. 237-242.
- [Ko04] Koschmider, A. et al.: Entwicklung eines Informationsportals für die Informatik, In P. Dadam; M. Reichert, INFORMATIK 2004 - Informatik verbindet, GI-Jahrestagung, Ulm, volume 50 of LNI, pp. 208-213. September 2004.
- [Ley02] Ley, M.: The DBLP-Computer Science Bibliography: Evolution Research Issues, Perspectives. In 9th International Symposium SPIRE, 2002; S. 1-10.
- [LF04] Lehti, P.; Fankhauser, P.: XML Data Integration with OWL: Experiences and Challenges. In Proceedings of the 2004 Symposium on Applications and the Internet (SAINT 2004); 26-30 January 2004, Tokyo, Japan
- [LSF04] Lehti, P.; von Stackelberg, S., Fankhauser, P.: The Semantic Web Query Language SWQL; 2004, <http://www.ipsi.fraunhofer.de/oasys/projects/semiport/SWQL-WD0704.doc>
- [SJ04] Schwarzkopf, E.; Jameson, A.: Personalized Support for Interaction with Scientific Information Portals. AMR 2003, LNCS3094, pp. 58-71, 2004
- [Sc04] Schwarzkopf, E.: Enhancing the Interaction with Information Portals. Proceedings of IUI 2004, pp. 322-324, 2004
- [WK98] Weibel, S.; Kunze, J.; Lagoze C.; Wolf, M.: Dublin Core Metadata for Resource Discovery, RFC 2413, September 1998, <http://www.ietf.org/rfc/rfc2413.txt>