

sources erst nach der eigentlichen Extraktion zulassen¹, zielt der von uns entwickelte Ansatz auf eine Bewertung von Ressourcen während der eigentlichen Suche ab. Dies ermöglicht eine effektive und effiziente Nutzung vorhandener Kapazitäten². Die fokussierte Suche nach Ressourcen basiert auf der Bestimmung der Relevanz einer Ressource zu einer bestimmten Domäne oder Teildomäne. Generell wird als Suchstrategie die Verfolgung von Relationen zu Ressourcen mit möglichst hoher Relevanz verwendet (fokussiertes Crawlen). Die Berechnung der Relevanz setzt sich zusammen aus der **inhaltlichen Analyse** des Dokumenteninhalts und einer **Bewertung der Verlinkungsstruktur** der Ressource. Das Ergebnis ist ein numerischer Wert, welcher zur Erzeugung einer sortierten Menge an Ressourcen verwendet wird. Diese Menge dient zur weiteren Suche, wobei Ressourcen mit höherer Relevanz vorrangig verarbeitet werden.

Zusammenfassend werden Ontologien zur Modellierung der Umgebungswelt und zur Modellierung der eigentlichen Anwendung (Domäne) eingesetzt. Die Domänen-Ontologie beschreibt dabei Konzepte zu denen weitere Instanzen identifiziert werden sollen. Die Suche nach diesen Instanzen wird mittels einer semantischen Bewertung durchgeführt.

2.2 Personalisierte Benutzerinteraktion mit wissenschaftlichen Informationsportalen

Die Interaktion mit Information besteht nicht nur aus Information Retrieval, sondern auch aus dem Organisieren und Verstehen der gesammelten Daten. Gerade bei wissenschaftlichen Informationssystemen ist eine entsprechende Werkzeugunterstützung wertvoll. Wird solch ein Werkzeug in ein Informationsportal integriert, bietet es zudem die Möglichkeit, dem Benutzer bei der Informationssuche gezielter zu unterstützen, da mehr Indizien über dessen Informationsbedarf im System vorhanden sind, beispielsweise der Teil der persönlichen Dokumentsammlung, mit der sich der Benutzer gerade beschäftigt. Diese zusätzlichen Daten können auch zur Unterstützung anderer Benutzer verwendet werden; hier bieten sich kollaborative Empfehlungssysteme oder sogar einfach nur das Verfügbarmachen der von anderen Benutzern definierten Annotationen, wie eine Themenzuordnung zu Dokumenten, an.

Im Rahmen von SemIPort wurde ein persönliches Dokumentmanagementwerkzeug und ein zugehöriges Empfehlungssystem entwickelt [SJ04,Sc04], welche in Informationsportale integriert werden können. Die Kernidee dabei ist, dass die Benutzer persönliche Wissensbasen auf Grundlage einer gemeinsamen Ontologie aufbauen und diese Wissensbasen vom Empfehlungssystem gesammelt und zur Generierung von Empfehlungen genutzt werden. Die gemeinsame Ontologie ermöglicht hier eine einfache Integration der Daten aus den verschiedenen Quellen, sowie die Herstellung des Bezugs zwischen den Benutzerdaten und dem Informationsraum des Portals auf einer abstrakteren als der reinen Inhaltsebene. Für Empfehlungen kann so z.B. das Wissen genutzt werden, dass

¹ An dieser Stelle sei auf Methoden wie HITS-Algorithmus und PageRank-Algorithmus verwiesen.

² Vorwiegend Reduktion des Bedarfs an Speicherplatz, Bandbreite und Rechenzeit.

sich der Benutzer gerade mit einem bestimmten Themengebiet beschäftigt, und das eine bestimmte Menge von Dokumenten im Portal unter dieses Themengebiet fallen.

Bevor Benutzerdaten genutzt werden können, müssen diese vorhanden sein. Dazu muss der Dokumentmanager von den Benutzern akzeptiert und dessen Annotationsfunktionalität genutzt werden. Bei der Entwicklung haben wir uns deshalb zum einen an Benutzeranforderungen orientiert, die wir anhand von Interviews und Literaturrecherchen identifiziert haben, und zum anderen gerade in Bezug auf die Annotationen die Richtlinie verfolgt, dass jegliche vom Benutzer investierte Arbeit einen direkten, unmittelbaren Nutzen für diesen haben soll. Einige Merkmale des resultierenden Systems sind die Möglichkeit zur räumlichen Anordnung von Dokumenten (siehe Abbildung 1), die Suche anhand von Annotationen, die kontextspezifische Relevanzbewertung, und das Durchstöbern der Dokumentsammlung basierend auf zwischen Dokumenten definierten Beziehungen.

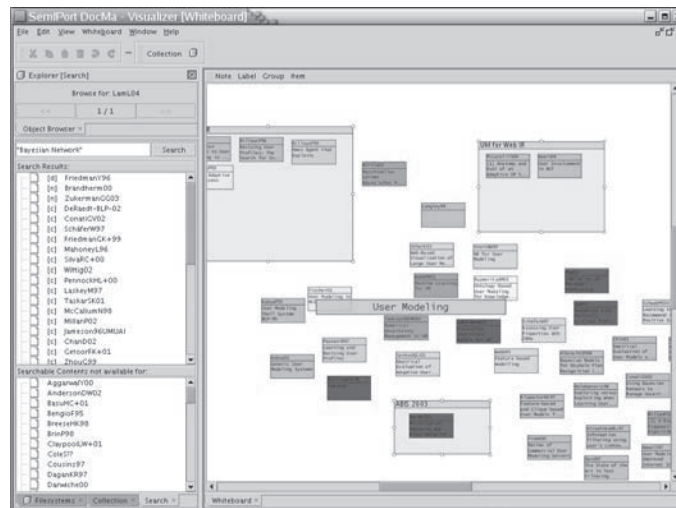


Abbildung 1: SemiPort Document Manager mit geöffnetem Whiteboard (rechts), auf dem Dokumente durch Karteikarten repräsentiert räumlich angeordnet werden können.

2.3 Finden und Browsen von bibliographischen Daten mit SWQL

Das Aufspüren von relevanten Informationen ist eine komplexe Aufgabe. Hierzu ist das Finden relevanter Publikationen wichtiger Konferenzen oder (Fach-)Journals zur Befriedigung eines Informationsbedürfnisses wichtiger denn je. Der DBLP-Bibliographie-Server mit seinen mehr als 650.000 Publikationen von über 400.000 Autoren ist ein Service, der von tausenden Wissenschaftlern auf der ganzen Welt genutzt wird und eine fundamentale Hilfestellung für das Finden von Publikationen, Konferenzen/Journals oder anderer Wissenschaftlern in ähnlichen Gebieten bietet [Ley02].