

## **Wissenserschließung in der Community**

Frank Gronostay, Kristin Albert

This document appeared in

Bernhard Mittermaier (Eds.):

eLibrary - den Wandel gestalten

5. Konferenz der Zentralbibliothek

Proceedings of the WissKom 2010: 5. Konferenz der Zentralbibliothek, 08.-11. November 2010, Jülich

Schriften des Forschungszentrums Jülich / Reihe Bibliothek/Library, Vol. 20

Zentralbibliothek (ZB)

Forschungszentrum Jülich GmbH, Zentralbibliothek, Verlag, 2010

ISBN: 978-3-89336-668-2

## **Wissenserschließung in der Community**

Frank Gronostay, Kristin Albert

### **Zusammenfassung**

#### *Wissenserschließung in der Community*

Für wissenschaftliche Bibliotheken in Forschungsinstituten oder andere wissenschaftsgetriebene Institutionen ist es im Rahmen einer qualifizierten Informationsversorgung meist unerlässlich, eine intellektuelle Erschließung von Fachpublikationen durchzuführen. Durch das rapide Wachstum der digitalen Dokumente, steigen die Kosten für diese Erschließung.

Die Fraunhofer ESK und die KTS Informationssysteme GmbH wollen die Leser und damit die fachlichen Experten in die Verschlagwortung einbeziehen. Dafür soll eine Community aufgebaut werden, in der sich Leser ihre eigene virtuelle Bibliothek aufbauen und sich mit anderen Experten vernetzen und austauschen können. So wird durch die Sortierung und persönliche Verschlagwortung der geliehenen Bücher und gelesenen Dokumente die Gesamtbibliothek automatisch sukzessive indexiert. Ausgangspunkt für eine qualitativ hochwertige Erschließung ist ein voll integriertes Ontologie-Management-System. Anders, als beim social tagging, das weitgehend frei und unkontrolliert stattfindet, soll in der neuen Wissensplattform anhand der verfügbaren Schlagworte nach bestimmten Regeln indexiert werden. Das System unterstützt den Anwender hierbei, indem es bestimmte semantische Zusammenhänge aufzeigt, ferner dem Anwender zeigt, wie andere User ähnliche Dokumente verschlagwortet haben und nicht zuletzt selber Vorschläge unterbreitet. Neben der Verschlagwortung ist ebenso eine Kommentierung und Bewertung von Fachpublikationen möglich.

Gelesen wird aber in den seltensten Fällen direkt neben einem PC. Mobiltelefone bieten sich als Alternative an. Sie sind ständige verfügbare Begleiter und werden immer mehr zum allgegenwärtigen Assistenten. Da Mobiltelefone in der Regel kleine Displays und Tastaturen haben, sind ausgeklügelte Benutzeroberflächen für den Erfolg des mobilen Assistenten entscheidend. Zum einen können über die Handy-Kamera Bücher anhand ihres Einbands erkannt werden. Zum anderen eröffnet ein integrierter Mini-Projektor, wie er schon in einige Handys der neuesten Generation integriert ist, völlig neue Möglichkeiten bei der Bedienung mobiler Anwendungen. Die Informationen werden auf das Buch projiziert und der Benutzer kann sie mit seiner Hand direkt bearbeiten. Die Kamera erfasst den Finger, so dass Aktionen wie z.B.

das Hinzufügen eines Schlagworts nun direkt auf dem Buch ausgeführt werden können.

Es wird eine Plattform für die Wissenserschließung geschaffen, auf die die Benutzer immer dann Zugriff haben, wenn sie mit den Wissensdokumenten arbeiten - sei es im Büro, zu Hause, im Zug oder am See. Durch die Zusammenarbeit in einer Expertengruppe werden die Dokumente besser erschlossen als dies bisher möglich war.

### **Abstract**

Subject indexing with a community platform

For scientific libraries at research institutes or other knowledge driven institutes which are managing information it is necessary to index scientific publications intellectually. Because of the rapid increase of digital filed documents the costs of subject indexing are increasing.

The institute "Fraunhofer ESK" and company "KTS Informationssysteme GmbH" would like to include the readers and the scientific experts when it comes to "indexing" or identifying a document in terms of its subject content.

To facilitate this process it is planned to built a community platform which allows the readers to create their own virtual library and also to network and exchange together with experts.

This means that through each personal management of documents read or books lent the entire library will be automatically little by little indexed.

The starting point to assure a highly qualified indexing process is a completely integrated "ontology-management system". Different from "social tagging" which happens without any order or control the new community-driven knowledge management platform will be organized through a controlled vocabulary and rules of indexing which will guide the user through the indexing process.

The system will support the user by providing the specific semantic context. As well it will inform you about how other users indexed similar documents and even suggests possible taggs. Additionally it is also possible to add comments or give grades about scientific publications.

But usually we do not read right next to our computers. The new alternative here is the cell phone. We tend to have our cell phone always with us which became a constant assistant. Most of the cell phones still have small displays and keyboards. But there exist already new user interfaces on the market which offer many more practical and smart features.

As an example there are cell phone cameras which can recognize book titles. Most recent models have integrated mini-projectors which offer numerous new appliances. Information can be projected into the book and the user can use and work with it by just using his fingers. The camera captures the finger and in this way for example keywords can be easily added directly in context with the book.

A new type of knowledge management platform was invented which allows users to work with documents and organize information - no matter if they are at the office, at home, in the train or at a lake. As well due to the close cooperation with a group of experts the information will be better accessible and can be managed and organized in a smarter way than before.

### **1. Problemfeld**

Wissenschaftliche Information begegnet uns heute in zwei unterschiedlichen Ausprägungen.

Entweder als tiefererschlossene Fachinformation, die oftmals in Verbindung mit einem Thesaurus verschlagwortet ist und eventuell auch mit Abstracts versehen. Das Retrieval beschränkt sich meist auf eine oder wenige Datenbasen und liefert ein überschaubares und vollständiges Ergebnis mit durchwegs relevanten Fundstellen.

Im anderen Falle treffen wir auf Informationsangebote, die rein volltextbasiert sind. Hier werden möglichst viele Quellen in die Suche eingebunden, möglichst viele Treffer werden geliefert, ein intelligenter Algorithmus nimmt ein Ranking der Treffer vor und der Suchende vertraut darauf, dass die ersten 10 oder 20 Fundstellen wohl die relevantesten sein werden.

Schaut man sich heute in der Forschung um, findet man mittlerweile häufiger die volltextbasierten Dienste, gleichwohl – wenn man die Beteiligten befragt – überwiegend die Meinung vertreten wird, das wohl die erste Ausprägung eines Informationsdienstes, die qualitativ bessere ist.

Das Problem: Angesichts der Masse der Publikationen und der zunehmenden Kurzlebigkeit der Information wird die Tiefererschließung durch berufsmäßige Dokumentare und Bibliothekare immer schwieriger.

## **2. Das OWC-Projekt der Fraunhofer ESK und der KTS**

Die Fraunhofer ESK / München und die KTS informations-Systeme GmbH / München haben in einem gemeinschaftlichen Entwicklungsprojekt ein wissenschaftliches Community-System entwickelt, das zum Ziel hat, Fachliteratur und Informationen durch den Fachleser – also z.B. den wissenschaftlichen Mitarbeiter am Institut – verschlagworten und kommentieren zu lassen.

Folgende Überlegungen wurden angestellt:

- Vorausgesetzt, dass in der Community überwiegend Mitglieder vorhanden sind, die auf Ihrem jeweiligen Fachgebiet Experte sind, kann man davon ausgehen, dass sie nach dem Lesen einer Fachpublikation momentan am besten Auskunft über die Publikation geben können.
- Da aber ein wissenschaftlicher Mitarbeiter – anders als der professionelle Dokumentar – seine Bücher nicht am Schreibtisch sitzend vor dem Computer liest, sondern vielleicht im Zug, Zuhause oder im Urlaub, müssen wir über ein ubiquitär verfügbares, webbasiertes System mit ihm kommunizieren. Das ist in den Zeiten des mobilen Internets und der Netbooks und Smartphones kein Problem mehr.
- Der Fachleser kann nun unterschiedliche Eingaben zu der Publikation machen: Entweder eine Bewertung mit einem Kommentar, oder eine inhaltliche Zusammenfassung oder eine Verschlagwortung. Bei der Verschlagwortung kann er jedoch nicht frei taggen, sondern soll sich des vorhandenen Thesaurus oder einer Ontologie bedienen.

Daher leitet sich der Projekt-Name „OWC“ ab. OWC steht für „Ontologie gestützte Wissenserschließung durch ubiquitäre Communities“.

Ziel des Projektes ist es zu untersuchen, inwieweit solch eine Fachcommunity in der Lage ist, die professionelle Erschließungsarbeit des Dokumentars sinnvoll zu ergänzen oder gar zu ersetzen. Ist es möglich, durch ein solches System ohne großen Arbeitsaufwand zu einer tiefenerschlossenen Datenbasis zu kommen?

## **3. Vorgehensweise:**

Als kritische Faktoren für den Erfolg der Anwendung wurden folgende Punkte identifiziert:

- Die Motivation der Fachleser, mitzumachen
- Die Qualität der Eingaben
- Die Unterstützung des Fachlesers durch intuitive mobile Technologien

Zur Motivation der Fachleser:

Um in einer ausgewachsenen Bibliothek den Katalog mit benutzergenerierten Inhalten zu befüllen, bedarf es natürlich einer kritischen Masse an Usern, die sich an der Community beteiligen.

Ein Blick in die aktuelle Internet-Landschaft zeigt, dass sich kollaborative Systeme und Communities immer größerer Beliebtheit erfreuen. Das Parade-Beispiel ist sicherlich die Online-Enzyklopädie Wikipedia, die am 27. Dezember 2009 ihren einmillionsten deutschen Artikel feiern konnte und heute ca. 190 Millionen Besucher pro Tag hat.

Ein gewisse Grundmotivation, sich in eine Community einzubringen, kann man also voraussetzen.

Darüber hinaus wurden in unserem Projekt zwei grundlegende Konzepte verwirklicht, die die Motivation zur Teilnahme an der Community erhöhen sollen.

- Das Community-Mitglied hat einen persönlichen, praktischen Nutzen.

Die Anwendung wurde gezielt so gestaltet, dass der User eine persönliche Literatur- und Referenz-Verwaltung damit betreiben kann und damit auch einen praktischen Nutzen aus dem System zieht.

- Das Community-Mitglied erlangt durch seine Mitarbeit eine Reputation in der Fachcommunity.

Hierzu hat das Fraunhofer ESK ein kontextfähiges Reputations-System entwickelt, das es erlaubt zu bestimmten Themen nicht nur Fachliteratur zu finden, sondern auch Experten auszuweisen.

Zur Qualität der Eingaben:

Wenn ein Fachleser freiwillig eine Bewertung zu einer Fachpublikation abgibt oder den Inhalt der Publikation in kurzen Worten wiedergibt, kann man in der Regel davon ausgehen, dass die Eingabe des Users eine akzeptable Qualität aufweisen wird.

Schwieriger ist es bei der Verschlagwortung: Der professionelle Dokumentar erschließt Literatur anhand eines Thesaurus, mit dem er bestens vertraut ist, und anhand bestimmter Regeln, die ihm nach jahrelanger Arbeit in Fleisch und Blut übergegangen sind.

Der Fachleser ist zwar in seinem jeweiligen Fachgebiet Experte und kann daher das Gelesene gut beschreiben, er kennt jedoch das kontrollierte Vokabular nicht so gut und kennt auch keine Regeln zur Verschlagwortung.

Es sind daher in diesem Projekt zwei Methoden implementiert worden, um den Fachleser bei der Verschlagwortung zu unterstützen:

1. Vorschläge für Schlagworte:

Aufgrund statistischer Vergleiche mit ähnlichen Dokumenten wird dem Fachleser der Teil des kontrollierten Vokabulars präsentiert, der in diesem Fall relevant erscheint. Ausgehend von diesen Vorschlägen kann er blättern und stöbern, Unter- und Oberbegriffe, Hintergrundinformationen und Begriffe im Kontext ansehen und sich dann für die treffendsten Begriffe entscheiden.

2. Assistenz zur Verschlagwortung:

Anhand eines Schritt-für-Schritt Bildschirmdialoges werden bestimmte inhaltliche Aspekte der Publikation abgefragt, um vom Leser eine möglichst vollständige und qualifizierte Verschlagwortung zu erhalten.

Dass nun trotz aller Assistenten ganz unterschiedliche Ansichten über eine Publikation entstehen können, macht ein Community-System gerade erst interessant. Sollten aber schlichtweg falsche oder unpassende Eingaben entstehen, können diese durch andere Community-Mitglieder reklamiert werden.

Wie bereits erwähnt, haben Community-Mitglieder die Möglichkeit, sich durch ihre Mithilfe in der Community eine Reputation zu erarbeiten. Diese Reputation gibt an, wie qualitativ ein Community-Mitglied Dokumente verschlagwortet und lässt außerdem Schlüsse darüber ziehen, wer Experte in welchen Themengebiet ist.

Die Ziele des Reputationsmodells sind:

- Möglichkeit zur Qualitätsbewertung
- Kontextfähigkeit

Zur Qualitätsbewertung:

Jede Zuordnung eines Schlagwortes zu einem Dokument soll hinsichtlich ihrer Qualität bewertet werden können. Dazu wird unter anderem die Reputation des Benutzers betrachtet.

Zur Kontextfähigkeit:

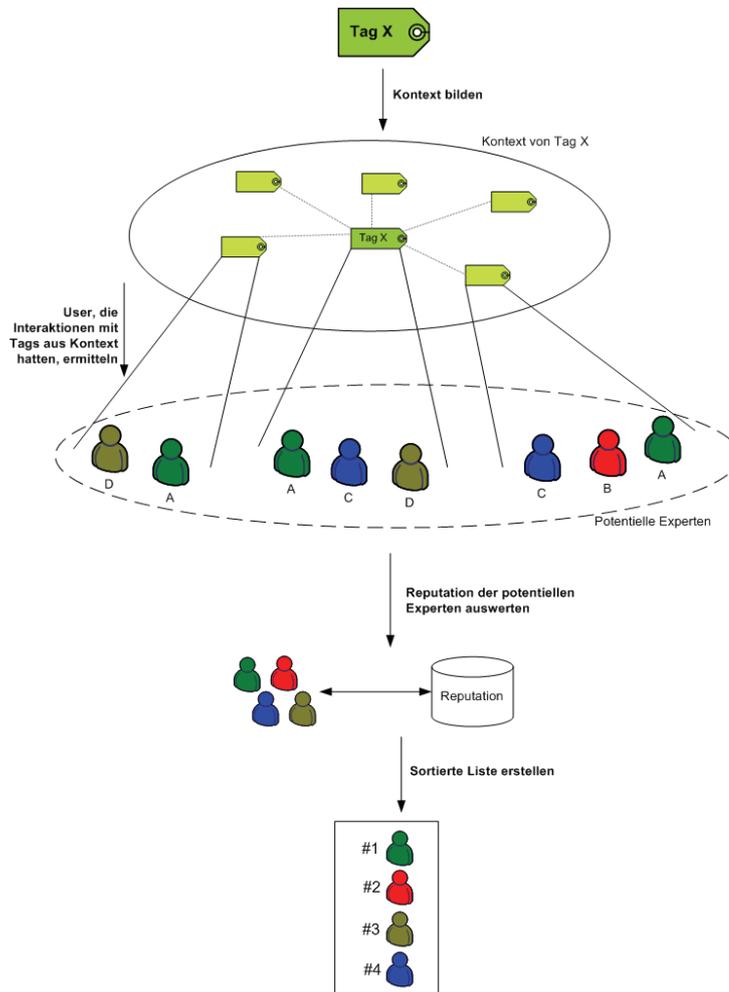
Die Reputation eines Benutzers lässt sich nicht durch einen einzelnen Zahlenwert ausdrücken. Es muss berücksichtigt werden, dass Dokumente und Schlagwörter aus völlig unterschiedlichen Fachgebieten in der Datenbasis vorhanden sind. Ein Benutzer, dessen Aktionen auf einem bestimmten Gebiet eine hohe Qualität aufweisen, darf nicht automatisch auf anderen Gebieten als ebenso kompetent gelten. Deshalb sind jedem Benutzer mehrere Reputationswerte zugeordnet, die jeweils für einen bestimmten Kontext (d.h. ein Fachgebiet) stehen.

Aus den oben genannten Zielen des Reputationsmodells ergeben sich folgende zwei Interaktionen für den Benutzer mit dem System:

- Ein Dokument verschlagworten, d.h. ein Dokument wird mit einem Schlagwort in Beziehung gesetzt, mit dem es vorher in keiner Beziehung stand
- Eine Verschlagwortung bewerten, d.h. ein bereits vorhandene Beziehung zwischen einem Dokument und einem Schlagwort wird bewertet

Ein zusätzliches Ergebnis, das sich bei der Auswertung der gesammelten Daten ergibt, ist die Expertensuche. Sie stellt die Funktionalität bereit, um innerhalb eines bestimmten Kontexts nach Experten zu suchen.

Als Eingabe wird ein Schlagwort an das System übergeben. Das System ermittelt dann zu diesem Schlagwort, unter Verwendung des Kontextmodells, den zugehörigen Kontext. Wie in Abbildung 1 ersichtlich bilden die Schlagwörter innerhalb des Kreises mit der durchgezogenen Linie den Kontext von Tag X. Der Kontext beschreibt hier eine Menge von Schlagwörtern, die eine gewisse thematische Nähe zueinander besitzen. Jeder Benutzer, der bereits ein Schlagwort dieses Kontextes einem beliebigen Dokument zugeordnet hat, gehört zur Gruppe der potentiellen Experten. In Abbildung 1 befinden sich die potentiellen Experten im Kreis mit der gestrichelten Linie. Für jeden Benutzer dieser Gruppe wird nun die Reputation in diesem Kontext ermittelt. Die potentiellen Experten werden absteigend nach ihrer Reputation sortiert.



**Abb.1: Expertensuche**

Zur Unterstützung des Fachlesers durch intuitive mobile Technologien:

- Mobiler Assistent
- Zukunftsvision: Projizierte Benutzeroberflächen

Zum Mobilen Assistenten:

Mobile Anwendungen greifen zunehmend auf Internetdienste zu. Diese Dienste sind zum großen Teil noch nicht an mobile Geräte angepasst. Verbindungsabbrüche und schwankende Datenraten führen zu Problemen. Im Rahmen des Projektes wurde das Mobility Gateway entwickelt, das zwischen Online-Dienst und mobilen Endgerät arbeitet. Es kann Daten zwischenspeichern und Protokolle übersetzen.

Darüber hinaus kann es Profile verwalten und rechenintensive Aktionen für den Client übernehmen. Zum Beispiel bietet es die Möglichkeit zusätzliche Plugins, wie eine Bilderkennung, zu aktivieren und dem Client die Funktionalität bereitzustellen. Im OWC Anwendungsfall übersetzt der Gateway zwischen dem Web Service Protokoll SOAP und REST. Für beide Protokolle und ihre spezifische Implementierung wurde eine eigene Softwarekomponente entwickelt.

Der mobile Client wurde für das Android Betriebssystem entwickelt und umfasst die wichtigsten Funktionalitäten des Desktop-Clients. Das User Interface ist eines der entscheidenden Kriterien für mobile Anwendungen, da insbesondere die Einschränkungen der mobilen Endgeräte (Bildschirmgröße, Texteingabe) beachtet werden müssen. Aufgrund der Einschränkungen der mobilen Endgeräte und der begrenzten Zeit kann ein mobiler Client nicht alle Funktionen der Community- Plattform abbilden. Der Schwerpunkt des Clients liegt auf einer schnellen Suche nach Dokumenten. Dabei wird der Nutzer so unterstützt, dass er zum Beispiel einen schnellen Zugriff auf die zuletzt verwendeten Suchbegriffe hat oder aber durch eine Autovervollständigung bei der Eingabe unterstützt wird (wie in Abbildung 4 dargestellt).

Außerdem können mit Hilfe des mobilen Clients Dokumente in Listen verwaltet und jederzeit an jedem Ort verfügbar gemacht werden. Die Synchronisation von mobilen und webbasierten System ermöglicht dabei den problemlosen Wechsel zwischen den Geräten.



Abb. 2: Der mobile Client

Zur Zukunftsvision:

Um neue intuitive Interaktionskonzepte für den mobilen Einsatz zu demonstrieren, wurde im Rahmen des Forschungsprojektes ein Hardwareprototyp entwickelt, der einen integrierten Mini-Projektor mit einer Kamera kombiniert. Verwendet man das Gerät wie ein Taschenlampe, dann wird die projizierte Oberfläche direkt auf die Seiten eines aufgeschlagenen Buches sowie auf freie Flächen neben dem Buch dargestellt. Zur Bedienung der Anwendung verwendet der Benutzer einen Finger, dessen Position durch die Kamera erfasst wird. Aktionen wie z. B. das Hinzufügen eines Schlagworts können nun direkt auf dem Buch ausgeführt werden: Aus einer Liste, die neben dem Buch dargestellt wird, zieht man ein Schlagwort direkt in das Buch hinein. Schon ist das Buch mit diesem Schlagwort verknüpft.



**Abb. 3: Eine Zukunftsvision ist die Interaktion mit projizierten Benutzeroberflächen**

#### *4. Ergebnis des Projektes:*

Das OWC-System wurde in verschiedenen Versuchen praktisch getestet und derzeit werden die ersten kundenspezifischen Anwendungen erstellt.

Es zeigt sich, dass durch die Mitarbeit der Fachleser die Datenbasis besser erschlossen werden kann, als es bisher möglich war. Anders als im klassischen Falle, wo ein Literaturnachweis fertig erschlossen in die Datenbank gestellt wird,

erschließen sich in der Community die Dokumente erst sukzessive, je nachdem wie viel Community-Mitglieder es gibt und wie viel Dokumente gelesen werden.

Die Verschlagwortungen der Fachleser verhelfen zu einem besseren Retrieval und beschleunigen das Auffinden relevanter Dokumente.

Die Meinungen und Kommentare der Fachleser werden in der Regel von anderen Usern als sehr interessant und nützlich empfunden. Aus einer Vielzahl von Treffern werden einzelne Titel hervorgehoben, die von Wissenschaftlern mit einer ähnlichen Interessenslage bereits als sehr relevant oder weniger relevant eingestuft wurden.

Insgesamt wird das OWC-System als deutliche Verbesserung des Informationsangebotes empfunden.

Besonders hervorzuheben ist die Rolle des Mobilien Assistenten, der von der Fraunhofer ESK entwickelt wurde. Dieser trägt maßgeblich dazu bei, dass das Gesamtsystem von dem potentiellen Community-Mitglied gerne angenommen wird und birgt noch sehr viel Entwicklungspotential für viele weitere Anwendungen.