

September 2001

Eine offene, dezentrale Architektur für das Suchen und Auffinden von Informationen in verteilten Systemen

Harald Häuschen

Universität Zürich, haeusche@ifi.unizh.ch

Follow this and additional works at: <http://aisel.aisnet.org/wi2001>

Recommended Citation

Häuschen, Harald, "Eine offene, dezentrale Architektur für das Suchen und Auffinden von Informationen in verteilten Systemen" (2001). *Wirtschaftsinformatik Proceedings 2001*. 23.
<http://aisel.aisnet.org/wi2001/23>

This material is brought to you by the Wirtschaftsinformatik at AIS Electronic Library (AISeL). It has been accepted for inclusion in Wirtschaftsinformatik Proceedings 2001 by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

In: Buhl, Hans Ulrich, u.a. (Hg.) 2001. *Information Age Economy*; 5. Internationale Tagung
Wirtschaftsinformatik 2001. Heidelberg: Physica-Verlag

ISBN: 3-7908-1427-X

© Physica-Verlag Heidelberg 2001

Eine offene, dezentrale Architektur für das Suchen und Auffinden von Informationen in verteilten Systemen

Harald Häuschen

Universität Zürich

Zusammenfassung: Das Angebot an Informationen, die via Internet bezogen werden können, ist immer noch stark im Zunehmen begriffen. Täglich kommen neue Informationsquellen hinzu. Die Informationen werden in verschiedenen Formaten angeboten (HTML, XML, PDF, Katalogen, branchenspezifischen Datenbanken, u.a.). Trotz Suchmaschinen können viele Informationen nicht oder nur schwer aufgefunden werden. Zudem braucht es für den Zugriff meist unterschiedliche Programme. Dieser Artikel stellt eine verteilte Sucharchitektur „DiSA“ vor, die es erlaubt, strukturiert Informationen aus bereits bestehenden unterschiedlichen Quellen zu veröffentlichen, so dass diese rasch gefunden werden können. Der Lösungsansatz besteht im wesentlichen darin, dass die weltweit verstreuten Informationsquellen mittels eines einfachen standardisierten Interfaces in eine Such-Architektur eingebunden werden. Die Lösung ist dezentral und weitgehend sprachneutral und ermöglicht sowohl standardisierte als auch individuelle Informations-, Produkt- und Dienstleistungsangebote. Die Architektur ist so einfach, dass bestehende Informationsquellen rasch eingebunden werden können.

Schlüsselworte: Information Retrieval, Informationssuche, Architektur

1 Einführung

Das Angebot an Services, Dienstleistungen und Informationen, die via Internet bezogen werden können, ist immer noch stark im Zunehmen begriffen. Täglich kommen neue Server, Datenbanken und Anwendungen hinzu. Eine ungeheure Vielfalt an Informationen ist vorhanden. Doch viele Informationen werden nicht aufgefunden und viele Angebote bleiben ungenutzt. Die Gründe hierfür sind rasch aufgezählt:

- Trotz des grossen Angebots an Suchmaschinen hat niemand mehr eine Übersicht, was wo und in welcher Form angeboten wird.
- Die Informationen werden in verschiedenen Formaten abgespeichert (HTML, XML, PDF, Katalogen, branchenspezifischen Datenbanken, u.a.).

- Für den Zugriff braucht es meist unterschiedliche Programme.
- Man kann stunden- und tagelang erfolglos suchen, obwohl die Anfrage einfach erscheint und viele scheinbar passende Seiten gefunden werden.
- Trotz Standards wie beispielsweise HTML und XML sind die Informationsseiten uneinheitlich.
- Die Informationen werden oft nur in einer Sprache angeboten.
- Eine Zusammenarbeit der Anbietern unterschiedlichen Plattformen gibt es praktisch nicht.
- Eine standardisierte Automatisierung von unternehmensübergreifenden Prozessen (z.B.: suchen, evaluieren, vergleichen u.a.) ist nur mit grosser Mühe realisierbar.
- Anbieter von Informationen verwenden mehr Zeit in die Gestaltung der Navigation und Angebote (multimediale Präsentation, Werbung etc.) als in die gezielte Strukturierung ihrer Information [ScLe99].

Die obigen Aussagen gelten nicht nur für das Internet, sondern auch häufig für das Intranet. Auch hier fallen grosse Informationsmengen an und die Benutzer klagen über fehlende Übersicht und fehlende Transparenz beim Zugriff auf die unterschiedlichen Informationsquellen.

Die Suche und die Organisation der Information ist bereits seit langem ein Thema. Die Forschung ist breit und die Ergebnisse dieser Forschung fliessen in verschiedene Anwendungen und Systeme ein. Die folgende Aufzählung dient als Beispiel und ist nicht vollständig.

- Im Internet sind vor allem die Suchmaschinen sehr beliebt. Diese verwalten in einer oder mehreren Datenbanken Millionen von Suchbegriffen und die dazugehörigen Verweise auf Internet-Seiten. Die Suche ist häufig eine einfache Volltextsuche und unterstützt keine strukturierte Verfeinerung (z.B.: www.google.de).
- Metasuchmaschinen beinhalten häufig keine eigene Datenbank, sondern suchen das Gewünschte auf den verschiedenen Suchmaschinen und fassen dann die Ergebnisse zusammen (z.B.: www.virtualfreesites.com)
- Im Business-to-Business Bereich haben sich vor allem Kataloge und branchenspezifische Datenbanken durchgesetzt. Diese haben den Vorteil, dass die Informationen strukturiert abgelegt und gesucht werden können. [HaSc97].
- In verschiedenen Branchen schliessen sich die unterschiedlichen Unternehmen zusammen und einigen sich auf einen Standard wie Informationen abgerufen werden können. Sogenannte „Information Resource Locator“ helfen dann die gewünschte Information zu finden (z.B.: www.nemonline.org/mirl).

- Such-Agenten sind autonome und individuelle Softwarekomponenten, die in der Lage sind, auf Wunsch des Benutzers automatisiert und benutzerorientiert Informationen im Internet zu suchen und damit Aktionen auszuführen, beispielsweise Produkte zu suchen und Verkaufsabschlüsse zu tätigen (z.B.: www.cdrom-guide.com/bargainfinder.htm).
- In den Umgebungen, bei denen besonders viele Informationen anfallen und verarbeitet werden (z.B.: Astronomie, Physik), wird innerhalb des internationalen Projektes „Grid“ eine spezifische Architektur ausgearbeitet, um die Informationsflut zu bewältigen [Moor99].

Jedes dieser Systeme besitzt Vor- und Nachteile. Allen gemeinsam ist, dass jeweils nur ein geringer Prozentsatz, der zur Verfügung stehenden Information, gefunden werden kann und dass das Gesuchte oft nur ungenügend präzisiert werden kann, da zu wenige Strukturierungskriterien zur Verfügung stehen [StHo98]. Es steht dem Anwender also eine breite Palette von Werkzeugen, Anwendungen und Systemen zur Verfügung, doch der Suchende kann diese nicht einheitlich, auf einfache Art und Weise gleichzeitig befragen.

Bisher bestehen keine befriedigenden Lösungen zur Verfügung und es sind auch keine in Sicht und dies obwohl das Suchen und Auffinden zentrale Funktionen für die Information Society sind, die Suchkosten und Qualität der Suchergebnisse immer relevanter werden und das Suchen zur Zeit einen hohen Interaktionsaufwand benötigt. Viele Vorschläge von Gruppen (z.B. ACM SIGIR) und Konferenzen (z.B.: Hypertext – Information Retrieval [FuDi97]), die sich mit Information Retrieval beschäftigen, gehen deshalb häufig dahin, dass man versucht eine gemeinsame Datenorganisation zu finden und zu definieren. Die geschieht auch unternehmensintern. Im Unternehmen werden hierfür Data Warehouses bereitgestellt, in denen alle wichtigen Daten, die im Unternehmen anfallen, zu finden sind. Doch dies ist nicht anzustreben, da eine Standardisierung schwierig und die Transformation aller Informationsquellen in neue Formate aufwendig und nicht immer wirtschaftlich ist.

Dieses Papier schlägt deshalb einen anderen Ansatz vor. Die Informationsquellen können so belassen werden wie sie sind. Lediglich die Suchmöglichkeiten und Strukturierung der Information wird öffentlich zugänglich gemacht. Dies geschieht mittels einer einfachen Architektur bestehend aus einer Middleware, wenigen Schnittstellen-APIs und gemeinsamen Suchservices. Es werden hierfür keine neuen Suchtechniken für Datenbanken, keine neuen Standards zur Speicherung von Informationen, kein neues Datenmanagement und auch kein Ersatz für bestehende Informationsformate vorgeschlagen.

2 Anforderungen

Die allgemeinen Anforderungen an einen Suchdienst sind offensichtlich. Informationen (z.B.: zu Produkten und Dienstleistungen) müssen rasch, weltweit oder regional und auf das eigene Bedürfnis abgestimmt gefunden werden können. Hierbei möchte der Benutzer wenige Ergebnisse, jedoch exakt abgestimmt auf seine Abfrage, erhalten. Dies sei an den folgenden Beispielen illustriert.

- Beispiel 1: Es gibt Produkte (z.B.: CD, Buch), die man weltweit suchen und von überall beziehen kann. Das Produkt ist häufig klar spezifiziert (es gibt nur eine Ausgabe, eine Qualität und nur einen Produzenten). Es würde also genügen, wenn der Benutzer seinen Wunsch formuliert und dann entsprechende Preis/Quelle/Lieferangebote auf seinem Desktop angezeigt werden.
- Beispiel 2: Sucht man ein klassisches Stück auf einer beliebigen CD, so ist die Anfrage bereits nicht mehr so klar definiert und der Suchende muss differenzierte Suchbegriffe (z.B.: Aufnahme, Dirigent, Solist und Orchester) eingeben können und erwartet eine detaillierte Übersicht. In diesem Beispiel muss der Benutzer bereits mehr Informationen eingeben können und erhalten, wie z.B. weitere Stücke auf der CD, Hörproben etc. Nützlich könnte es auch sein, ihm anzuzeigen, dass ein neues Produkt in Vorbereitung ist.
- Beispiel 3: Es gibt Produkte (z.B.: Hotelangebot), die man weltweit suchen, aber nur vor Ort nutzen kann. Der Preis und der Qualitätsvergleich sind auf die Region beschränkt. Zudem spielt hierbei die Verfügbarkeit eines Angebotes zu einem bestimmten Datum und der Abdeckungsgrad für ein individuelles Bedürfnis eine entscheidende Rolle.

Wie diese einfachen Beispiele zeigen, kann eine Anfrage aus der Sicht des Benutzers einfach und eindeutig sein, die Antwort sollte idealerweise den Kontext der Anfrage berücksichtigen. So kann es auch nützlich sein, über den Suchenden Informationen zu besitzen (Vorlieben, Ort etc.), um auf die Anfrage angepasst zu antworten.

Ein neues Konzept kann deshalb nur erfolgreich sein, wenn folgende Kriterien aus der Sicht der Benutzer beachtet werden:

- Einfachheit: Die Bedienung muss intuitiv und dem Benutzer adäquat sein
- Kultur- / Sprachunabhängig: Ob der Benutzer in englisch, deutsch oder in französisch sucht, die Sprache darf auf das Resultat keinen entscheidenden Einfluss haben, ausser der Suchende wünscht es explizit.
- Multimedial: Der individuellen Aufbereitung und der Präsentation der Informationen kommt ein wesentlicher Stellenwert zu.
- Personalisierbar: Der Benutzer muss seine eigenen Profile, die seine Wünsche und Bedingungen enthalten, erstellen und verwalten können.

- Anonym: Der Suchende sollte anonym bleiben können.
- Wiederverwendbar: Suchfilter und -ergebnisse müssen wiederverwendbar sein.
- Sicher: Die Verfügbarkeit, Integrität und Vertraulichkeit müssen gewährleistet sein.
- Automatisierbar: Suchprozesse müssen automatisiert werden können.

Von Seiten der Informationsanbieter ist folgendes zu beachten:

- Integration: Es muss ohne grossen Einsatz gelingen, bestehende Systeme einzubinden.
- Standard: Bestehende Standards sollten integriert und genutzt werden (HTML, SQL, XML).
- Skalierbarkeit: Informationen und Prozesse müssen beliebig verteilt werden können.
- Synchronität: Es müssen sowohl asynchrone als auch synchrone Anfragen bearbeitet werden können.
- Zugriff: Es müssen Pull und Push Prinzipien Beachtung finden.
- Flexibilität: Es müssen Versionen von Informationsobjekten verwaltet werden können.

Die vorgeschlagene Architektur versucht diese Anforderungen zu erfüllen. Im Zentrum steht nicht, möglichst viele Treffer und Informationen anzuzeigen, sondern nur jene, die der Benutzer auch tatsächlich sucht. Im Zentrum steht also der individuelle Benutzer, der bestimmte Vorlieben und Verhaltensweisen besitzt, die bei jeder Suche mitberücksichtigt werden. Die Daten zum Benutzer werden nur vom diesem verwaltet und können auf Wunsch weiteren Beteiligten zur Verfügung gestellt werden.

3 DiSA Architektur

Die „DiSA“ (Distributed Search Architecture) definiert eine Architektur, die dem Suchen und Auffinden von Informationen dient (Abb. 1). Die Architektur umfasst:

- Anwendungen und Services, die das Organisieren und Suchen erleichtern,
- Basisservices und Bibliotheken für Netzwerke und Schnittstellen und
- eine Middleware, um Services unabhängig von Plattformen miteinander zu verbinden.

Die Architektur ist so gestaltet, dass sie offen für beliebige Erweiterungen ist.

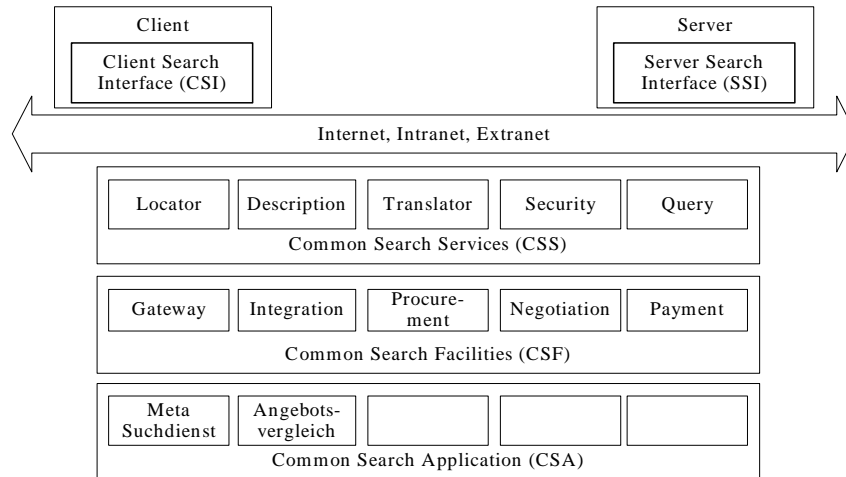


Abbildung 1: DiSA Architektur

- **CSS :** Die Basis der Architektur bilden die CSS-Services. Sie dienen dazu, Informationen, Produkte und Dienstleistungen zu beschreiben, Informationen zu transformieren, Sicherheitsanforderungen zu realisieren, Informationen zu lokalisieren und komplexe Abfragen zu definieren.
- **CSF :** Die Facilities stellen vor allem Services zur Verfügung, die im e/mBusiness Verwendung finden. Hierzu zählen Gateways zu anderen Suchsystemen, Integrationsdienste, Vermittlungsdienste und Bezahlendienste.
- **CSA :** Aufbauend auf den Basisservices und Facilities können typische Anwendungen definiert werden.
- **SSI:** Damit die bestehenden Informationsquellen eingebunden werden und diese die angebotenen Services nützen, aber auch zur Verfügung stellen können, wird eine Standardschnittstelle eingesetzt. Diese stellt alle Funktionalitäten zur Verfügung, um die Dienste CSS, CSF und CSA zu nutzen.
- **CSI:** Dem Benutzer wird ebenfalls ein Standardinterface zur Verfügung gestellt. Es besteht aus einem Java-Applet. Dieses Java-Applet kann interaktiv benutzt oder mittels Programmierung genutzt und somit erweitert werden.

In einer ersten Phase werden keine zusätzlichen Server oder zentralen Dienste benötigt. Lediglich die Server, die bereits heute Informationen bereitstellen, müssen drei Grunddienste implementieren: Registration, Lokalisation und Zugriff von Informationen (siehe weiter unten). Jeder kann einen oder mehrere Server betreiben. Das System ist vollständig dezentral und benötigt keine zentralen Instanzen.

4 Informationsobjekt

Im Zentrum der Architektur steht das Informationsobjekt. Es beschreibt eine Klasse von Informationen mit denselben Eigenschaften (z.B.: Die Klasse der Hotels mit den Attributen Anzahl Zimmer, Kategorie u.a.). Der Benutzer kann die Informationsobjekte von einem Registrations-Server abrufen und damit Suchanfragen formulieren (Beispiel siehe Abbildung 2).

```

<Information-Object >
  <Name>CD</Name>
  <Version>1.0</Version>
  <Create> <Name>Harald Häuschen</Name>
           <Mail>haeuschen@haeuschen.ch</Mail></Create>
  <Standard>
    <Attribute> <Name>Gruppe</Name>
                <Typ>string</Typ><Size>50</Size></Attribute>
    <Attribute><Name>Titel</Name>
                <Typ>string</Typ></Attribute>
    <Attribute ><Name>Preis</Name>
                <Typ>float</Typ></Attribute></Standard>
  <Dictionary>
    <us>"Gruppe" "Group";"Titel" "Title";"Preis" "Price"</us>
    <uk><like>us</like></uk></Dictionary>
</Information-Object>

```

Abbildung 2: Beispiel einer Definition eines Informationsobjektes, das für eine einfache CD-Suchanfrage verwendet werden kann

Das Informationsobjekt wird mittels XML vollständig beschrieben. Die Definition eines Informationsobjektes ist vielfältig und lässt viel Gestaltungsfreiraum. Folgende ist bereits vorgesehen: Attribute, die jedes Informationsobjekt dieses Typs besitzt; Informationsobjekte können aus anderen Informationsobjekten bestehen; Versionen von Informationsobjekten; XSLT Regeln wie ein Informationsobjekt zur nächst höheren Version transformiert werden kann; Hinweise zur Übersetzung von Attributen und Informationen; Beschreibung des Layouts; Informationen zum Ersteller.

5 Registrieren

Bevor ein Informationsobjekt vom Benutzer verwendet werden kann, muss dieses bei einem oder mehreren Servern, die eine Registration anbieten, registriert werden (vergl. Abbildung 3).

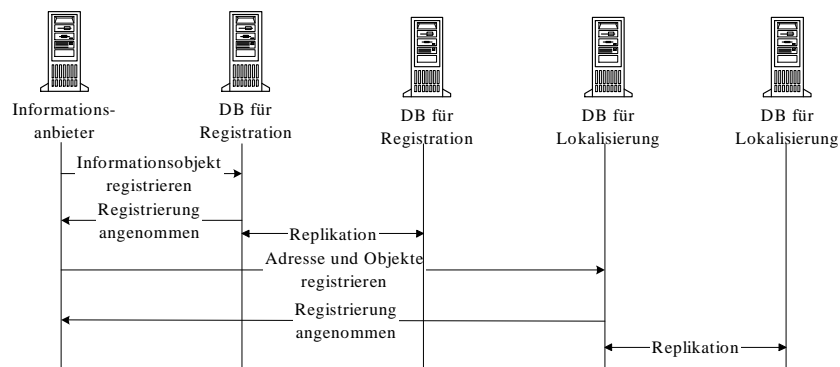


Abbildung 3: Prozess der Registrierung und Lokalisierung aus der Sicht des Informationsanbieters

Diese Registrierung geschieht mit der Übertragung der XML-Definition. Sofern das Objekt mit dieser Versionsnummer nicht existiert, trägt der Server das Objekt in seiner Datenbank ein und antwortet mit einer Bestätigung. Existiert bereits ein solches Objekt mit derselben Versionsnummer, so wird die Definition dieses Objektes als Antwort zurückgegeben und es wird kein Eintrag gemacht. Der Anbieter hat nun die Möglichkeit, eine andere Versionsnummer zu vergeben und gleichzeitig zu beschreiben, wie die bereits bestehenden gleichen Objekte in sein Objekt transformiert werden können. So ist sichergestellt, dass Veränderungen und unterschiedliche Definitionen jederzeit möglich sind und der Suchende dennoch nur ein Informationsobjekt ausfüllen muss, um die gewünschte Information zu finden. Die Server, die eine Registration anbieten, verteilen die registrierten Informationsobjekte auf weitere Server, die ebenfalls eine Registration anbieten und als solche bekannt sind.

6 Lokalisieren

Zu jedem Informationsobjekt muss zum einen festgehalten werden, wo Informationsquellen zu finden sind, die Informationen dazu enthalten, und zum anderen müssen die Informationsanbieter angeben, welche Informationsobjekte sie unterstützen. Hierzu wird ein Server benötigt, der einen Lokalisierungsdienst anbietet.

Der Informationsanbieter schickt seine Adresse und den Namen des Informationsobjekts an den Server. Der Server registriert dies und sendet die Daten an weitere ihm bekannte Lokalisierungsdienste.

7 Auffinden

Möchte nun der Benutzer eine bestimmte Information im Netz finden, so sucht er zuerst ein passendes Informationsobjekt (vergl. Abbildung 4).

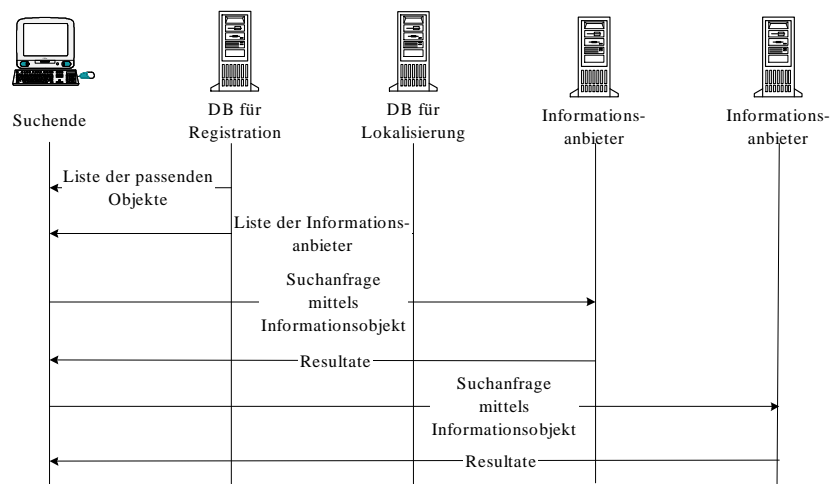


Abbildung 4: Prozess des Auffindens von Informationen aus der Sicht des Suchenden

Anschließend füllt er die Felder mit den gewünschten Werten aus. Nun wird der Lokalisierungsdienst angefragt, welche Server das gewünschte Informationsobjekt unterstützen. Die genannten Server erhalten dann die Anfrage. Hierbei kann der Benutzer selber bestimmen, welche Server von der Anfrage ein- bzw. ausgeschlossen werden. Die Anfragen können parallel verschickt und bearbeitet werden. Alle Antworten aller Server werden in einer Gesamtliste zusammengefasst. Die Antwort eines Datenservers kann beliebig viele Daten enthalten, die zu dieser Anfrage passen. Die Antwort kann hierbei dynamisch mit zusätzlichen Eigenschaften erweitert werden. So ist es dem Anbieter jederzeit möglich, mehr Informationen zu liefern als gewünscht.

8 Auswerten

Es ist offensichtlich, dass man die Anbieter nicht zwingen kann, nur die Daten zu liefern, die angefragt werden. Zudem ist auch der Kontext der Frage nicht immer bekannt. Deshalb wird die Auswertung der Antworten immer auf dem Client ausgeführt. Das Clientinterface filtert zuerst einmal die gesuchten Informationen gemäss der Anfrage und berücksichtigt hierbei das persönliche Profil des Benutzers. Der Benutzer kann nun alle Antworten beliebig filtern, sortieren und miteinander vergleichen, ohne dass er eine weitere Kommunikation mit den Servern benötigt. Erst wenn er seine Anfrage verändert, muss die Anfrage erneut verschickt werden.

9 Profil

Das Clientinterface interpretiert die Eingaben, die der Benutzer macht. Hierzu gehören beispielsweise Server, die bevorzugt angefragt werden, Filterkriterien, Sortierung u.a. Diese Angaben werden lokal abgespeichert. Diese abgespeicherten Zusatzinformationen spielen eine wichtige Rolle. Sie werden als Profil des Benutzers bezeichnet. Der Benutzer kann sein Profil einsehen und beliebig verändern. Er kann aber auch Teile seines Profils weitergeben, damit der Anbieter dieses auswerten und ein individuelles Angebot machen kann.

9.1 Wiederverwendung

Alle Anfragen können beim Client gespeichert werden. So ist es möglich, dass der Benutzer nicht jedes Mal erneut seine Anfrage formulieren muss bzw. dass er eine bestehende Abfrage in sehr kurzer Zeit abändern kann. Mit Hilfe seines Profils kommt er nun rascher zu brauchbaren Ergebnissen. Es können auch einfache zeit- oder ereignisgesteuert Anfragen realisiert werden.

9.2 Abonnieren

Mit Hilfe der Profile und den wiederverwendbaren Informationsobjekten kann nun ein einfacher Push-Dienst realisiert werden. Für den Benutzer wird es so ermöglicht, seine Angaben bei einem Anbieter zu hinterlegen und damit Informationen regelmässig via Push zur Verfügung gestellt zu bekommen. Z.B.: Der Benutzer sucht eine Information über wissenschaftliche Literatur im Bereich Robotik. Dabei bevorzugt er Ausgaben in deutsch und solche, die autonome Roboter abhandeln. Da solche Artikel unregelmässig erscheinen, schickt er sein Profil und das Informationsobjekt an die Anbieter. Diese können nun die Anfrage individuell zugeschnitten, in beliebigen Zeitabständen dem Anfragenden zukommen lassen.

10 Agenten

Bisher wurde davon ausgegangen, dass der Benutzer immer online ist und alle Informationen erhält und selber auswertet. Dies ist nicht immer der Fall. So wäre es wünschenswert, dass man auch in mobilen Situationen, in denen man nicht immer online sein kann, Informationen suchen kann. Z.B.: Ein Aussendienstmitarbeiter fliegt von Zürich nach Denver. Vor dem Abflug startet er eine Abfrage für ein Hotel. Nach der Landung möchte es das Resultat seiner Suche erhalten. Dies kann auf zwei Arten realisiert werden. Entweder kann er seine Anfrage einem Broker überlassen, der die Suche für ihn übernimmt (siehe weiter unten) oder er benutzt hierfür einen Software-Agenten. Ein Agent ist eine Softwarekomponente, die selbstständig im Auftrag des Erstellers einen Dienst ausführt [NwNd98]. Der Agent benötigt das Informationsobjekt und gewisse Profildetails des Anwenders. Danach ist er in der Lage, im Netz zu migrieren und die Informationen zusammenzutragen. Die Agenten können so gestaltet werden, dass komplexe Abfragen effizient abgewickelt werden können. Damit dies möglich wird, müssen einige Anbieter oder Broker eine Agentenplattform anbieten.

11 Informationsbroker

Wie bereits oben erwähnt, ist der Benutzer nicht immer in der Lage oder auch nicht Willens, die Suche und Auswertung selber vorzunehmen. Hier tut sich nun ein Feld für Broker auf, die Informationen gezielt vermitteln. In der bestehenden Architektur ist diese Tätigkeit leicht zu unterstützen. Der Broker muss lediglich angeben, welche Informationsobjekte er anbietet. Erhält er dann eine Anfrage, so startet er selber eine entsprechende Suche und wertet die Ergebnisse aus. Diese stellt er für einen bestimmten Betrag dem Suchenden zur Verfügung. Alle Dienste sind hierfür in der Architektur vorgesehen.

12 Ein Informationsnetz entsteht

Bisher wurden vor allem die technischen Aspekte, die für den Aufbau eines solchen Suchnetzwerkes benötigt werden, beschrieben. Im folgenden wird nun gezeigt, wie mit der Zeit ein Verbund entsteht.

12.1 Schrittweiser Aufbau

Wie bereits erwähnt, gibt es keine zentralen Instanzen. Somit braucht es auch keine ersten Standardisierungen oder Gremien, welche die Struktur von Informationsobjekten oder Infrastrukturen definieren. Die Idee ist vielmehr, dass die bestehenden Informationsanbieter in einem ersten Schritt die Dienste Registration, Lokalisation und Zugriff für ihre eigenen Informationsobjekte anbieten. Dabei können diese auch die neu erstellten Informationsobjekte anderen Registrations-Servern zukommen lassen. Bereits dieser erste Schritt stellt für den Suchenden und den Informationsbroker einen grossen Gewinn dar, da nun Informationen strukturiert über mehrere Server hinweg gesucht werden können. Am Anfang muss der Suchende jedoch noch selbständig die Server-Adressen dem System bekanntgeben. Dies ist jedoch nur für eine kurze Zeit notwendig, denn die Vergangenheit zeigt, dass wenn ein Dienst erfolgreich eingesetzt wird, zentrale Server aufgebaut werden.

12.2 Informationsinseln

Natürlich werden nicht alle Anbieter bestrebt sein, sich mit anderen zu vernetzen, da vielleicht eine gewisse Transparenz und somit auch Konkurrenz gefürchtet wird. Dennoch werden mit der Zeit kleinere und grössere unabhängige Informationsverbunde entstehen. Innerhalb solcher Verbunde wird man sich auch darüber verständigen können, wie ein Informationsobjekt für einen bestimmten Informationstyp strukturiert sein sollte. Die Anbieter müssen dann nicht, wie bereits erwähnt, alles ändern, sondern können via Transformationsbeschreibung ihre Definition in die allgemein anerkannte Form bringen. Der Aufwand für die Transformation ist gering.

Mit der Zeit werden auch die Informationsinseln zusammenwachsen, da der Suchende jede dieser Inseln sowieso ansprechen kann.

12.3 Konflikte

Da es keine zentrale Stelle für die Definition eines Informationstyps (z.B.: Hotelangebot) gibt, kann es passieren, dass es verschiedene Definitionen für ein und denselben Informationstyp gibt. Beim automatischen Verteilen der Definitionen kann es deshalb zu einem Konflikt kommen. Hierfür ist eine automatische Konfliktlösung vorgesehen. Jeder Registrations-Server akzeptiert die Definitionen, erzeugt jedoch eine spezielle Versionsnummer, damit eine Unterscheidung möglich wird. Gleichzeitig werden die Ersteller via Mail darüber informiert. Diese haben nun die Möglichkeit, entweder eine Transformation zu beschreiben oder sich mit den Anderen in Verbindung zu setzen und sich auf eine Version zu einigen. Mit der Zeit entsteht dadurch eine gewisse Einheitlichkeit. Dies zeigen die bereits vor-

handenen XML-Definitionen für bestimmte Branchenzweige, die in virtuellen Katalogen veröffentlicht werden (z.B.: The XML Catalog unter www.xml.org/xmlorg_registry/index.shtml).

13 Zusammenfassung

Das hier vorgestellte System löst nicht die bestehenden Systeme ab noch verdrängt es diese. Sondern es wird aufgezeigt, wie bestehende Informationsdatenbanken auf einfache Art und Weise miteinander verbunden werden können. Das System integriert und erleichtert somit dem Suchenden und Anbieter die Arbeit. Es werden in dieser Architektur keine Vorschläge gemacht, wie Informationen zu strukturieren und organisieren sind, damit diese effizient verwaltet werden können. Die Architektur beschränkt sich bewusst auf die Verwaltung von Informationsobjekten und deren Lokalisierung im Netz. Zusätzlich erhält der Suchende eine besondere Unterstützung, um die Informationen lokal selber nach seinen Vorstellungen auswerten zu können. Die vorgeschlagene Architektur verstärkt die Bemühungen um eine Homogenisierung der Suche in einer heterogenen Infrastruktur.

14 Ausblick

Es ist offensichtlich, dass die hier vorgestellte Architektur nicht nur zur Informationssuche, sondern vor allem auch im e/mCommerce zum raschen Auffinden von Produkten und Dienstleistungen verwendet werden kann. Der Vorteil ist hierbei, dass der Anbieter keinen kompletten Shop bereitstellen muss, sondern lediglich seinen Produktkatalog in das Suchsystem einbinden muss. Viele KMUs, die sonst nicht in der Lage sind einen adäquaten virtuellen Laden zu eröffnen, erhalten so die Möglichkeit, ihre Produkte und Angebote zu veröffentlichen. Zudem gibt es immer Situationen, in denen das virtuelle Einkaufserlebnis auf besonders effektvollen Seiten gar nicht wichtig ist, sondern dass vielmehr die gewünschte Aktion rasch erledigt werden kann.

Das elektronische Einkaufen könnte also eine neue Dimension erhalten. Der Benutzer besucht nicht virtuelle Geschäfte, sondern formuliert eine Anfrage und erhält wenige Zeit später individuelle Angebote auf seinem Desktop. Diese Angebote sind von verschiedenen Unternehmen, entsprechen den individuellen Bedürfnissen des Benutzers und sind trotzdem so detailliert wie nötig, d.h. sie nutzen Multimedia und andere moderne Technologien, um dem Benutzer den Kaufentscheid zu erleichtern.

Literatur

- [FuDi97] Fuhr N., Dittrich G., Tochtermann K. (Hg.): Hypertext – Information Retrieval – Multimedia 97. UVK Universitätsverlag Konstanz, 1997.
- [HaSc97] Handschuh S., Schmid B.F., Stanoevska-Slabeva K.: The Concept of a Mediating Electronic Product Catalog. In: International Journal of Electronic Markets, Vol. 7, No.3, vdf Zürich, 1997, S. 3-5.
- [Moor99] Moore R.W. et al.: Data-Intensive Computing. In: The Grid: Blueprint for a New Computing Infrastructure (Ed.: Foster I., Kesselmann C.), Morgan Kaufmann Publisher San Francisco, 1999, S. 105-130.
- [NwNd98] Nwana H.S., Ndumu D.T.: A Brief Introduction to Software Agent Technology. In: Agent Technology: Foundation, Application, and Markets (Ed.: Jennings N.R., Wooldridge M.J.), Springer-Verlag Berlin Heidelberg New York, 1998, S. 29-48.
- [ScLe99] Schneider B., Lederbogen K.: Navigationskonzepte für Internet Anwendungen. In: Information Management and Consulting, Bd. 14, 1999, S. 103-109.
- [StHo98] Stanoevska-Slabeva K., Hombrecher A.: Vor- und Nachteile verschiedener Suchmethoden und das Beispiel eines Attributbasierten Suchsystems, HSG St. Gallen, 1998.