



neofonie



Für das DFN

DFN Science-To-Science Abschlussbericht



Team: s2s-I (s2s-I@neofonie.de)

Stand: 15.9.2004, Version 1.0

Historie

Datum	Änderungen	Version	Bearbeiter
5.9.2004	Erstversion	0.1	rw
6.9.2004	Anhangskonzept	0.2	rw
7.9.2004	Zweitversion	0.3	rw
9.9.2004	AP6 / Statistik	0.4	rw
10.9.2004	Korrektur	0.5	ag
13.9.2004	Quellenverzeichnis	0.6	rw
14.9.2004	Korrektur	0.7	uf, rw, ag
15.9.2004	Eingehende Korrektur	0.8	rw
15.9.2004	Final Version	1.0	

Inhaltsverzeichnis

Historie	iii
Inhaltsverzeichnis	v
Liste der Anhänge	ix
1 Einleitung	11
1.1 Hintergrund	11
1.2 Zielsetzung	12
1.3 Wissenschaftlich-technischer Stand zu Projektbeginn	12
1.4 Planung und Ablauf	13
1.5 Projektergebnisse	13
1.5.1 Nicht-technische Ergebnisse	14
1.5.2 Wissenschaftlich-Technische Ergebnisse	14
1.6 Projektnutzen und -Zukunft	15
2 Arbeitspakete	17
2.1 AP 1 – Grundlagen	17
2.1.1 Hintergrund	17
2.1.2 Ergebnisse	18
2.1.3 Fazit	19
2.2 AP 2 - Provider	19
2.2.1 Zielsetzung	19
2.2.2 Ergebnisse	20
2.2.3 Fazit	22
2.3 AP 3 – Hub	22
2.3.1 Zielsetzung	22
2.3.2 Ergebnisse	23

2.3.3	Fazit	25
2.4	AP 4 - Consumer Server.....	25
2.4.1	Zielsetzung	25
2.4.2	Ergebnisse	25
2.4.3	Fazit	27
2.5	AP 5 - Consumer Client.....	28
2.5.1	Zielsetzung	28
2.5.2	Ergebnisse	28
2.5.3	Fazit	29
2.6	AP 6 – Betrieb	30
2.6.1	Zielsetzung	30
2.6.2	Ergebnisse	30
2.6.3	Netzwerk-Nutzung.....	33
2.6.4	Fazit	35
3	Dissemination	37
3.1	Vorträge.....	37
3.2	Kontakte und Zusammenarbeiten.....	38
3.2.1	DFN.....	38
3.2.2	HBZ NRW	38
3.2.3	FU Berlin, Neurobiologie.....	38
3.2.4	EDUTELLA / Uni. Hannover	38
3.2.5	DFN SINN.....	38
3.2.6	DFKI	39
3.2.7	DFN JOIN	39
3.2.8	ZIB.....	39
3.2.9	regioIT Aachen	39
3.2.10	Beta-Test + Nutzer	39
3.3	Veröffentlichungen	40

4	Quellenverzeichnis	41
----------	---------------------------	-----------

Liste der Anhänge

Anhang 1: S2S qua Verity Federator

Besuchen Sie bitte <http://s2s.neofonie.de/>

Weitere technische Informationen befinden sich in den Meilensteinberichten und deren Anhängen¹.

¹ Diese Informationen können von der neofonie oder dem DFN nachgefordert werden.

1 Einleitung

Dieser Bericht wurde gemäß der Richtlinien des DFN verfasst. Er dient als Abschlussbericht des Projektes DFN Science-To-Science (S2S). Die nachstehenden Informationen erfüllen formal die Anforderungen von Teil I und II, BEBF-AN 98, §9 Abs. 2.

Der Bericht besteht aus insgesamt 4 Teilen:

- Die Einleitung vermittelt einen Überblick über das Projekt, stellt den Ablauf des Projektes dar und erläutert die Ergebnisse.
- Der zweite Teil berichtet über die einzeln ausgeführten Arbeitspakete. Besondere technische Ergebnisse des Projektes werden vorgestellt. Tiefergehende Details befinden sich in den jeweiligen Meilensteinberichten.
- Der dritte Teil berichtet über weitere Resultate des Projektes, d.h. Vorträge, Kontakte, Zusammenarbeiten/ Kooperationen und Publikationen in der Fachpresse.
- Abschließend erscheint das Quellenverzeichnis des Projektes.

1.1 Hintergrund

Die Suche in wissenschaftlichen Quellen wird immer schwieriger. Das Internet wächst viel zu schnell, um sämtliche verfügbaren Quellen zentralistisch indexieren zu können (wie es z.B. Google praktiziert). Viele weitere Quellen sind im „Deep Web“ versteckt, obwohl sie für die Öffentlichkeit zugänglich sein dürfen. Es fehlt jedoch an der erforderlichen Organisation, diese Daten sichtbar machen zu können.

Vor allem in den Bereichen, in denen es um gemeinschaftliche Arbeit geht, wie insbesondere in der Wissenschaft, ist es vorteilhaft, die Quellen direkt von den Experten im Rahmen ihrer täglich zu verrichtenden Arbeit organisieren zu lassen. Dafür müssen jedoch einige Hürden überwunden werden. Unter anderem muss ein geeignetes System derartig einfach sein, dass auch IT-Laien in die Lage versetzt werden, übersichtlich die notwendigen Aufgaben ohne Mühe abzuschließen.

Mit dem Peer-To-Peer (P2P) Ansatz begegnet das Projekt DFN S2S diesen Herausforderungen. Es zeigt durch seine Implementierung einen möglichen Ansatz zur Bewältigung der aktuellen Probleme der Internet-Suche. Ein besonderes Merkmal des P2P-Ansatzes ist die Selbstverwaltung, die freie Ressourcen – auch organisatorische – verknüpft, um

einen Mehrwert zu erzielen. DFN S2S ist ein Team-Werkzeug, das in den Community Bereich skaliert. Es können Tausende von Teilnehmern im Netzwerk ihre Daten auffindbar machen, sich in Gruppen gliedern und eigene Interessen verfolgen.

1.2 Zielsetzung

Das Hauptziel des Projektes war es, ein Suchnetzwerk für die Wissenschaft auf Basis der neofonie Suchmaschine (neofonie search Produktlinie; vgl. [w2, w9, w10]) und der JXTA P2P-Technologie [w3], zu entwickeln und zu betreiben. Die entwickelten Netzwerk-Komponenten sollten den oben genannten Anforderungen gerecht werden und praktisch einsetzbar sein. Es sollte anhand der Projekt-Software möglich sein, eine eigene Suchmaschine aufzusetzen, die den eigenen Teil des unsichtbaren Web sichtbar machen sollte.

Es sollte anhand der Software auch möglich sein, für andere Bereiche oder ganz andere Nutzergruppen eigene Netzwerk aufzusetzen.

Ein weiteres Ziel bei der Arbeit mit der JXTA-Technologie war es, auch einen Beitrag zur JXTA Search [w1]-Spezifikation zu leisten.

Die Ziele des Projektes stimmen überein mit den Zielen des Entwicklungsbereichs „Informationsdienste“ des DFN. Die Entwicklungen, die in diesem Projekt realisiert worden sind, sollten Suchmaschinen und -Technologien, die Integration zwischen Diensten, Dataming-Technologien sowie die Arbeit an Standards fördern.

1.3 Wissenschaftlich-technischer Stand zu Projektbeginn

Zu Projektbeginn wurde durch populäre Projekte, wie z.B. Napster, das Interesse an der P2P-Technologie gesteigert (vgl. [b2]). Eine Reihe von Forschungsprojekten beschäftigte sich mit diesem sehr neuen Forschungsthema (vgl. [3, 7, 11, 12, 14, 19, 25, 27]) und der zugrunde liegenden mathematischen Basis [2, b1].

P2P-Netzwerke waren hybride (wie Napster) oder benutzten einfache Algorithmen (Gnutella) [6]. Die sogenannte dritte Generation der P2P-Netzwerke, auf Basis von verteilten Hash-Tabellen gebaut, war noch nicht praktisch erschienen [14, 19, 23, 25].

Die JXTA-Software hatte die Version-Nr.1.0 (b57a_040402). [26, w3]

Im Bereich Suche in P2P-Netzwerken gab es noch keine Varianten mit Volltext-

Unterstützung² (vgl. [3, 10, 13, 15]).

1.4 Planung und Ablauf

Schon im Projektantrag wurde die Projektplanung ausführlich skizziert. Die technische Projektplanung erfolgte ohne größere Abweichungen vom Plan. Auch die erwarteten Aufwände wurden eingehalten, allerdings mit anderer Verteilung als ursprünglich geplant. Änderungen vom Plan wurden zeitnah dem DFN gemeldet, bzw. mit dem DFN abgesprochen. Zu Beginn des Projektes wurde etwas mehr am Projekt gearbeitet als erwartet, im mittleren Bereich hat die Flaute in der Branche auch die neofonie beeinträchtigt und die Aufwände im Projekt verringert. Demgegenüber wurden abschließend die Arbeiten unmittelbar vor der Pilotphase noch einmal deutlich intensiviert.

Im voraus sollte festgestellt werden, wer als möglicher Betreiber des Systems agieren könnte. Identifiziert wurden Internetportale und das DFN selbst. Einige Internetportale³ wurden mit vorformulierten Projektideen diesbezüglich kontaktiert. Im allgemeinen wurde das P2P-Marktmodell⁴ vom Projekt DFN S2S als „vielversprechend“ bewertet. Die JXTA-Technologie wurde als „unausgereift“ beurteilt (vgl. [9]). DFN und neofonie sind zur Zeit Betreiber des Projektes.

1.5 Projektergebnisse

Technisch und wissenschaftlich ist das Projekt ein Erfolg. Es wurden alle im Rahmen der Meilensteine umzusetzenden Vorgaben realisiert und es ist eine nutzbare neue Technologie entstanden.

Auch an der Anzahl der Kontakte, die durch Umsetzung dieses Projekts entstanden sind, sowie an der Generierung eines Auftrags, der eine Weiterentwicklung der Software sichert, zeigt sich der Erfolg dieses Projekts⁵. Bei der Bekanntmachung und Verbreitung der Projektergebnisse wurde das Projektziel allerdings nicht vollständig erreicht. Im Sommer 2003,

² [24] hat Volltext-Unterstützung, erlaubt jedoch nicht, dass der Nutzer selber Quellen auswählt.

³ Im Laufe des Projektes wurde z.B. ein Seminar „P2P @ AOL“ (2002.09.11) gehalten.

⁴ Das Marktmodell basiert auf Einsatz-Szenarien beschrieben in [b2, b6]

⁵ Diese Bewertung erfolgt anhand eines Vergleichs mit anderen geförderten Forschungs- und Entwicklungsvorhaben der neofonie.

als die Beta-Test Phase startete, sorgte die Pressestelle des BMBF für eine große Nachfrage von (potentiellen) Beta-Testern (insgesamt über 400 Anmeldungen aus Wissenschaft und Industrie wurden entgegengenommen). Für die Pilotphase konnte diese Resonanz nicht erzielt werden, obwohl neofonie und das DFN sich intensiv darum bemüht hatten. Somit war die Anzahl der Nutzer auch eher minimal. Das Netzwerk konnte dadurch nicht signifikante Ausmaße erreichen, wodurch eine automatisierte bzw. selbständige Verbreitung des Konzeptes in dem gewünschten Umfang stattgefunden hätte.

1.5.1 Nicht-technische Ergebnisse

- Das DFN S2S-Suchnetzwerk steht seit dem 11. Februar allen Wissenschaftlern im deutschsprachigen Raum zur Verfügung. Durch Herunterladen⁶ der S2S-Software kann jeder Teilnehmer ganz einfach seinen Arbeitsrechner als Suchmaschine an das Suchnetzwerk anbinden.
- neofonie hat anhand dieses Projektes viele wirtschaftlich und technologisch wichtige Kontakte geknüpft (vgl. Abschnitt 3) und viel Aufmerksamkeit erzielt. Dies ist ein wichtiges Ergebnis des Projektes, das jedoch schwer zu quantifizieren ist. Das Projekt wurde in zwei uns bekannten Dissertationen erwähnt (Claus Eikemeier, Universität Bremen; Sam Joseph, Universität Tokyo), sowie in vielen Artikeln, die auch online zu finden sind. Das Projekt ist gut im Internet repräsentiert, vgl. Google Suchanfrage⁷ mit ungefähr 450 Treffern.

1.5.2 Wissenschaftlich-Technische Ergebnisse

Folgend werden zentrale technologische Ergebnisse des Projektes aufgeführt und erläutert (weitere Details können dem folgenden Abschnitt entnommen werden):

- Ein Portal im Internet erlaubt allen interessierten Suchenden Zugang zum Netzwerk. Die Portal-Software mitsamt Design offenbart auf einfache Weise, wie S2S funktioniert.

⁶ Mitarbeiter des BMBF und ihre Gutachter können die Downloadseite <http://s2s.neofonie.de/navigation/index.jsp?nav=39265> mit folgenden Zugangsdaten erreichen: Als „Email-Adresse“ geben Sie bitte bmbf ein; als „Kennwort“ geben Sie bitte 3027407 ein.

⁷ Bei einer Suche nach *“DFN Science To Science” OR “DFN S2S” OR “DFNS2S”*, <http://www.google.com/search?q=%22DFN+Science+To+Science%22+OR+%22DFN+S2S%22+OR+%22DFNS2S%22>

- Komponenten, die das S2S Netzwerk aufbauen und funktionieren lassen, sind Hub, Provider und Consumer. Die Komponenten wiederum basieren auf einem gemeinsamen P2P-Rahmenwerk, das die kommunikativen Aufgaben im Netzwerk regelt.
- Die neofonie search-Suchsoftware wurde erweitert und im Hub und Provider eingesetzt. Damit wurden wichtige Fortschritte in der neofonie - Suchtechnologie erzielt.
- Anhand der neofonie XML Indexierer-Technologie konnten neue Verfahren für die Kommunikation von Suchanfragen durch Hubs entwickelt werden. Diese Verfahren erlauben eine verbesserte Inhalt-basierte Suche in P2P-Netzwerken (vgl. AP3 – Hub). Die sogenannte Auto-Filter-Technologie ist eine Spam-Technik, die Ergebnisse von Providern nutzt, um sie später automatisch zu filtern. Es erfolgt im Netzwerk ein automatischer Spamschutz.
- Durch Nutzung des XML-Indexierers in allen Providern sind neue Möglichkeiten für die Sortierung und Aggregation von Daten aus heterogenen Datenquellen in Echtzeit eröffnet worden. Diese Technologie erlaubt, dass dezentrale Suchmaschinen über diverse Datenquellen effizient funktionieren (vgl. AP3 – Hub).

1.6 Projektnutzen und -Zukunft

Der Projektnutzen ergibt sich unmittelbar aus den weiter oben genannten Projektergebnissen. Die neofonie- Suchtechnologie wurde schon mehrfach anhand des Portals demonstriert. Auf Basis der iDB Technologie (vgl. AP2 – Provider) wurde z.B. „SmartSearch“ für WEB.DE realisiert. Auch andere Teil-Entwicklungen des Projektes haben bereits wirtschaftliche Erträge gebracht. Die neuen Suchprodukte der neofonie :suchexpress⁸ und :newsexpress⁹ profitieren von den Ergebnissen des Projektes.

Auch die Projektsoftware als Ganzes hat weiteren Nutzen außerhalb des Projektes erzielt. Bereits im August 2004 wurde z.B. eine Anpassung von S2S für Rechtsämter im öffentlichen Dienst von der RegioIT in Aachen beauftragt. Ein eigenes, direkt auf S2S basierendes Netzwerk wird in Aachen entstehen, mit Providern in einigen Rechtsämtern. Sollte das Projekt WikoR (ein

⁸ <http://www.suchexpress.de>

⁹ <http://www.newsexpress.de>

BMWA-gefördertes Projekt) in dem vorgesehenen Rahmen bewilligt werden, wird S2S auch noch erweitert werden, so dass Sicherheitsaspekte, Benutzergruppen, Rollen und Ontologien aus dem Rechtsumfeld berücksichtigt werden (vgl. [8, 21, 22, 26, w8]).

In April 2004 haben sich neofonie, das DFKI und die Universität Hannover (L3S) zu einem Konsortium konstituiert, um die S2S Technologie zu verbessern und zu erweitern. Dieses Projekt passt auch sehr gut zu den eScience-Vorhaben des BMBF. Kontakt wurde mit dem DLR schon aufgenommen und es besteht die Möglichkeit, die erzielten Ergebnisse im Rahmen von eScience zu verbessern und weiter zu entwickeln. Ein Konzept für die Einbindung der Technologie in eine Grid - Umgebung (auf Basis von [28]) wurde von neofonie nach Projektschluss erarbeitet.

2 Arbeitspakete

2.1 AP 1 – Grundlagen

2.1.1 Hintergrund

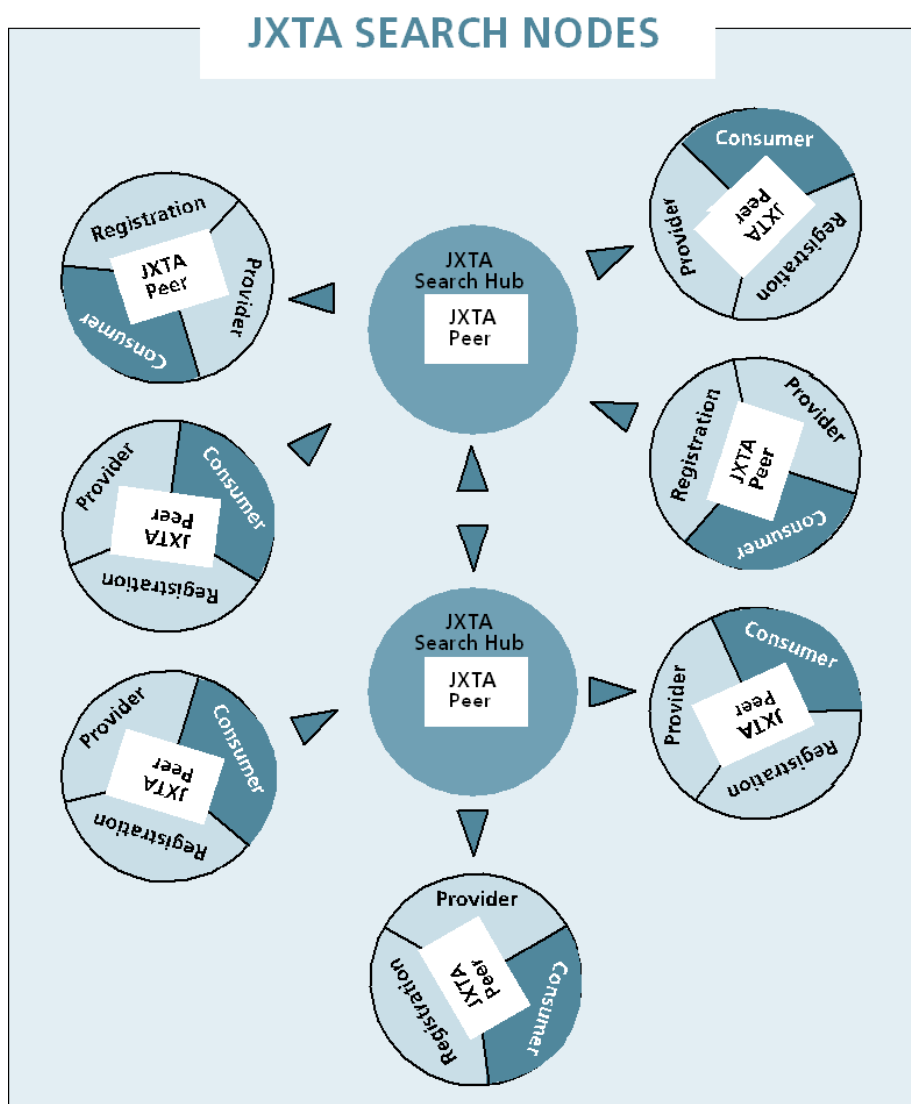


Abbildung 1. Beispiel JXTA Search-Architektur

JXTA ist ein Community Projekt (von Sun Microsystems geleitet), das eine gleichnamige P2P-Kommunikationsspezifikation und einige Referenzimplementierungen der Spezifikation präsentiert. Auf Basis der Referenzimplementierung gibt es ein öffentliches JXTA-Netzwerk.

Die JXTA Search Spezifikation wurde als Basis des S2S Netzwerkes ausgewählt. Noch nicht als Standard akzeptiert, aber durch Unterstützung von Sun bekannt, definiert die Spezifikation, wie ein Suchnetzwerk spontan aufgebaut werden kann und welche Nachrichten zwischen Peers übermittelt werden sollten. Die XML Definition war möglichst einfach ausgelegt, um auch über Web Services oder einfachem HTTP zu funktionieren, allerdings war die hauptsächlich vorgesehene Anwendung im JXTA P2P-Netzwerk vorgesehen.

JXTA Search definiert einige Rollen von Peers, die entweder Suchanfragen stellen dürfen (Consumers - AP4/5), antworten können (Providers - AP2) oder das Netzwerk aufbauen und Suchanfragen nach einfachen Heuristiken weiterleiten (Hubs - AP). In der Abbildung sieht man ein typisches JXTA Search Netzwerk mit der Rollenverteilung nach Funktion.

Für das Projekt Science-To-Science reichte das Protokoll allerdings nicht. Wesentliche Änderungen in Form von Ergänzungen zum Protokoll wurden identifiziert und entwickelt. Der direkte Austausch von Information zwischen den Peers und die Kommunikation von Bewertungen wurden ermöglicht.

Auch ein projektinternes einheitliches Metadaten-Format musste gewählt werden.

2.1.2 Ergebnisse

Es wurde ein leichtes P2P-Kommunikations-Rahmenwerk konstruiert, das die Nachrichtensendung und den Nachrichtenempfang von und zum JXTA-Netzwerk einheitlich kapselt. Das Rahmenwerk wurde unverändert von allen Java Peers genutzt, um einen Zugang zum P2P-Netzwerk zu bekommen. Änderungen an JXTA konnten so leicht gehandhabt werden.

Die im Rahmen des Projektes entwickelten Änderungen an JXTA Search sind in [The DFN Science-To-Science JXTA Application](#) (vgl. Abschnitt 3.3) definiert.

Anhand dieser Änderungen konnte eine Suche im P2P-Netzwerk besser organisiert werden. Eine Suche im Netzwerk liefert nicht die Originaldokumente von den Peers (die beachtlich groß sein können, wie z.B. bei der Neurobiologie FU, wo eine Mehrzahl der Dokumente um 1Mb Größe beträgt), sondern liefert kleine Ausschnitte aus den Dokumenten, wie auch Usus in gewöhnlichen Internet Suchmaschinen. Geeignete Dokumente können in einem weiteren Schritt vom Provider Peer heruntergeladen werden. Die Last, die auf dem Netzwerk liegt, ist somit viel geringer und der Nutzer wird nicht mit unnötiger Information überflutet - die Suche

ist also gezielter.

Zusätzlich kann der Nutzer einen Provider (Quelle von Dokumenten) bewerten. Diese Bewertung ist auch für andere Nutzer sichtbar. Sie wird auch intern vom System für die Weiterleitung (Routing) von Suchanfragen genutzt (s.a. AP3).

Als Basis der Metadaten in S2S wurden die Standard-Metadaten des neofonie Robotersystems mit dem Standard-Metadatensatz der Dublin Core [w7] Initiative zusammengefügt. Dieses Ergebnis wurde erzielt nach einer eingehenden Untersuchung der vorhandenen Optionen (vgl. [4, 11, w5]). Die resultierende Metadatenammlung hat sich während der Ausführung bewährt. Datensätze Dritter konnten problemlos an das System angeschlossen werden, z.B. MARC Daten des Hochschulbibliothekzentrums des Landes Nordrhein-Westfalens (HBZ-NRW), PDF Dateien des Neurobiologischen Instituts der FU Berlin sowie einige OAI Daten (z.B. die von Citeseer, vgl. AP6).

Es wurde auch ein Bezug zur QEL Query Sprache des Projektes Edutella [w4] hergestellt. Dieser neue Standard dient als hochwertige Sprache für den Austausch von Suchanfragen zwischen heterogenen verteilten Systemen. Die QEL in ihrer Version 1.0 gab es erst in Q3 2003.

2.1.3 Fazit

Die Ziele dieses Arbeitspakets, eine Verbindung zum JXTA-Netzwerk zu gewährleisten und ein Metadaten-Schema zu definieren, sind frühzeitig im Projekt erreicht worden. Das Metadaten-Schema erwies sich als flexibel genug für alle Daten, die während des Projektes aufgetreten sind.

Die JXTA-Anbindung an das Netzwerk änderte sich während der Ausführung geringfügig, aber da Änderungen an den Schnittstellen vorgesehen waren, wurden alle JXTA-Versionen von 1.0 bis 2.2.1 problemlos vom Projekt genutzt.

2.2 AP 2 - Provider

2.2.1 Zielsetzung

Ziel dieses Arbeitspakets war es, ein Werkzeug bereitzustellen, welches dem Nutzer eine eigene steuerbare Volltext Suchmaschine, Plattform unabhängig, zur Verfügung stellt. Die Software musste dezentral laufen und über ein P2P-Netzwerk von Wissenschaftlern genutzt werden können ohne Mithilfe eines Administrators. Die Wissenschaftler sollten ohne spezielle IT

Kenntnisse in der Lage sein, das System zu verstehen.

Bei der Umsetzung dieses Pakets wurde auch auf die Wünsche der Nutzer eingegangen. In der letzten Version der P2P-Software wurde ein einfacher Sicherheitsmechanismus zum Schutz von Datenrechten installiert. Zudem wurde ein Provider Peer für den Mac entwickelt.

2.2.2 Ergebnisse

2.2.2.1 :robot

Das neofonie search:robot Software-Paket (ein Dokumentverarbeitungs-Roboter, der beliebig erweitert werden kann) wurde für dieses Projekt kostenlos zur Verfügung gestellt. Folgende Erweiterungen wurden vom Projekt benötigt und im Rahmen des Projektes vorgenommen.

- Generierung von Dokumenten-IDs aus den Dateinamen der Dokumente.
- Unterstützung für die Traversierung von FTP Servern.
- Unterstützung für die Traversierung von OAI Repositories. [w6]
- Verwendung einer eingebetteten Java Datenbank (HSQLDB) für die Verwaltung von Zustandsinformationen über Dokumente.
- Scheduling-Funktionalität, um die Aktionen des Systems auf vom Endnutzer festgelegte Zeitfenster einzuschränken.
- Unterstützung der Dokumentenformate LaTeX und PDF und die Abbildung dieser Dokumentenformate in die interne Dokumentenstruktur.
- Unterschiedliche Wartungszeiten je nach Quelle: so mussten Festplatten („shared Verzeichnisse“ viel öfter besucht werden als eine Webseite).

2.2.2.2 :engine

Auch das neofonie search:engine Software-Paket (der flexible und skalierbare XML Indexierer der neofonie) wurde für dieses Projekt kostenlos zur Verfügung gestellt. Folgende Erweiterungen wurden vom Projekt benötigt und im Rahmen des Projektes vorgenommen:

- Austausch der Berkeley DB durch eine Eigenimplementierung: diese Änderung war aus mehreren Gründen wünschenswert. Die verbreitete Berkeley DB Software ist nicht auf unseren Bedarf optimiert und die notwendige Lizenz für S2S hätte mindestens \$30'000 gekostet.

- Portierung (der C++-Anteile) auf die Plattform Intel/Win32: der Indexierer ist nicht einfach portierbar wegen seiner Plattform-Optimierten Umsetzung. Die Portierung auf Win32 (die Schnittstelle, die alle Anwendungen auf Windows 98, ME, 2000, XP benutzen müssen) wurde über die Cygwin Interoperabilitäts-Software vorgenommen.
- Portierung auf Mac OS X: der Indexierer wurde zusätzlich auf Mac OS X portiert und ein Mac-Peer wurde auf Nachfrage von Nutzern entwickelt.
- Es wurden auch Relevanzbewertungsoptimierungen, zum Teil nach PageRank [16], vorgenommen. Diese hingen zusammen mit einem Mechanismus, der die relevanten Terme eines Providers vermittelt, zwecks Registrierung beim Hub [18].

2.2.2.3 Sicherheitsmechanismus

Ziel war es, auf Grund von Nutzeranregungen eine Möglichkeit zu schaffen, den Download von geschützten Dokumenten zu unterdrücken. Für den Dokumenten-Download können zwei Zugangsmethoden genutzt werden:

- Die direkte Methode über eine URL,
- Die indirekte Methode über das P2P Netzwerk (über Proxy-Rechner falls keine direkte Verbindung existiert).

Dokumente, die in einem Intranet vor direkten (über ihre URL) externem Zugriff geschützt, sind können prinzipiell in einem P2P Netzwerk über die zweite Zugangsmethode heruntergeladen werden. Der im Projekt entwickelte Sicherheitsmechanismus schaltet nun diese Zugriffsmethode aus. Externe Nutzer können geschützte Dokumente über das P2P Netzwerk zwar finden und Metainformationen über deren Inhalte erhalten, ein weiterer Zugriff ist aber nur über Absprachen mit dem Provider möglich.

2.2.2.4 Installationsmechanismus

Als die funktionale Hauptentwicklung zu Beginn des Beta-Tests abgeschlossen war, fing die Arbeit an den Oberflächen und an den Installationsmechanismen an. Diese Arbeit nahm eine zentrale Rolle in Q3/2003 ein. Das Ergebnis war eine leicht zu bedienende und auf Standard-Verfahren basierte Installationssoftware. Die Entwicklung erfolgte auf Basis der Installshield Software.

Während der Installation mussten Nutzer einer S2S-Lizenz¹⁰ zustimmen. Die Lizenz bestimmt die Rechte und Pflichten von Nutzern und Anbietern bei der kostenlosen Nutzung der Software.

Während der Installation wurde auch die initiale Konfiguration der Software vorgenommen. Die flexiblen unterliegenden Komponenten haben XML-Konfigurations Dateien, die für Nutzer eine unzumutbare Komplexität aufweisen. Diese können jetzt (natürlich sehr beschränkt) über einfache Oberflächen gesteuert werden. Auch während des Betriebes können alle nutzerbezogenen Einstellungen geändert werden.

2.2.3 Fazit

Das Provider-Software-Paket formte aus Nutzersicht das Herzstück des Projektes und wurde 600 Mal heruntergeladen.

Die Erweiterungen der neofonie search Produkte, die im Rahmen des Projektes durchgeführt wurden waren erforderlich um die Funktionalität der Peers auf unterschiedlichen Plattformen den Anforderungen entsprechend zu gewährleisten. Sie sind darüber hinaus auch für andere Anwendungen einsetzbar. Folgende Beispiele sind erwähnenswert: Die Einbettung der Java Datenbank HSQLDB hat die Effizienz des Roboters um eine Ordnungsgröße gesteigert. Vorherige Versionen ohne angeschlossene Datenbank konnten nur 1000 Dokumente bearbeiten, mit HSQLDB können bis zu 50'000 Dokumente bearbeitet werden.

Auch im Bereich der nutzerfreundlichen Installation sind wesentliche Ergebnisse erzielt worden.

2.3 AP 3 – Hub

2.3.1 Zielsetzung

Der Hub nimmt eine wichtige Rolle im Netzwerk ein. Hier befinden sich die zentralen Prozesse wie

- Steuerung des Datenflusses im Netzwerk (Routing) – die Suchanfragen und - Ergebnisse sollen möglichst optimal im Netzwerk fließen,
- Sicherheitsvorkehrungen durch Aufzeichnung der Provider Aktionen (Provider

¹⁰ Die S2S-Lizenz kann unter <http://s2s.neofonie.de/navigation/index.jsp?nav=333265> eingesehen werden.

Profiling) wie z.B. Kontrollen der Nutzung um Missbrauch aufzufinden, und

- Daten-Management – nicht alle Provider antworten gleich schnell, die Ergebnisse müssen den Nutzer auf sinnvolle Weise aggregiert geliefert werden.

Vorrangig versucht man folgende Frage zu beantworten: welche Peers, die zurzeit online sind, sind am besten geeignet, eine bestimmte Suchanfrage zu beantworten?

2.3.2 Ergebnisse

2.3.2.1 Provider Profiling

Zu jedem Provider Peer werden im Netzwerk eine Reihe von Merkmalen gespeichert. Diese Merkmale werden vom System sowohl aus den Registrierungsangaben des Providers als auch aus den gelieferten Suchergebnissen extrahiert und in einem :engine Index durchsuchbar aufbewahrt. Diese Merkmale werden von den Hubs für die Auswahl geeigneter Provider für eine Suchanfrage ausgewertet bzw. dem Nutzer zusammen mit den gefundenen Treffern dargestellt.

Auch statistische Merkmale der Provider wie Anzahl der gelieferten Treffer, Verbindungsgeschwindigkeit (Anzahl der Bytes geliefert durch Zeit in Millisekunden), die Anzahl der aufgenommenen Verbindungen, Nutzerbewertungen, seit wann der Provider bekannt ist und wie lange er ununterbrochen online ist werden gespeichert und ausgewertet.

2.3.2.2 Daten-Management: Caching und Aggregation

Es wurde ein eigener Cache entwickelt, der die Suchergebnisse eines Providers auf der Festplatte speichert. Suchergebnisse werden einer Suchanfrage zugeordnet. Wenn die Suchanfrage veraltet ist, werden die gesamten Informationen für diese Suchanfrage gelöscht. Wie lange Suchergebnisse für eine bestimmte Suchanfrage angenommen werden, kann separat gesteuert werden. Wird die gleiche Suchanfrage mehrfach gestellt, können die Suchergebnisse viel schneller zusammengestellt werden, da das Netzwerk nicht angefragt wird. Gleich wichtig ist, dass langsame Provider auf diese Art auch mit in den Resultaten erscheinen – der Nutzer braucht nur seine Suchanfrage zu wiederholen. Ergebnisse, die aus dem Cache zusammengestellt waren, werden mit Cache-Zeiten angereichert. Dies ermöglicht, dass das Frontend die neuen Treffer hervorheben kann.

Zur Aggregation wurde ein vergleichendes Paper zwischen S2S und einem Konkurrenzprodukt (Verity Federator) geschrieben. Das Paper „S2S qua Verity Federator“ ist im

Anhang beigefügt worden.

2.3.2.3 Routing

Nach dem JXTA Search Verfahren werden folgende Schritte ausgeführt, um eine Auswahl an passenden Providern für die Suchanfrage zu treffen:

1. Eingehende Suchanfragen werden mit Provider Registrationen verglichen.
2. Die Suchanfrage wird ein weiteres mal mit der Registrierung verglichen, weil Provider auch Suchanfragen ausschließen dürfen sollen¹¹.
3. Falls zu wenige Provider ausgewählt wurden, müssen die ähnlichen Provider erneut aufgespürt werden.

Das S2S Verfahren wählt ähnliche Provider anhand der Provider Profile aus. Zusätzlich kann ein Provider auch anhand seiner gelieferten Suchergebnissen ausgewählt werden. Das Netzwerk lernt praktisch mit der Zeit, welche Inhalte wo zu finden sind. Die Nutzung der :engine Technologie an dieser Stelle bedeutet, dass Provider in Reihenfolge von Relevanz mit Bezug auf das Gesuchte gewählt werden. Wenn mehrere Provider als geeignet gefunden werden, können die Bestgeeigneten z.B. durch Merkmale wie Geschwindigkeit und Beliebtheit ausfindig gemacht werden. Diese Algorithmen sind besonders dann von Bedeutung, wenn in einem P2P-Netzwerk sehr viele Peers mit begrenzten Ressourcen bedient werden müssen.

2.3.2.4 Hub Chaining Protokoll

Es wurde im Projekt auch ein Konzept für ein Protokoll erstellt, womit sich die Hubs allgemein verketteten können. Dieses Protokoll ist, wie JXTA selber, unabhängig von der Anordnung der Peers (Topologie). Eine beliebige Topologie kann auf diesem Grundgerüst aufsetzend, entwickelt werden. Das HCP ist in einem nicht veröffentlichten Paper zusammengefasst. Die Arbeit wurde im Rahmen einer Diplomarbeit angefangen¹².

¹¹ Dieser Schritt wird nicht von dem S2S Hub ausgeführt, da er vergleichsweise wenig Nutzen im System bietet: die Komplexität des Matchings steigt sehr – was einen potenziellen Bottleneck einführt und Nutzer wollen keine solche Mehraufwand eingehen.

¹² Robert Haensel war als Diplomand an der neofonie im Rahmen des Projektes von April bis August 2003 tätig (haenselr@cs.tu-berlin.de). Er hat die Arbeit aus persönlichen Gründen abgebrochen. Durch seinen Abgang sind einige Baustellen bis zum Projektende offengeblieben. Sie können aber jederzeit im Rahmen eines wissenschaftlichen Projektes wieder aufgenommen werden.

2.3.3 Fazit

Der Hub dient als bewegliches Zentrum des S2S Netzwerkes. Durch Installation eines Hubs kann ein unabhängiges S2S Netzwerk mühelos eingerichtet werden. Der Hub basiert auch auf neofonie search:engine und kann voraussichtlich als eigenes Produkt für die Aggregation von Daten in Wissensmanagement Systemen benutzt werden, allerdings müssten einige Erweiterungen vorgenommen werden (vgl. Anhang I).

Folgende Erkenntnisse, die auf sehr interessante zukünftige Forschungsfelder deuten, wurden gewonnen:

- Ein heuristisches Routing wie das verwendete, kann gute Ergebnisse liefern.
- Die Routing-Methode mittels verteilter Hash-Tabellen kann bei der Verkettung von Hubs die Netzwerk Effizienz steigern, obwohl Hash-Tabellen für eine Volltext Suche nicht optimal geeignet sind [3].

2.4 AP 4 - Consumer Server

2.4.1 Zielsetzung

Unter Verwendung des Server-basierten Search-Consumer können Endnutzer über ein Web-Interface im DFN S2S Netzwerk suchen und gefundene Dokumente laden. Aus Erfahrungswerten wissen wir, dass die Mehrzahl der Nutzer in einem P2P-Netzwerk stöbern und Informationen herunterladen will. Der Search Consumer soll diesen Bedarf abdecken.

2.4.2 Ergebnisse

2.4.2.1 ATPS¹³ Handler

ATPS ist das CMS¹⁴-Anbindungs Rahmenwerk von neofonie, auch als Teil des Produktes neofonie content [w11] vorhanden. Mit ATPS können diverse dynamische Datenquellen wie Datenbanken, andere Websites, etc. in eine Seite integriert werden. Die Gestaltung der Seiten bleibt einheitlich und wird durch eine eigene Taglib ermöglicht.

¹³ Active Template Publishing System

¹⁴ Content Management System

Für die Anbindung an den Hub wurde die JXTE Micro Edition (JXME) Schnittstelle benutzt, um einen ATPS Handler zu realisieren. Es gibt also einen JXTA-fähigen ATPS Handler, der aus einem beliebigen JXTA-Netzwerk Inhalte in eine Website einbinden kann. In diesem Falle ist der Handler auf das S2S Netzwerk ausgerichtet und erlaubt der Website, diverse Informationen über das Netzwerk darzustellen.

Auf der S2S Website <http://s2s.neofonie.de/> kann eingesehen werden, wie viele Anbieter gerade online sind. Auch die Suche selbst erfolgt über diesen Weg. Hier kann die Last, die auf den Hub ausgeübt wird, geregelt werden.

2.4.2.2 Webanwendung

Eine eigene Webanwendung wurde entwickelt, um die Suche und den Download aus dem Internet zu erlauben. Die Suchanwendung ist in der Struktur ähnlich wie bekannte Web-Suchanwendungen, liefert allerdings zusätzliche Informationen. Zu jeder Suchliste wird die Routing Statistik ausgegeben: die Anzahl der Provider, die angefragt wurde, die Anzahl, die geantwortet hat und die Anzahl derjenigen, die Treffer für die Suche hatte. Zu jedem Suchtreffer werden zusätzlich zu Titel, Autor, Kurzfassung, etc. auch Informationen zu den liefernden Providern angegeben. Deutlich zu sehen ist die S2S-Bewertung des Providers (vgl. „Provider Profiling“). Diese Information kann bei der Auswahl der Treffer zum Herunterladen wichtig sein. Es ist auch möglich, die Suche auf einen besonderen Provider einzuschränken.

The screenshot shows a web browser window displaying the DFN S2S Consumer Server interface. The browser's address bar shows the URL: `http://s2s.neofonie.de/navigation/index.jsp?nav=39268&q=title%3A+%28%22query+routin`. The page features a navigation menu with tabs for 'Home', 'Suche', 'News', 'Interaktiv', and 'Hilfe'. Below the menu, there are search options: 'Einfache Suche', 'Erweiterte Suche', and 'Dokumente anbieten'. The main content area displays search results for the query 'query routing'. It shows 2 documents found at 1 provider. The first document is titled 'Citations: Scalability issues for query routing service discovery - Gibbins, Hall [...]' and is 7 KB in size, dated 27.05.2004. The second document is titled 'Citations: Query routing in large-scale digital library system - Liu (ResearchIndex)' and is 15 KB in size, dated 28.05.2004. On the left side, there are sections for 'S2S News - Aktuelles' with a list of news items and a small image of a satellite, and 'Was ist S2S?' with a diagram of the network structure. A recommendation section at the bottom left suggests visiting 'neofonie search.newsexpress'.

Abbildung 2. DFN S2S Consumer Server

2.4.3 Fazit

Die Websuche im S2S-Netzwerk ist so gestaltet, dass die Features des Netzwerkes optimal angezeigt werden können. Diese Oberfläche wurde bei allen Demonstrationen der Software verwendet und von den Beobachtern sehr positiv bewertet, da der Ablauf im Netzwerk gut nachvollzogen werden konnte.

2.5 AP 5 - Consumer Client

2.5.1 Zielsetzung

Ziel dieses Arbeitspakets war es, über die Möglichkeiten des Consumer Servers hinaus zwei weitere Zugriffsmöglichkeiten auf das Netzwerk bereitzustellen:

1. Zugriff direkt über einen nativen JXTA-Consumer. Da dieser Peer direkt an das JXTA-Netzwerk angebunden ist, stehen ihm mehr Möglichkeiten zur Verfügung – diese sollten exemplarisch entwickelt werden.
2. Zugriff über mobile Geräte. Es sollte möglich sein, im Netzwerk per Handy zu suchen.

2.5.2 Ergebnisse

2.5.2.1 Nativer JXTA-Consumer Client

Ein nativer Consumer Client wird automatisch installiert, wenn man einen Provider installiert¹⁵. Dieser Consumer ist eine Java Swing Anwendung, die direkt auf den Arbeitsplatzrechnern der Endnutzer ausgeführt wird. Die Funktionalität ist identisch mit dem Server-basierten Search Consumer, allerdings mit anderer Ausführung. Ein nativer Consumer Client lädt zum Beispiel Dateien direkt vom Provider herunter.

Folgende Erweiterungen sind in der Entwicklung vorgenommen worden:

Ein Provider Profiling auf dem Consumer erlaubt es dem Nutzer, direkt an bestimmte Provider Suchanfragen zu stellen.

Der Zugriff auf den eigenen Provider ist immer gewährleistet (wenn das Internet und damit auch der Hub nicht erreichbar sind, kann man normalerweise den Provider nicht erreichen, obwohl er auf dem gleichen Rechner läuft).

¹⁵ Die Anwendungen teilen viel Code, also stellt dies keine Ressourcen-Verschwendung dar.

2.5.2.2 Mobiler Zugriff Auf S2S

Die Marktforschung¹⁶ zeigt, dass z.B. in 2003 deutlich mehr sogenannte „Smart Phones“ (Handys die Code ausführen können) als PDAs verkauft wurden. Um ein möglichst breites Publikum zu erreichen, lag der Fokus bei der Realisierung des mobilen Zugangs auf der Nutzung des Netzwerkes mit Handys. Dafür musste die Menge der Feature eingeschränkt werden, da die Geräte in der Regel nicht die Fähigkeiten eines PCs haben. Um die entwickelte Verfahren auch für PDAs nutzbar zu machen wurde die MIDP Programmier-Schnittstelle verwendet. Dementsprechend kann der mobiler Zugriff für Handys mit kleinen Anpassungen auch für PDAs benutzt werden.

Ein MIDP Programm wurde entwickelt, das die Suche im S2S Netzwerk erlaubt. Das Programm wurde in einem MIDP-Simulator von Motorola getestet. Es ist möglich mit einem MIDP-fähigen Gerät eine Suche im S2S Netzwerk zu starten und kurzgefasste Suchergebnisse anzuschauen. Dateien kann man nicht herunterladen, da diese in der Regel zu groß sind für mobile Geräte. In der nahen Zukunft ist zu erwarten, dass auch download-fähige mobile Geräte unterstützt werden können.

2.5.3 Fazit

Der direkte Zugriff per JXTA auf das Netzwerk wurde plangemäß umgesetzt.

Es wurden Erkenntnisse gewonnen, die zu neuartigen Anwendungen führen können. Vor allem scheinen Innovationen auf folgenden Gebieten vielversprechend zu sein:

- Bei der kollaborativen Netzwerk Arbeit (z.B. im Grid), kann mittels des P2P-Ansatzes dem Nutzer eine Privatsphäre vorbehalten werden. Diese Eigenschaft ist bei kollaborativen Werkzeugen sehr wünschenswert.
- Im mobilen Bereich müssen die im Projekt erzielten Ergebnisse untersucht werden. Durch die erhöhte Leistung mobiler Geräte können die hier entwickelten Verfahren in der Praxis eingesetzt und in sinnvolle Anwendungen eingebaut werden. Man kann z.B. selber den mobilen Zugang zum eigenen Büro PC einrichten, um so von unterwegs an Informationen zu gelangen, die im Büro liegen.
- Im Bereich des persönlichen Wissensmanagement bietet es sich an, einen Consumer

¹⁶ ZDNET Journal, 23.4.2003, http://zdnet.com.com/2100-1103_2-997981.html

Client einzusetzen.

2.6 AP 6 – Betrieb

2.6.1 Zielsetzung

Ziele dieses Arbeitspakets waren:

- Eine Beta-Test Phase ab dem zweiten Meilenstein durchzuführen, die einen kleinen Kreis von Nutzern die Möglichkeit geben sollte, das System zu benutzen. Damit sollte auf eine weiter gefasste Veröffentlichung des Netzwerkes vorbereitet werden.
- Eine Pilotphase ab dem dritten Meilenstein durchzuführen, in der das Netzwerk mit voller Funktionalität betrieben werden sollte.
- Für den Betrieb einen kommerziellen Anbieter-Partner zu finden.
- Die Nutzerbetreuung durchzuführen.

Verbesserungen der Software anhand des Nutzer-Feedbacks durchzuführen.

2.6.2 Ergebnisse

Es wurde zusammen mit dem DFN entschieden, nur Wissenschaftler und Studenten in öffentlichen Wissenschaftseinrichtungen Zugang zur Software zu geben. Als Gebiet wurde der deutschsprachige Raum gewählt. Die Inhalte des Netzwerkes sollten auf diese Weise qualitativ hochgradig gehalten werden.

Die Beta-Test Phase lief planmäßig ab Mitte 2003. Es erfolgte ein Ansturm von Nutzeranmeldungen. 20 wurden ausgewählt und nahmen an der Testphase teil. Einige der Anwender haben mit viel Feedback, Änderungswünschen und Vorschlägen sehr zur Verbesserung des Systems beigetragen.

Um die Kommunikation zwischen allen Teilnehmern zu verbessern wurde eine Mailing-Liste für die Nutzer eingerichtet. Es wurde auch ein Bugzilla Bug-Tracking-System eingerichtet, um Fehler und Änderungswünsche zu verfolgen. Diese Mittel wurden allerdings nur mäßig benutzt.

Die an die Beta-Test Phase folgende Pilotphase wurde durch Pressemeldungen des DFN-Vereins und der neofonie angekündigt. Es wurden auch persönliche eMails mit Zugangsdaten (die Software ist Passwort geschützt) an alle 400 Personen verschickt, die sich für den Beta-Test angemeldet hatten. Die Software wurde in der Pilotphase 72 Mal heruntergeladen (23 Linux,

49 Windows).

Als externer Provider wurde das Neurobiologische Institut an der Freien Universität zu Berlin an das Netzwerk gebunden. Der Provider Peer war ein gewöhnlicher Arbeitsplatzrechner einer Wissenschaftlerin. Es sollten 3 Gb an Dateien indexiert werden, die sich als zu viel für einen solchen PC erwiesen. Das Programm konnte 1Gb der Daten – über 1000 Dokumente - bearbeiten. Die Installation eines effektiveren Peer an der Neurobiologie soll noch geschehen¹⁷. Es hat sich hiermit gezeigt, dass Peers Spezial-Konfigurationen beherrschen müssen. Sehr große PDF Dateien benötigen eine besondere Konfiguration des Peers. Bei grossen Datenmengen muss damit gerechnet werden, dass während des initialen Index-Aufbaus nicht nebenher gearbeitet werden kann – die Ressourcen des Rechners werden voll benötigt.

Diese Arbeit war im wesentlichen darauf ausgerichtet, das Netzwerk für Wissenschaftler interessant zu gestalten. Es wurden zusätzliche Zugangsarten entwickelt und zusätzliche Datenquellen angebunden.

Ein Zugang wurde von der neuen Mac OS X Plattform ermöglicht. Diese innovative Plattform genoss während der Ausführungszeit großes Aufsehen und es kamen einige Nachfragen nach einem entsprechenden Peer. Dieser Peer kann ebenfalls auf Anfrage heruntergeladen werden.

Der Zugang über das IPv6 Netzwerk des DFN wurde über eine zusätzliches JXTA-Relay realisiert. In der Abbildung unten ist dargestellt, wie die IPv4 und IPv6 Welten nahtlos miteinander verbunden werden. Das Testen dieses Zugangs aus dem 6-WiN wurde von Mitarbeitern des DFN-JOIN-Projekts - durchgeführt.

¹⁷ Diese Arbeit soll auf freiwilliger Basis durchgeführt werden.

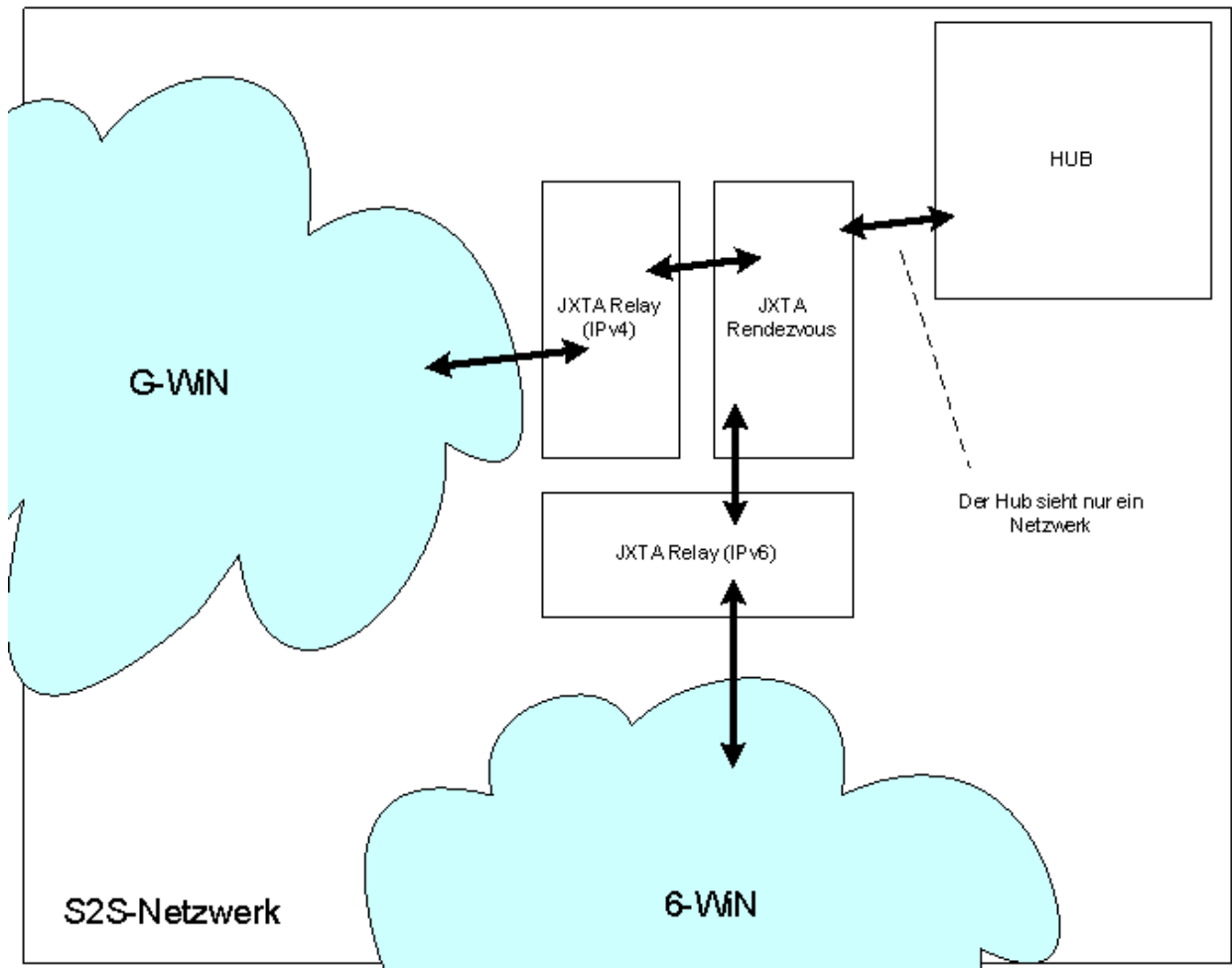


Abbildung 3. 6-WiN und G-WiN

Es wurde ein OAI-Adapter [w6] für Provider Peers entwickelt. Mit einer Konfigurationsänderung ist es demnach möglich, mit einem eigenen Provider Peer auch OAI-Quellen anzubinden. Exemplarisch wurden Citeseer und ein Europäisches Dokumentenverzeichnis angebunden.

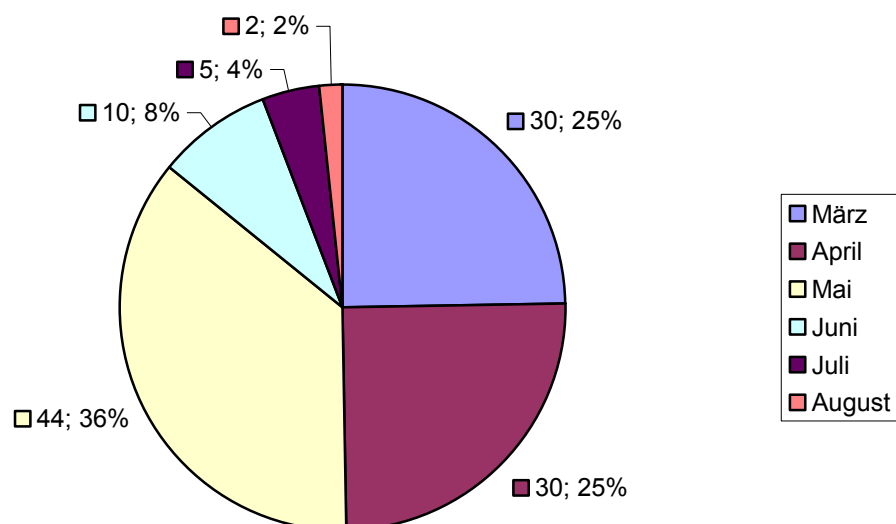
Die digitale Bibliothek des Hochschulbibliotheksentrums des Landes Nordrhein-Westfalen (HBZ-NRW), hatte Ende 2002 eine beträchtliche Summe an Dokumenten (10.000.000 heute sind es 13.000.000), aber keine Suche. S2S wurde prototypisch als Suche eingesetzt. Die web Seiten wurden auch bei Google gemeldet. Die S2S Peers konnten einzeln solch eine Summe nicht bewältigen, aber es wurde gezeigt, dass durch Verteilung des Dokumentenraums das Problem gelöst werden könnte. Exemplarisch hat die neofonie zwei Peers aufgesetzt, die

zusammen über eine Million der Bibliographischen Dokumenten indexierten. Das war mehr als Google beim HBZ-NRW indexiert.

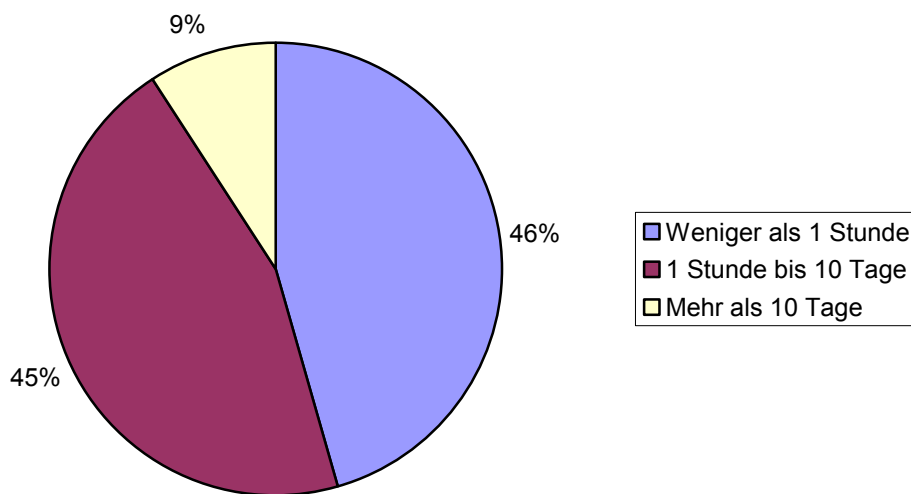
2.6.3 Netzwerk-Nutzung

Die Nutzung des Dienstes während der Pilotphase wurde aufgezeichnet. Eine Auswertung des Nutzungs-Logs folgt.

Anzahl der Erstanmeldungen pro Monat



Peer Aufenthaltszeit im Netzwerk



Zeit im Netzwerk: Avg (Stunden / Tage)

68 2

Zeit im Netzwerk: Top Providers (Stunden / Tage)

1.	neofonie	1320	55
2.	neofonie	957	39
3.	neofonie	833	34
4.	neofonie	785	32
5.	neofonie	756	31
6.	neofonie	582	24

...

10.	neuro	265	11
-----	-------	-----	----

Man sieht, dass fast ausschließlich die Peers die von der neofonie aufgesetzt waren, längerfristig im Netzwerk aktiv waren. Provider von Nutzern waren insgesamt sehr kurz online – ein Verhalten, das schon mehrfach auch in anderen P2P-Netzwerken beobachtet wurde [1,27].

Insgesamt erschienen 121 Peers¹⁸ im Profile Log.

2.6.4 Fazit

Benutzer sind grundsätzlich gegenüber P2P-Software misstrauisch gestimmt. Einerseits hat diese Technologie einen schlechten Ruf wegen Rechtsverletzungen beim illegalen Musikausch, andererseits besteht wegen der Möglichkeit, von außen auf lokale Rechner zuzugreifen, ein technologisches Sicherheitsrisiko. Während der Projektlaufzeit wurden mit den Sasser, MyDoom und w32.blast Viren (bzw. Würmer) entwickelt, die sich nicht nur per eMail verbreiteten, sondern auch Sicherheitslücken in anderen Diensten ausnutzen. Dies steigerte das Misstrauen der Nutzer.

Die S2S Software basiert auf JXTA und nimmt nur Suchanfragen entgegen und gilt deshalb noch¹⁹ als technologisch sicher. S2S ist aber organisatorisch problematisch. Dokumente, die im Intranet liegen, können prinzipiell von außen heruntergeladen werden, wenn der S2S-Nutzer sie sammeln und indexieren lässt. Der Nutzer muss dann die Download-Funktion ausdrücklich unterbinden.

Die Erwartungen waren sehr hoch. So wollte z.B. das HBZ-NRW 10 Million Records indexieren. Eine solche Leistung kann man von kostenloser Software (noch) nicht erwarten.

¹⁸ Für jeden Installationsvorgang wird eine neue Peer ID vergeben, so können Nutzer mehrfach in den Logs vertreten sein.

¹⁹ Es wurden bisher in der JXTA Plattform keine Sicherheitslücken entdeckt. Die JXTA Referenz-Anwendung läuft in der als sicher geltenden Java Virtual Maschine. Direkter Zugriff auf Rechen-Ressourcen sind also nicht möglich.

Auch die 3 Gb an neurobiologischen Daten waren einfach zu viel für das System auf den Desktop PC der Wissenschaftler.

Es hat sich gezeigt, dass es bei Projekten dieser Art notwendig ist, die Presse für die Ergebnisse zu interessieren. Ohne entsprechendes Presse-Echo schenkt auch die Zielgruppe der Wissenschaftler einem solchen Projekt wenig Aufmerksamkeit.

3 Dissemination

3.1 Vorträge

20.11.2001, H.Oertel, DFN Symposium, Berlin, "Peer To Peer 2001 'The next big thing' oder Kalter Kaffee? Forschungsthemen, JXTA, DFN Science to Science"

28.5.2002, R.Wertlen, DFN-Presseseminar "Studieren im Netz", München, „DFN Science-to-Science“

22.9.2002, R.Wertlen, Neurobiologisches Institut FU Berlin, Brükentin, "P2P Seminar with the staff of the department of Neurobiology"

20.5.2003, R. Wertlen, TERENA Konferenz, Zagreb, „DFN S2S: Peer-to-Peer Scientific Research“

12.5.2003, R.Wertlen, <xmlcity:berlin> Kolloquium, FU-Berlin, „DFN Science-To-Science: Suche auf Peer-to-Peer Basis für die Wissenschaftliche Gemeinschaft“

15.9.2003, R. Wertlen, SWITCH Steering Committee Meeting, Maddlingen-Ob-Biel CH, "DFN Science-To-Science"

6.10.2003, R. Wertlen, BMBF Tagung "E-Library & Innovative Dienste", Die Deutsche Bibliothek, Frankfurt am Main, "DFN Science-To-Science: P2P Networking for Research"

14.10.2003, R. Wertlen, BMBF <xml:city> Tage - Wirtschaftsforum, HU zu Berlin, "XML based peer-to-peer networking: DFN Science-to-Science and JXTA Search"

11.11.2003, R. Wertlen, JXTA Workshop des Wirtschafts-Informatik-Instituts Universität Oldenburg, TU zu Berlin, "The DFN Science-To-Science JXTA Application"

20.3.2004, H.Oertel, FUTURE TALK – Innovatives Internet, CeBIT 2004, "DFN Science-To-Science: Kooperatives wissenschaftliches Arbeiten in Kommunikationsnetzen"

3.2 Kontakte und Zusammenarbeiten

Im Rahmen des Projektes wurden folgende Kontakte gemacht²⁰. Untenan werden zu jeder Kontaktart der Zusammenarbeit aufgelistet.

3.2.1 DFN

Zusammen mit dem DFN wurde an der Infrastruktur des Projektes gearbeitet. Der S2S Hub ist direkt am DFN G-WIN geknüpft, um für Wissenschaftler einen Performanz-Vorteil zu schaffen.

3.2.2 HBZ NRW

Mit dem HBZ der Landesbibliothek Nordrhein-Westfalen wurde sehr früh ein möglicher Anwendungspartner gefunden. [Zur Zeit sind mehr Daten vom NRW durch S2S indexiert als von Google.]

3.2.3 FU Berlin, Neurobiologie

Mit der FU Berlin Neurobiologie wurde früh ein zusätzlicher Anwendungspartner gefunden. Bei Vorgesprächen zeigten sich die Biologen eher skeptisch, aber einen Bedarf für solch eine Software hatten die Biologen, da Ihr Archiv keine Suche bietet.

3.2.4 EDUTELLA / Uni. Hannover

Mit dem EDUTELLA Vorhaben wurden auch sehr früh Kontakte geknüpft. Wissen wurde zwischen den beiden Projekte ausgetauscht vor allem im Bereich des P2P-Ansatzes. Die Möglichkeit einer Zusammenarbeit über den Abschluss des Projektes hinaus wird derzeit aktiv angestrebt.

3.2.5 DFN SINN

Ein Wissensaustausch fand mit dem DFN SINN Projekt statt. Die Möglichkeit über die OAI

²⁰ Aufgelistet sind meist Kontakte mit öffentlichen Einrichtungen, DFN S2S wurde auch bei Wirtschaftspartnern der neofonie demonstriert, u.a. auch Daimler Chrysler Services und AOL.de.

Protokolle an Daten aus dem SINN Netzwerk zu kommen, führte dazu, dass eine generische OAI Anbindung von S2S entwickelt wurde.

3.2.6 DFKI

Ein Wissensaustausch mit dem Deutschen Forschungszentrum für Künstliche Intelligenz fand statt. Dabei wurden Metadaten und die automatische Verarbeitung von Daten als Schwerpunkt gesetzt. Die Möglichkeit einer zukünftigen Zusammenarbeit wird derzeit untersucht.

3.2.7 DFN JOIN

Im Rahmen einer Zusammenarbeit wurde eine IPv6 Infrastruktur für das DFN S2S Netzwerk geschaffen. Spezial konfigurierte Peers können also über das IPv6 Netzwerk zusammen mit dem gewöhnlichen S2S Netzwerk verbunden werden. Diese technologische Unabhängigkeit ist ein Merkmal des S2S Software.

3.2.8 ZIB

Im Rahmen des D-GRID Programms wurden Kontakte mit dem Konrad Zuse Zentrum für Informatik in Berlin geknüpft. Ein Wissenstransfer fand statt und eine Absicht für Zusammenarbeit im Rahmen des D-GRID Programms wurde geschaffen.

3.2.9 regioIT Aachen

Im letzten Monat des Projektes fand ein Wissensaustausch mit der regioIT Aachen statt. Es soll ein Wissensnetzwerk für Anwälte auf Basis der S2S Software aufgebaut werden. (s.u.)

3.2.10 Beta-Test + Nutzer

Zusätzlich zu den oben aufgelisteten wurden 20 Beta-Tester (von 400 angemeldeten) in 2003 und während der Pilotphase 70 zusätzliche Nutzer aufgenommen. Die Beta-Tester haben sich teilweise sehr große Mühe mit dem Testen der Software gegeben und ließen Ihre Peers über mehrere Monate laufen.

3.3 Veröffentlichungen

Ronald Wertlen. DFN Science-To-Science: ein DFN Projekt erschließt peer-to-peer Technologie für die Wissenschaft. DFN Mitteilungen 59. Nov. 2002.

Ronald Wertlen. Position Paper: neofonie and DFN Science-To-Science. First International Workshop for Peer-To-Peer Computing (IPTPS'02). unpublished – see <http://s2s.neofonie.de>.

Ronald Wertlen. DFN Science-to-Science: Peer-to-Peer Scientific Research. Proceedings of the 8th Terena Conference in Zagreb. 2003.

Ronald Wertlen. DFN Science-To-Science: wissenschaftliche Suchmaschine geht in den Testbetrieb. DFN Mitteilungen 62. Juni 2003.

Ronald Wertlen. The DFN Science-To-Science JXTA Application. In N. Gronau, A. Benger (Ed.). JXTA Workshop: Potenziale, Konzepte, Anwendungen. 2003.

Ronald Wertlen. Search 'n Share mit DFN Science-To-Science: erste Erfahrungen mit Peer-to-Peer Technologie für die Wissenschaft. DFN Mitteilungen 65. Juni 2004.

4 Quellenverzeichnis

- [1] Eytan Adar and Bernardo A. Huberman. Free Riding on Gnutella. <http://citeseer.nj.nec.com/465529.html>
- [2] A.-L. Barabasi and R. Albert, Science 286, p. 509 (1999).
- [3] Brian Cooper, Hector Garcia-Molina. Studying Search Networks with SIL.
- [4] Dornfest R. and Brickley, D. Metadata. (In Peer to Peer: Harnessing the power of disruptive technologies.) Ed. Oram, A., O'Reilly & Associates (2001) 191-202.
- [5] Rael Dornfest. JXTA takes its position. O'Reilly Open P2P [Online], Apr 2001.
- [6] Ian Foster. Peer-to-peer architecture case study: Gnutella network. University of Chicago, Dec 2001. Cited 11 March 2002. <http://www.cs.uchicago.edu/files/trauthentic/TR200126.pdf>.
- [7] Prasanna Ganesan, Qixiang Sun, Hector Garcia-Molina. YAPPERS: A Peer-to-Peer Lookup Service over Arbitrary Topology. IEEE INFOCOM 2003
- [8] Guha, R. Open Rating Systems. <http://tap.stanford.edu/wot.pdf>
- [9] Emir Halepovic, Ralph Deters. Building a P2P Forum System with JXTA. <http://citeseer.nj.nec.com/halepovic02building.html>
- [10] M. Harren, J. M. Hellerstein, R. Huebsch, B. T. Loo, S. Shenker, and I. Stoica, Complex queries in DHT-based peer-to-peer networks, in The 1st International Workshop on Peer-to-Peer Systems (IPTPS), Cambridge, MA, March 2002.
- [11] Sam Joseph. Distributed Metadata Strategies, is an excellent introduction to the topic of metadata in a P2P search environment. http://www.neurogrid.net/Decentralized_Meta-Data_Strategies-neat.html
- [12] Jan Mischke and Burkhard Stiller. Rich and Scalable Peer-to-Peer Search with SHARK. <http://citeseer.nj.nec.com/565289.html>
- [13] Nejdli, W., Wolf, B., Qu, C., Decker, S., Sintek, M., Naeve, A., Nilsson, M., Palmer, M. and Risch, T.: Edutella: A P2P Networking Infrastructure Based on RDF (2001) <http://edutella.jxta.org/>
- [14] Sylvia Ratnasamy, Paul Francis, Mark Handley, Richard Karp, Scott Shenker. A Scalable Content Addressable Network. <http://citeseer.nj.nec.com/460728.html>
- [15] Matei Ripeanu Adriana Iamnitchi and Ian Foster. Locating data in small-world peer-to-peer scientific collaborations. Electronic Proceedings for the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02), Mar 2002. <http://www.cs.rice.edu/Conferences/IPTPS02/172.pdf>.
- [16] Page, L., Brin, S., Motwani, R., and Winograd, T.: The PageRank citation ranking: Bringing order to the web. Technical report, Stanford University, (1998)
- [17] Plaxton, C.G., Rajaraman, R., Richa, A.W.: Accessing nearby copies of replicated objects in a distributed environment. In Proceedings of ACM SPAA. ACM, June (1997).
- [18] Robertson, S.E. and Sparck-Jones, K.: Relevance Weighting of Search Terms. Journal of the

American Society for Information Science, (1976) 129-146.

[19] A. Rowstron and P. Druschel. Pastry: Scalable, distributed object location and routing for large-scale peer-to-peer systems. In Proc. IFIP/ACMMiddleware 2001, Heidelberg, Germany, Nov. 2001.

[20] S. Saroiu, K. Gummadi, and S. Gribble. A measurement study of peer-to-peer file sharing systems. In MMCN, Jan. 2002.

[21] Mario Schlosser, Michael Sintek, Stefan Decker, and Wolfgang Nejdl. HyperCuP—Hypercubes, Ontologies and Efficient Search on P2P Networks. In International Workshop on Agents and Peer-to-Peer Computing, Bologna, Italy, July 2002.

[22] Sepandar D. Kamvar, Mario T. Schlosser, Hector Garcia-Molina. The EigenTrust Algorithm for Reputation Management in P2P Networks (2003). <http://citeseer.nj.nec.com/551866.html>

[23] C. Tang, Z. Xu, and M. Mahalingam, pSearch: Information retrieval in structured overlays, SIGCOMM Computer Communication Review, vol. 33, no. 1, January 2003.

[24] Chris Sherman. Help LookSmart Crawl the Web.
<http://searchenginewatch.com/searchday/article.php/2177021>

[25] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan. Chord: A scalable peer-to-peer lookup service for Internet applications. In Proc. ACM SIGCOMM 01, San Diego, CA, Aug. 2001.

[26] Bernard Traversat, Ahkil Arora, Mohamed Abdelaziz, Mike Duigou, Carl Haywood, Jean-Christophe Hugly, Eric Pouyoul, Bill Yeager. Project JXTA 2.0 Super-Peer Virtual Network.
<http://www.jxta.org/project/www/docs/JXTA2.0protocols1.pdf>

[27] Wilcox-O'Hearn B. (2002) Experiences Deploying a Large-Scale Emergent Network, in Electronic Proceedings for the 1st International Workshop on Peer-to-Peer Systems (IPTPS '02). <http://www.cs.rice.edu/Conferences/IPTPS02>

[28] ZIB, Florian Schintke, Thorsten Schütt, Alexander Reinefeld. A Framework for Self-Optimizing Grids Using P2P Components. Proceedings of the 14th International Workshop on Database and Expert Systems Applications (DEXA'03). 2003

Bänder

[b1] Albert-Laszlo Barabasi; Linked: How Everything Is Connected to Everything Else and What It Means; Perseus Books Group; 2002.

[b2] Andy Oram; Peer-to-Peer : Harnessing the Power of Disruptive Technologies; O'Reilly; 2001.

[b3] Daniel Brookshier, Darren Govoni, Navaneeth Krishnan, Juan Carlos Soto; JXTA: Java P2P Programming; Sams; 2002.

[b4] Sing Li; Early Adopter JXTA: Peer-to-Peer Computing with Java; Peer Information Inc.; 2001.

[b5] Brendon J. Wilson ; JXTA ; New Riders Publishing; 2002.

[b6] Detlef Schoder, Kai Fischbach, Rene Teichmann; Peer-to- Peer: Ökonomische, technologische und juristische Perspektiven; Springer, Berlin; 2002.

[b7] John W. Muchow; CORE J2ME Technology & MIDP; Prentice Hall PTR; 2001.

[b8] Deepak Alur, et al; Core J2EE Patterns: Best Practices and Design Strategies, Second Edition; 2003

Online Quellen

- [w1] JXTA Search Home Page <http://search.jxta.org/>
- [w2] neofonie Home Page <http://www.neofonie.de/>
- [w3] Project JXTA Home Page <http://www.jxta.org/>
- [w4] The Edutella Project. <http://edutella.jxta.org/>, 2002.
- [w5] Project Neurogrid Home Page <http://www.neurogrid.net/>
- [w6] Open Archives Initiative Home Page <http://www.openarchives.org/>
- [w7] Project Dublin Core Home Page <http://dublincore.org/>
- [w8] JXTA Security Home Page <http://security.jxta.org/>
- [w9] neofonie search robot documentation.
http://www.neofonie.de/loesungen/search/robot_dokumentation.csp
- [w10] neofonie search engine documentation.
http://www.neofonie.de/loesungen/search/engine_dokumentation.csp
- [w11] neofonie content delivery documentation.
<http://www.neofonie.de/loesungen/content/delivery.csp>