

Seminararbeit

YaCy - eine P2P-basierte Suchmaschine



Vorgelegt von:
Daniel Rebei

März 2007

Inhaltsverzeichnis

1	Einleitung	4
2	Komponenten	5
2.1	Rollenverteilung	5
2.2	Information Provider	6
2.3	Indexer	7
2.4	Datenbank	8
2.5	P2P Netzwerk	9
2.6	Oberfläche	9
3	Vor- und Nachteile	12
4	Aktueller Stand	14
5	Fazit	15

1 Einleitung

YaCy¹ wurde 2003 von Michael Peter Christen mit dem Ziel entwickelt eine freie, unabhängige, nicht zensierbare und auf dem P2P-Prinzip basierte Suchmaschine zu erstellen. Zu dieser Zeit existierten schon mehrere ausgereifte P2P-Techniken, sowie Crawler und Indexierer, die das Web durchsuchten. Jedoch gab es noch keine Technik, die diese beiden Welten miteinander verband.

YaCy wurde in Java geschrieben um die plattformunabhängigkeit der Programmiersprache zu nutzen und um die Probleme beim Portieren zu minimieren. Durch die Veröffentlichung des Sourcecodes unter der General Public License (GPL) steht es jedem Freiwilligen frei sich an dem Projekt zu beteiligen und Ideen, sowie Verbesserungen am System einzubringen. Durch die Veröffentlichung des Quellcodes besteht zudem die Möglichkeit der Erreichung einer großen Community von Entwicklern. Zu beachten ist jedoch, dass es sich bei YaCy weder um ein Portal, noch um eine Portalsoftware handelt sondern es weiterhin nur eine Suchmaschine auf P2P-Basis darstellt.

Die Untersuchung des Systems umfasst die unterschiedlichen eingesetzten Komponenten sowie die Vor- und Nachteile. Dabei werden die Daten-Beschaffung, -Speicherung und -Verteilung sowie das zugrundeliegende P2P-Netzwerk erläutert. Durch die fehlenden wissenschaftlichen Veröffentlichungen mussten jedoch viele Details mittels Diskussion mit den Entwickler erörtert und in dieser Arbeit zusammengefasst werden. Zusätzlich wurden einige Vorträge von der dritten SuMa²[3] und dem Linuxtag 2006 [4] sowie des 22. Chaos Computer Congress (CCC) [2] verwendet.

Für die Zukunft ist jedoch ein Buch geplant, dass die unterschiedlichen Komponenten in aller Tiefe beschreibt und die jeweiligen Codeabschnitte erklärt und dem Leser nahe bringen soll.

¹Yet another CyberSpace

²SuMa-eV ist ein gemeinnütziger Verein zur Förderung der Suchmaschinen-Technologie und des freien Wissenszugangs

2 Komponenten

Da es sich bei YaCy um eine Kombination des P2P-Prinzips und einer Suchmaschine handelt werden aus beiden Bereichen Ansätze verwendet um Daten zu besorgen, zu speichern und zu verteilen. So wird die grundlegende Funktion des Crawlers einer Suchmaschine mit der P2P-basierten Daten-Distribution und -Suche kombiniert.

Der Datenfluss lässt sich dahingehend differenzieren, als dass zuerst die Daten besorgt werden müssen (Information Provider), diese dann auf den Inhalt hin überprüft werden (Indexer), um anschließend in einer Datenbank gespeichert zu werden. Diese Datenbank hält alle Informationen bereit, die zur Suche und Datenhaltung nötig sind. Das P2P-Netzwerk speist unter anderem den Information Provider sowie die Datenbank und interagiert mit der Suche mittels des YaCy-Protokolls. All diese Vorgänge sind in Abbildung 2.1 schematisch abgebildet und werden in diesem Kapitel erklärt. Zusätzlich lassen sich noch weitere Informationen aus der Veröffentlichung in der Datenschleuder[1] beziehen.

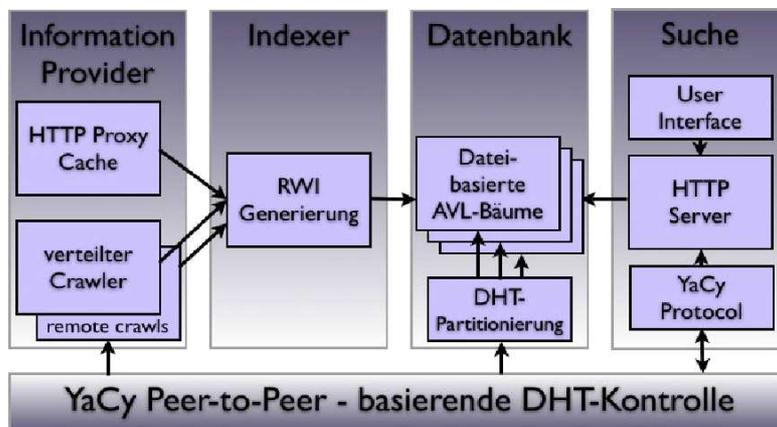


Abbildung 2.1: Schematische Darstellung der Komponenten und deren Zusammenspiel

2.1 Rollenverteilung

Die strikte Rollenverteilung und deren Unterscheidung spiegelt die Funktionen und Möglichkeiten eines Peers wieder. Dabei ist entscheidend ob ein Peer sich hinter einer

Firewall befindet oder nicht und somit, ob ihm Index-Daten aus dem DHT zugewiesen werden können oder nicht.

Virgin-Peer: Unter einem Virgin-Peer versteht man einen Peer, der keinen Kontakt zum Netzwerk hat und somit keinerlei Informationen dazu beitragen kann und auch keine Anfragen stellen kann.

Junior-Peer: Ein Junior-Peer hingegen hat eine Verbindung zu dem Yacy-Netzwerk, befindet sich jedoch hinter einer Firewall, sodass keine Verbindung aus dem Internet zu dem Peer aufgebaut werden kann um Suchanfragen zu starten und Daten zu speichern.

Senior-Peer: besteht eine Verbindung zum Internet und lässt sich der Peer dadurch direkt von anderen Peers erreichen, so wird dieser als ein Senior-Peer bezeichnet. Dies ist die angestrebte Rangstufe, da dieser Peer Suchanfragen entgegen nehmen kann und auch Informationen zum Netzwerk beitragen kann.

Principal-Peer: Ein Principal-Peer ist ein Senior-Peer, der zudem die Möglichkeit bereithält, Daten auf einen FTP¹-Server zu legen, so dass andere Peers die keinen Zugang zu Netzwerk erlangen vom FTP Server die hochgeladenen Daten beziehen können und somit Suchanfragen auf diese Daten stellen können.

2.2 Information Provider

Um eine Suchmaschine betreiben zu können, müssen Webseiten durchsucht und Daten beschafft werden, auf denen dann der Indexer Analysen durchführen kann. Dazu werden zwei unterschiedliche Methoden verwendet.

Die eine besteht darin eine Anfrage direkt zu starten (**local crawl**²) und somit eine Webseite, bzw. deren Links bis zu einer bestimmten Tiefe hin zu crawlen. Ein Crawler ist ein Programm, das Webseiten rekursiv aufruft und deren Hyperlinks ebenfalls in rekursiver Weise auswertet. Die Methode des lokal getriggerten Crawls kann bei einer großen Menge an Links zu einem verteilten Crawl führen, der dann auf andere Peers verteilt wird.

Die zweite Methode besteht darin, YaCy als einen **Proxy**³ zu verwenden um über diesen an Daten zu gelangen. Dabei wird YaCy als Proxy in den Browser eingetragen und dieser startet bei jedem Seitenaufruf einen Crawl, der nur die besuchten Webseiten indiziert. Dies bietet den Vorteil, dass nicht jedes mal eine neue Suche von Hand ausgelöst werden muss, sondern dass sich der Datenbestand mit jeder aufgerufenen Seite erhöht oder aktualisiert (on-the-fly Aktualisierung). Durch die Verwendung eines Proxy besteht zudem die Möglichkeit innerhalb eines Heim- oder Büronetzwerkes die Seiten vorzuhalten (prefetching) um somit die Zugriffszeit zu minimieren. Zudem besteht die Möglichkeit

¹File Transfer Protocol

²aus dem engl. durchwühlen, krabbeln, kraulen

³aus dem engl. proxy = Stellvertreter, Vertreter

Filter auszuwählen, die die vorgehaltenen Daten durchsuchen und unerwünschte Inhalte ausblenden. Dabei können ganze Seiten unzugänglich gemacht werden, wie auch nur Teile geblockt werden, wie das beispielsweise bei Werbefilter geschieht.

2.3 Indexer

Unter einem Indexer wird in diesem Fall ein Datenaufbereiter verstanden, der die vom Information Provider gesammelten Daten (URLs, Wörter, Sätze, ...) auf den Inhalt hin überprüft und diese in einem Reverse Word Index (RWI) zuführt. Die RWIs werden in gehashter Form in der Datenbank gespeichert was dazu führt, dass die Informationen nicht im Klartext hinterlegt sind und dadurch die Sicherheit der Indexhalter erhöhen, da es somit nicht möglich ist, darauf zu schließen wer die Daten erstellt hat.

Abbildung 2.2 zeigt den Ablauf der Indexierung, sowie die Datenflüsse im Netzwerk. Dabei wird das Netzwerk in das Backbone⁴, spezialisierte Peers und den persönlichen Peer unterteilt. Dabei stellt der persönliche Peer eine Anfrage und kann die Antworten entweder von spezialisierten Peers oder aus dem P2P-Netz (Backbone) erhalten.

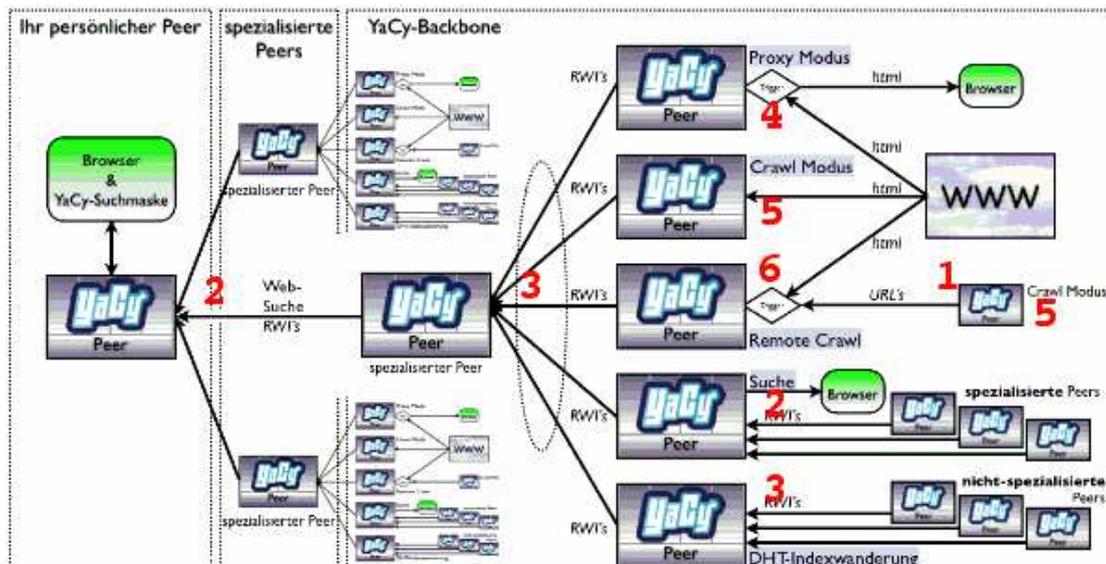


Abbildung 2.2: Datenfluss, der durch die Index Erwerbung über das Internet entsteht

Das Backbone kann mittels sechs unterschiedlicher Varianten Daten erhalten, weiterleiten, anfragen und Anfragen beantworten.

1. Eine Remote Crawl Rückmeldungen ist eine Liste von Indexe und Hashwerten, bei der der Peer den Crawl initiiert hatte, jedoch von anderen Peers gecrawlt wurden.

⁴aus dem engl. backbone = Basisnetz

2. Dieser Index Transfer wurde von anderen Peers gecrawlt und nun dem eigenem wieder zugestellt.
3. Die URL-Indexierung wurde von anderen Peers initiiert und durchgeführt. Danach wurden die Indexe und Links dem eigenen Peer mittels Indexwanderung zugeführt, da der eigene Peer die optimale Position in der Distributed Hash Table (DHT) für die Indexe besitzt.
4. Diese Indexierung wurde durch die Verwendung des Proxies auf Seiten des Peerbetreibers ausgeführt. Dabei werden aber **keine persönlichen oder geschützten Daten** indexiert.
5. Wird mittels YaCy eine Indexierung von Hand gestartet, so wird dies als lokal erzeugter Index bezeichnet.
6. Dies ist ein remote crawl, der von einem anderen Peer getriggert wurde und auf dem eigenen Peer ausgeführt wird. Dies stellt somit das Gegenstück zu Punkt eins dar.

2.4 Datenbank

Die Datenbank speichert alle indexierten Daten, die durch den Indexer und dem P2P-Netzwerk hinzugefügt werden. Dabei werden jedem Peer die Daten passend zu seiner Stelle im DHT zugewiesen und mittels Indexwanderung zugestellt. Das Durchsuchen der Datenbank erfolgt mittels des eingebauten Web-Oberfläche, die ähnlich zu den üblichen Suchmaschinen gestaltet ist (siehe Abbildung 2.3).



Abbildung 2.3: Suche in der YaCy Oberfläche mit zwei Treffern

Die Struktur der Datenbank wird in balancierte, binäre Suchbäume (AVL-Bäume) abgebildet, um eine logarithmische Suchzeit zur Anzahl der Elemente in dem Baum zu garantieren. Der Name AVL leitet sich dabei von den Erfindern Adelson-Velsky und Landis ab, die diese Datenstruktur für ausgeglichene Bäume 1962 entwickelten. Eine ausführliche Beschreibung von AVL Bäumen lässt sich in der Arbeit von Qiang Wang [5] finden.

Einige der Methoden die durch die Datenstruktur⁵ verwendet werden, sind umfassen den `URL-Hash`, die `Qualität`, das `Alter` und wann es zuletzt verändert wurde. Zudem werden noch der `Hitcount` und die `Positionen` innerhalb des Textes und des Satzes gespeichert, sowie die `Wörter` und die `Sätze` im Text. Die Datenstruktur bietet auch Informationen über die Position innerhalb des DHT, den `Worddistanzen` sowie `min/max RWI-Einträgen`.

2.5 P2P Netzwerk

Die Kontrolle der Peers wird mittels des P2P-Prinzips vollzogen, wobei dafür eine DHT-Struktur benutzt wird um eine optimale Indexverteilung zu gewährleisten. Dabei werden die Hosts anhand von individuellen Keys in der DHT an die richtige Stelle platziert. Dazu wird ein zufälliger Schlüssel mittels `final String key = crypt.randomSalt();` generiert, der daraufhin erst ein mal zu einem `SeedString` umgewandelt wird und dann in die `SeedDB` eingefügt wird (`yacyCore.seedDB.mySeed.genSeedStr(key)`).

Dabei werden RWIs innerhalb der Distributed Hash Table so organisiert, dass die Indexe mittels eines Wanderungskonzeptes immer an der optimalen Stelle im DHT vorhanden sind. Zudem können Crawls an anderen Peers abgegeben, oder von denen empfangen werden um einen Lastausgleich zu garantieren falls diese nichts zu tun haben. Zwei weiterer Mechanismen sind die Überprüfung der Eindeutigkeit der vorhandenen URLs im DHT einerseits, sowie die Pflege von Kontaktdaten aller Peers im Netzwerk, andererseits. Die Kontaktdaten werden dazu benutzt, dass eine eindeutige Zuordnung zwischen IP-Nummer und der `<peername>.yacy` Domain gegeben ist. Dies bietet bei Benutzung eines Proxies die Möglichkeit den Namen eines mittels DNS-Veränderung gesperrten Peers zu erreichen, ohne explizit die IP-Nummer zu kennen.

2.6 Oberfläche

Die Oberfläche mittels der die Funktionalität von YaCy gesteuert wird ist eine Webseite, die über die URL `http://localhost:8080` erreicht werden kann, falls YaCy lokal installiert wurde. Nach der Eingabe des Administratorpasswortes, kann der Benutzer den Peer verwalten und die Suchfunktion benutzen. Die Abbildungen 2.4 (a-d) zeigen die hauptsächlichen Überfunktionen die dem Benutzer zur Konfiguration bereit stehen. Elemente die mit einem Schloss versehen sind, wie es beispielsweise fast alle in 2.4 (b) sind, sind nur für den Administrator zugänglich und erfordern die Eingabe des Passwortes.

⁵siehe `src/yacy/source/de/anomic/index/indexRWIEntry.java` im Quellcode

Globaler Index	Crawler Steuerung	Kommunikation / Publikation	Peer-Verwaltung
Suchseite	Index erzeugen	Home Page	Status & Shutdown
Detailsuche	Index verwalten	Dateifreigabe	Grund-einstellungen
Surftipps	Index Monitor	Wiki	erweiterte Einstellungen
Lesezeichen	Crawler Monitor	Blog	Netzwerk
Hilfe	URL Blacklist	Nachrichten	News System
			Log
			Performance
			Verbindungen

Abbildung 2.4: GUI: (a) Globaler Index zur Suche und für alle zugänglich. (b) Crawler Steuerung um Indexe hinzuzufügen. (c) Kommunikation & Publikation zur Benutzung von Wiki, Blog & co. (d) Peerverwaltung

Globaler Index: Dient zur Suche und als Startpunkt für die Benutzung für YaCy.

Crawler Steuerung: Bietet die Möglichkeit Indexe zu erzeugen und zu verwalten. In der Index Verwaltung können zu bestimmten Suchwörtern die Hashes (Abb. 2.6) sowie die hinterlegten Informationen (Abb. 2.7) angezeigt werden.

Kommunikation & Publikation Mittels dieser Menüpunkte lassen sich die integrierte Webseite, das Wiki⁶, das Blog⁷ und die interne Nachrichtenversendung benutzen und konfigurieren.

Peer-Verwaltung Die Peer-Verwaltung bietet Zugriff auf den Status und gibt die Möglichkeit die grundlegenden Funktionen, wie beispielsweise Sprache, Aussehen und Passwörter fest zu legen. Des weiteren lassen sich noch das Netzwerk, sowie die Performance und der Proxy einstellen (Abb. 2.5).

Server Zugangs-Einstellungen
Proxy Zugangs-Einstellungen
Inhalt Parser Einstellungen
Crawler Einstellungen
HTTP Netzwerk
Remote Proxy (optional)
Port Weiterleitung (optional)
Seed Upload Einstellungen
Nachrichten Weiterleitung (optional)

Abbildung 2.5: Erweiterte Einstellungen

⁶Ein Wiki ist eine im World Wide Web verfügbare Seitensammlung, die von den Benutzern nicht nur gelesen, sondern auch online geändert werden kann.

⁷Ein Weblog (engl. Wortkreuzung aus Web und Log), häufig abgekürzt als Blog, ist ein digitales Tagebuch.

Wort:	<u>freiburg</u>	URL Einträge für Wort zeigen
Wort-Hash:	nINVGjFjGNaz	Liste erstellen
		URL Einträge für Wort-Hash zeigen
		An anderen Peer senden
		nX4oXc1zx8FN jarjar
URL:		Details für URL zeigen
URL-Hash:		Liste erstellen
		Details für URL-Hash zeigen

487 URL entries related to this word hash nINVGjFjGNaz

- y_4V5xcYfvEB <http://cone.informatik.uni-freiburg.de/>, pos=7
- OfkfqxcYfvEB <http://cone.informatik.uni-freiburg.de/activities-e.html>, pos=5
- RJFqixcYfvEB <http://cone.informatik.uni-freiburg.de/activities-g.html>, pos=7

Abbildung 2.6: Suche nach nach den Hashwerten die für das Wort 'freiburg' hinterlegt sind

URL String	<u>http://cone.informatik.uni-freiburg.de/</u>
Hash	<u>y_4V5xcYfvEB</u>
Description	Rechnernetze und Telematik, Albert-Ludwigs-UniversitA euro t Freiburg
Modified-Date	Thu Mar 01 01:00:00 CET 2007
Loaded-Date	Thu Mar 01 01:00:00 CET 2007
Referrer	
Doctype	h
Language	de
Size	6583
Words	743

Abbildung 2.7: Informationen die für den Hash 'y_4V5xcYfvEB' hinterlegt sind

3 Vor- und Nachteile

Bisher wurden nur die Komponenten des Systems sowie einige technische Spezifikationen erläutert, es wurde jedoch noch kein Bezug auf die Vor- und Nachteile des Systems genommen.

Die besonderen **Vorteile** von YaCy sind:

- Durch die dezentrale Struktur des P2P-Ansatzes besteht eine praktisch ausfallsichere Infrastruktur.
- Es können keinerlei Zensuren von einer zentralen Stelle vorgenommen, da jeder Peer-Betreiber beliebige Webseiten indexieren kann und somit persönliche Vorlieben in das Netz einfügen kann.
- Durch die Benutzung eines Proxies besteht eine hohe Aktualität des Indexes, da oft besuchte Seiten auch oft aktualisiert werden.
- Dadurch, dass Benutzer eigene Seiten hinzufügen können, besteht die Möglichkeit Seiten zu indexieren, die nicht durch Linkverweise aus dem öffentlich zugänglichen Web erreichbar sind, sondern nur dadurch dass jemand die spezifische Adresse kennt (deep-web).
- Das Projekt basiert auf einer plattformunabhängigen Programmiersprache (SUN Java¹) und bietet somit Benutzern vieler Betriebssysteme (Windows, Linux, MacOS) die Möglichkeit das System zu benutzen.
- Durch die Open Source Lizenz der General Public Licens (GPL)² besteht für jeden die Möglichkeit sich den Sourcecode zu besorgen und Veränderungen daran vor zu nehmen. Zudem bietet die Offenlegung des Codes die Möglichkeit Fehler und Sicherheitslücken schneller zu finden und zu beseitigen.
- Selbst durch ein kleines Netzwerk von ca 60-80 Peers ist es zur Zeit schon möglich 200-300 Millionen Links bereit zu stellen.

¹<http://www.java.com/>

²Quelle: <http://www.gnu.de/documents/gpl.de.html>

Die **Nachteile** des Systems lauten:

- Durch den P2P-Ansatz und somit der Verteilung der Indexe ist eine schnelle Suche nicht zu gewährleisten. Es werden im Moment Suchzeiten von 3-4 Sekunden angenommen, wohingegen Google³ Suchzeiten von ca. 0,2 bis 0,5 Sekunden aufweist.
- Momentan sind noch zu wenige Peers vorhanden⁴ um eine große Anzahl von durchsuchbaren Links anbieten zu können (siehe dazu auch Abbildung 4.1).
- Dadurch kann es passieren, dass beim Ausfall einiger großer Peers das Netzwerk stark leiden muss und somit eine große Anzahl von Links verloren gehen.
- Durch die noch größtenteils fehlenden Sicherheitsmechanismen lassen sich zur Zeit eine große Menge Angriffsmöglichkeiten finden.

Die angesprochenen Sicherheitsmängel sind den Entwicklern bekannt, wobei im Moment scheinbar noch kein Bedarf dazu besteht diese zu schließen. Dies rührt daher, dass die momentan am Netzwerk teilnehmenden Peers ausschließlich positive Absichten haben und meistens von den Entwicklern betrieben werden. jedoch muss in der Zukunft damit gerechnet werden, dass die Anzahl der Peers steigt und somit bösartige Peers auch daran teilnehmen werden, was dazu führt, dass über ausgefeilte Sicherheitskonzepte nachgedacht werden muss um beispielsweise Manipulationen von Rankinginformationen zu verhindern.

³<http://www.google.com>

⁴60-100 Senior- und Principal-Peers (Stand 15.03.2007)

4 Aktueller Stand

Die Abbildung 4.1¹ zeigt die Netzwerkstatistik des YaCy-Netzwerkes in der Woche vom 8. März 2007 bis 15. März 2007. Dabei ist zu erkennen, dass Schwankungen in der Anzahl der Hosts auch einen Einfluss auf die Anzahl der verfügbaren Links haben. Jedoch ist auch erkennbar, dass die wenigen Peers die aktiv am Netzwerk teilnehmen eine große Anzahl an Links (200-300 Millionen) bereitstellen.

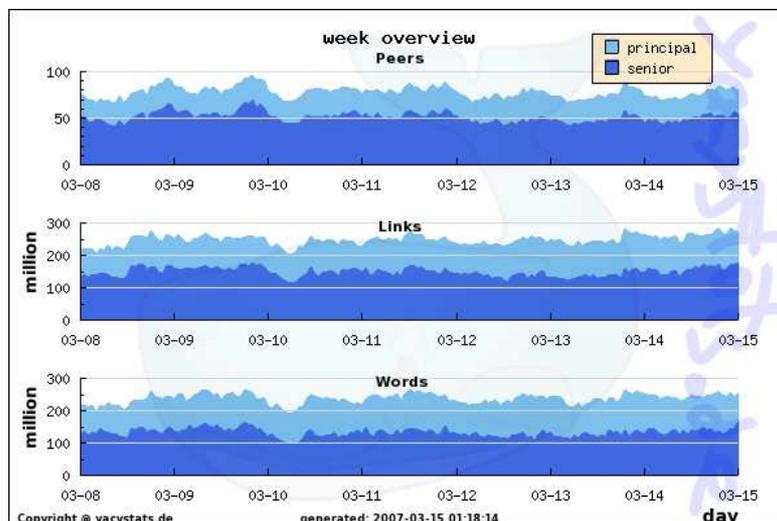


Abbildung 4.1: Netzwerkübersicht der Woche vom 8.3.07 - 15.3.07

YaCy liegt zur Zeit in der Version 0.50 vor (Stand 16.03.2007) und kann von folgenden Seiten bezogen werden:

- <http://www.yacy.net/yacy/Download.html> (Hauptseite)
- <http://www.yacy-websuche.de/index.php/download.html> (Deutsche YaCy Seite)
- http://developer.berlios.de/project/showfiles.php?group_id=3505 (Berlios Mirror)

¹Quelle: <http://www.yacystats.de/network.html>

5 Fazit

Die P2P-basierte Suchmaschine YaCy bietet zur Zeit eine innovative Mischung aus P2P-Prinzip und freier Suchmaschine, wobei der interessanteste Teil darin besteht, dass jeder jederzeit beliebige Seiten indexieren kann. Dies bietet den Vorteil, dass die individuellen Wünsche und Vorlieben der Benutzer und Betreiber in den Suchergebnissen abgebildet werden und somit eine Alternative zu den von Google und anderen Such-Portalen zentralen Modellen vorstellt.

Durch die verhältnismäßig leichte Installation und den minimalen Systemanforderungen eignet sich YaCy besonders gut zum Einsatz auf Heimrechnern und InternetCafes, da das System sowohl eine Suchoberfläche, als auch einen Proxy sowie eine Blacklist-Funktion bietet um Seiten zu sperren die nicht erwünscht sind (bspw. Sex-Seiten). Zudem bietet sich an YaCy innerhalb eines Büronetzwerkes zu benutzen, da es als kostenlose Alternative zu proprietären Suchmaschinen fungieren kann.

Leider fehlt es YaCy zur Zeit an der Anzahl der Benutzer die es effektiv am laufen haben, jedoch wird sich das mittelfristig ändern können, wenn die entsprechenden Sicherheitsmechanismen implementiert wurden und eine Steigerung in der Performance durchgeführt wurde. Doch sind die Zahlen die das System aufweist ein großer Schritt in die richtige Richtung, da sich die Anzahl der Peers innerhalb eines Jahres verdoppelt hatte¹ und einzelne Peers bis zu 15 Millionen Links verwalten.

Andere Arbeiten die sich mit diesem Thema beschäftigen liegen nicht vor, es existieren bisher aber andere interessante Themen, wie z.B. ein P2P-basierter Webcrawler (Grub²) sowie ein spezielles Suchprogramme für Bilder (Eikon³). YaCy scheint zur Zeit einzigartig zu sein im Bereich der P2P-basierten Suchportale.

¹Quelle: http://www.yacystats.de/tmp/networkplots_overviewyear.png (Stand 16.03.2007)

²<http://grub.looksmart.com/>

³<http://eikon.recursion.org/>

Literaturverzeichnis

- [1] *YaCy - Peer-to-Peer Web-Suchmaschine.*
<http://www.yacy.net/yacy/material/YaCy-Datenschleuder086.pdf>
- [2] CHRISTEN, Michael: *Suchmaschinen: In Technik, Wirtschaft und Medienkuns.*
<http://www.yacy.net/yacy/material/YaCy-22C3Speech.pdf>. – 22nd Chaos Computer Congress - Speech paper
- [3] CHRISTEN, Michael: *Suchmaschinen: In Technik, Wirtschaft und Medienkuns.*
<http://www.yacy.net/yacy/material/YaCy-Entwicklungen2005-2006.pdf>.
Version: September 2006
- [4] CHRISTEN, Michael: *YaCy Peer-to-Peer Web-Suche: Informationsfreiheit und freie Information Retrieval Softwar.* Version: 2006.
<http://www.yacy.net/yacy/material/YaCy-PracticalLinuxForumLT2006.pdf>.
– LinuxTag 2006 - Speech paper
- [5] WANG, Qiang: *Der AVL-Baum.* Version: 2002.
<http://www.fmi.uni-passau.de/wangq/Proseminar/Proseminar.pdf>. – Proseminar - Algorithms and Data Structures for Database Systems