

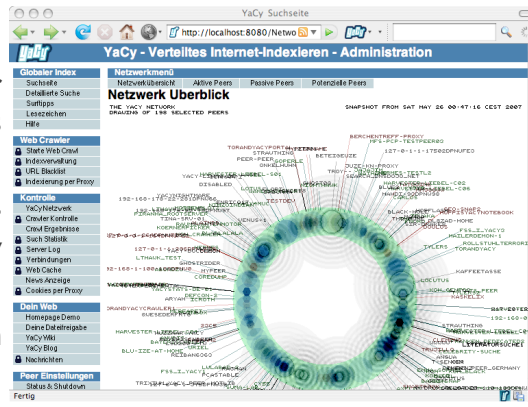
YaCy Besonderheiten

Dezentrales Netz

Im freeworld-YaCy Netz ist jeder Teilnehmer gleichberechtigt, es gibt keinen zentralen Server.

Rechts: Darstellung der angeschlossenen und aktiven YaCy Rechner im freeworld-Netz.

Die pulsierenden Punkte zeigen Crawlende Peers.

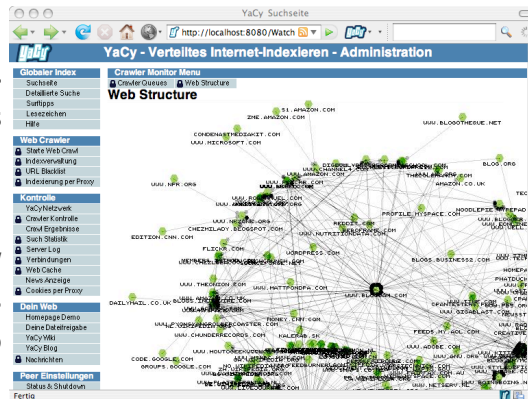


Web-Visualisierung

YaCy kann die durchsuchten Webseiten und deren Verlinkung als Grafik darstellen.

Data Mining

Verlinkungsstrukturen (forward-/backlinks), Linklisten und Domainlisten können als xml über http exportiert werden.



Performance - Leistungssteigerungen gegenüber vorherigen Versionen:

- nahezu keine IO-Last mehr. Man kann YaCy im Hintergrund betreiben.
- Der Crawler erfasst und parst 2000 Web-Seiten pro Minute; der Indexierer verarbeitet bis zu **30.000** Dokumente **pro Minute** (aus XML, Surrogate).
- Bis zu **30** Suchanfragen **pro Sekunde** sind möglich.
- Ein Peer kann **20 Millionen Dokumente** erfassen bei einem RAM-Bedarf von 2GB. Im Peer-to-Peer - Verbund kann eine unbegrenzte Zahl von Dokumenten erfasst werden, dabei skaliert YaCy dynamisch mit neuen Peers.

Download <http://yacy.net>

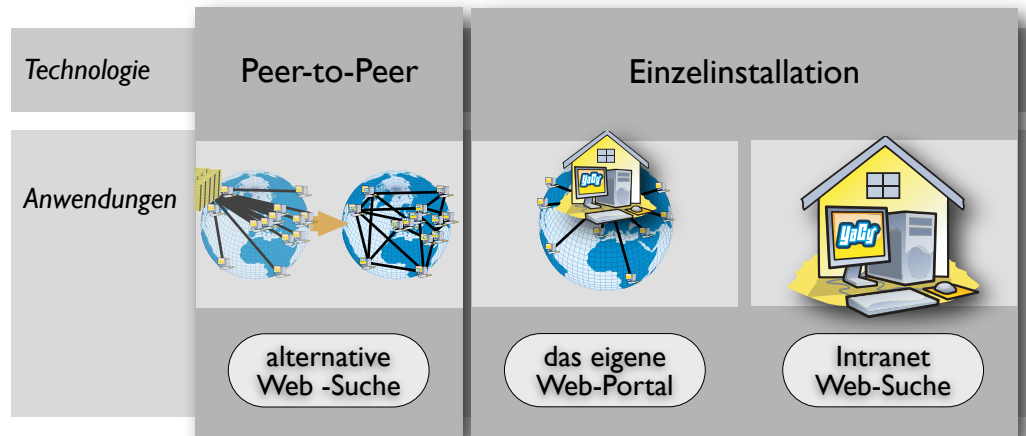
Free Software - Open Source / GPL-Lizenz - Support über forum.yacy.de



Suchmaschinensoftware

- ▶ Unzensurierbare Suche im P2P-Verbund
- ▶ Dein eigenes Suchportal, deine Inhalte
- ▶ Intranet- und Filesystemsuche
- ▶ API zur Integration (xml/json über http)

Wer viele Daten hat oder Daten des World Wide Web indexieren will, braucht eine eigene Suchmaschine. YaCy ist die Software für ein öffentliches oder privates Suchportal bei der *Du* entscheidest was im Suchindex ist. Weil dies dann *Dein* Suchportal ist, ist jede Suche für dich **anonym und unzensurierbar**.



Web-Suche im P2P-Netz

Deine YaCy-Instanz kann sich mit denen der anderen YaCy-Betreiber verbinden. Jede Web-Suche in deinem Peer nutzt dann auch den Web-Index der anderen Peers – und umgekehrt. Diese Funktion ist beim Programmstart des Standard-Releases bereits voreingestellt. Die Installation ist sehr einfach: nur herunterladen, auspacken, starten.

Suchmaschine für deine Daten

Du hast viele Dokumente und willst diese entweder öffentlich anbieten oder im Intranet durchsuchen können:

- Starte einen Web-Crawl für dein eigenes Suchportal. Wenn du ein Forum oder Wiki hast: YaCy bietet spezialisierte Erfassungsfunktionen.
- Im Intranet kannst du das eigene Filesystem oder ein Netzlaufwerk erfassen und über die YaCy-Suchseite für alle im Intranet zugänglich machen.

YaCy Indexierer und Suche

Der Such-Index in YaCy wird durch einen Web/File-Crawl, durch Import von Texten in XML (Surrogate) oder durch direkten Export aus einer SQL Datenbank gefüllt. Parser: doc/ppt/xls/pdf/rss/swf/rtf/odt etc. werden gelesen.

Daten-Eingabe: verschiedene Möglichkeiten um den Such-Index aufzubauen

Web-Crawler

Import aus XML

Dublin Core Metadata (Surrogate)

```
<?xml version="1.0" encoding="utf-8"?>
<!-- YaCy surrogate using dublican core notion -->
<surrogates
  xmlns:dc="http://purl.org/dc/elements/1.1/">

  <record>
    <dc:title><![CDATA[Alan Smithee]]></dc:title>
    <dc:identifier>http://de.wikipedia.org/wiki/Alan_Smithee</dc:identifier>
    <dc:description><![CDATA[Der als Filmregisseur oft genannte ''Alan Smithee'' ist ein Anagramm von „The Alias Men“.]></dc:description>
    <dc:language>de</dc:language>
    <dc:date>2009-04-14T00:00:00Z</dc:date>
    <!-- date is in ISO 8601 -->
  </record>
</surrogates>
```

SQL-Export

- phpBB3
- mediaWiki
- vBulletin
- ...
- ...
- ...
- ...

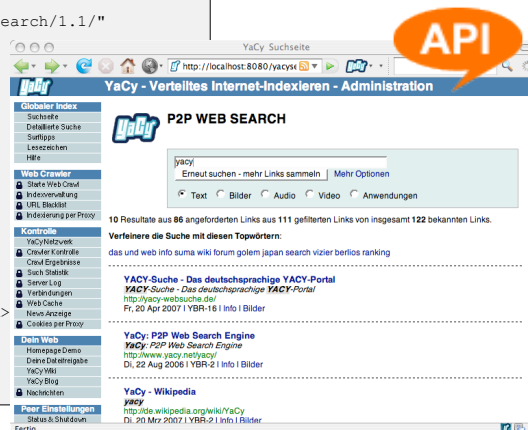
mehr Exporte in Arbeit

Die Suche steht über den in YaCy integrierten Webserver als ‚normale‘ Websuche, als Such-Widget und in Form einer RSS Ausgabe entsprechend opensearch.org zur Verfügung. Die Funktionen des APIs können als XML und

```
> curl http://localhost:8080/yacysearch.rss?query=foaf&maximumRecords=10

<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="/yacysearch.xsl" version="1.0"?>
<rss version="2.0"
  xmlns:yacy="http://www.yacy.net/"
  xmlns:opensearch="http://a9.com/-/spec/opensearch/1.1/"
  <!-- stark verkürztes Beispiel! -->
<item>
  <title>Friend of a Friend (FOAF) project</title>
  <link>http://www.foaf-project.org/</link>
  <pubDate>Fri, 23 May 2008 02:00:00 +0200</pubDate>
</item>
<item>
  <title>FOAF - Wikipedia</title>
  <link>http://de.wikipedia.org/wiki/FOAF</link>
  <pubDate>Tue, 08 Jan 2008 01:00:00 +0100</pubDate>
</item>
<link>http://microformats.org/wiki/xfn-to-foaf</link>
<pubDate>Fri, 09 May 2008 02:00:00 +0200</pubDate>
</item>
</rss>
```

JSON abgefragt und in eigene Programme eingebunden werden.



YaCy Anwendungen

Suchportalbetreiber nutzen YaCy als Content-Lieferant:

- **MetaGer** betreibt YaCy-Peers und feeded Suchresultate in die Metasuchergebnisse – metager.de
- **Fireball** realisiert eine Expertensuche mit YaCy durch den Aufbau von spezialisierten Suchindexen – fireball.de
- Das **Forschungszentrum Karlsruhe** betreibt ein YaCy-Netz aus 30 Rechnern zur Erfassung universitärer und wissenschaftlichen Webseiten weltweit – sciencenet.fzk.de



Integration in Wikis, Foren, Blogs mit einem Such-Widget:

YaCy bietet hierzu vorgefertigte Code-Snippets. Zur Erfassung der Content-Systeme (Blogs, Wikis, Foren) sind spezialisierte Harvester vorhanden. Die YaCy-Suche dient dann als Meta-Suche über die verschiedenen Quellen und bietet dazu spezifische **Navigatoren**.

YaCy Wiki Search ✕

- [De:Erweiterungen - YaCyWiki](#)
<http://www.yacy-websuche.de/wiki/index.php/De:Erweiterungen>
Wed, 29 Apr 2009
- [Re: YaCy-Widget](#)
<http://forum.yacy-websuche.de/viewtopic.php?t=7867>
Sun, 21 Jun 2009
- [Re: API-Webseiten kenntlich machen](#)
<http://forum.yacy-websuche.de/viewtopic.php?t=11731>
Sun, 21 Jun 2009
- [Re: Linuxtag 2009](#)
<http://forum.yacy-websuche.de/viewtopic.php?t=14161>
Sun, 21 Jun 2009

Displaying result 1 to 4
Prev
Next
@

Die **Analysefunktion** zur Verlinkungsstruktur ist für Webdesigner interessant: über das [http-API](http://http-api) können XML-Dokumente mit in- und out-Links zu beliebigen Domains abgefragt werden.

YaCy enthält ausserdem ein **Bookmark-System** mit sharing-Funktionen und entsprechendem XBEL-API.

```
> curl http://localhost:8080/api/webstructure.xml?about=yacy.net

<?xml version="1.0"?>
<webstructure maxhosts="20000">
<references direction="out" count="1" maxref="300">
<domain host="yacy.net" id="FhlyhQ" date="20090618">
  <reference id="VRAHIA" count="5">suma-ev.de</reference>
  <reference id="EMALDQ" count="3">www.kit.edu</reference>
  <reference id="sX4ozA" count="15">liebel.fzk.de</reference>
</domain>
</references>
<references direction="in" count="1">
<domain host="yacy.net" id="FhlyhQ" date="20090618">
  <reference id="a_bYbr" count="32">de.wikipedia.org</reference>
  <reference id="DWDqha" count="1">hwiki.fzk.de</reference>
  <reference id="4JR9RA" count="1">wiki.yacy.de</reference>
  <reference id="wgcWfA" count="1">www.itgrl.de</reference>
  <reference id="P290EA" count="128">www.heise.de</reference>
  <reference id="z4bRCA" count="1">blog.suma-ev.de</reference>
  <reference id="sX4ozA" count="5">liebel.fzk.de</reference>
  <reference id="FXg39Q" count="3">www.yacy.net</reference>
</domain>
</references>
</webstructure>
```