



Freier Wissenszugang mit der Suchmaschine YaCy

Michael Christen, mc@yacy.net, IP-ShareMedia.de

26. Oberhofer Kolloquium 2011

Web 3.0 - wird es das Web der Informationsspezialisten?

Inhalt

Teil 1: **10 Thesen zum freien Wissenszugang**
freie Inhalte, Bürgerrechte, Privatsphäre, Datenschutz

Teil 2: **Suchmaschinenteknik**

Grundfunktionen, Skalierung, dezentrale Architektur

Teil 3: **Freier Wissenszugang mit YaCy**

freie Such-Appliance für alle





- (a) **Wissen** ist Erbe und Besitz der Menschheit und damit **frei**.
- (b) Der **Zugriff** auf Wissen muss **frei** sein.
- (c) **Alle Menschen** haben das Recht auf Kommunikation und Informationsfreiheit.
- (d) Das Recht auf Achtung der **Privatheit ist ein Menschenrecht** und ist unabdingbar für die freie und selbstbestimmte Entfaltung von Menschen in der Wissensgesellschaft.

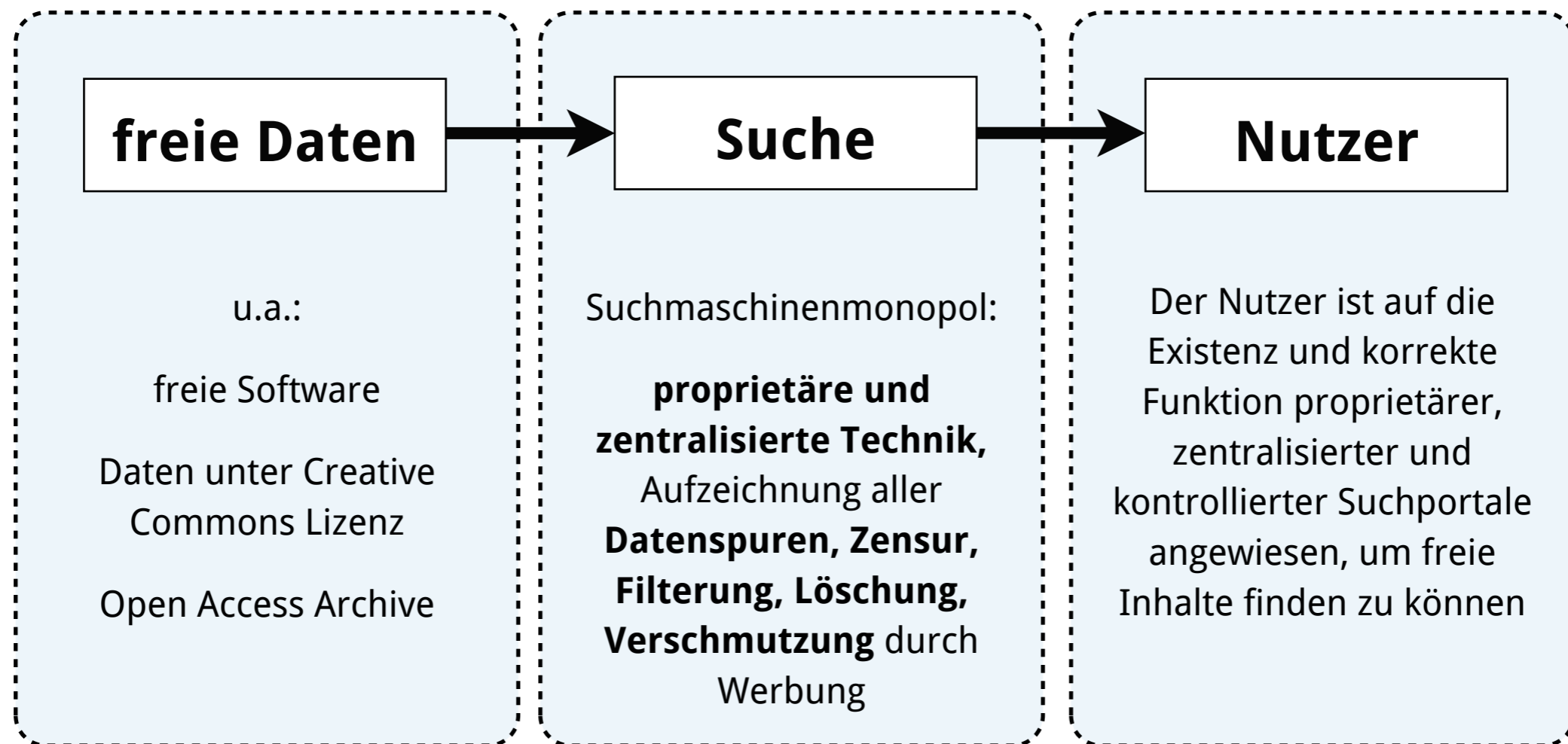
aus: <http://www.worldsummit2003.de/de/web/52.htm>

Vorgehensweise:



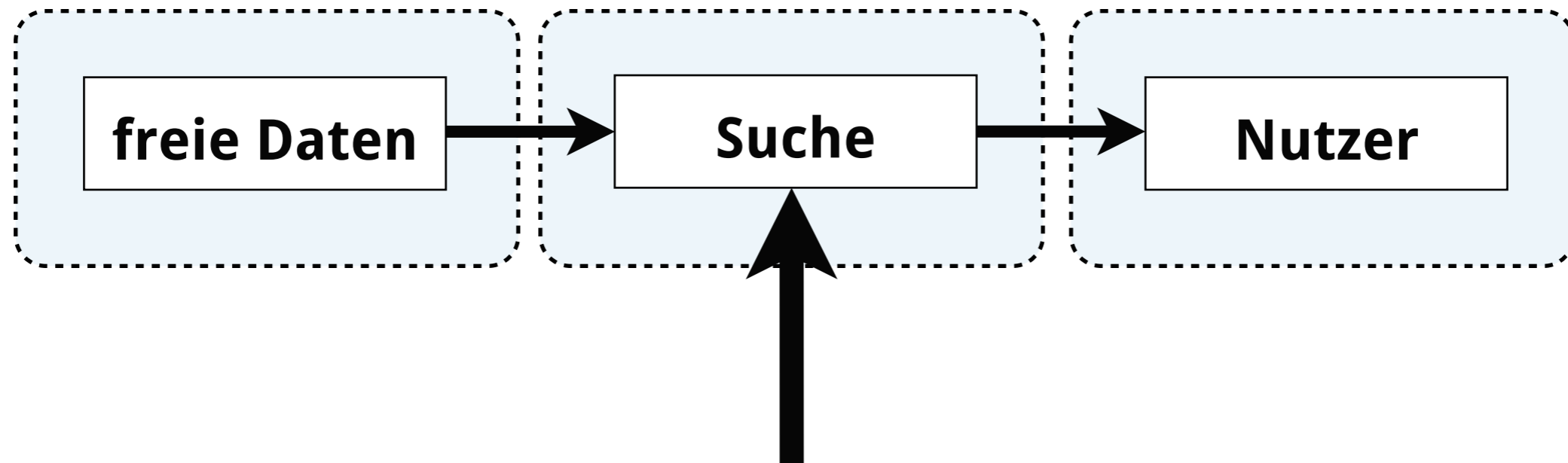
Suchmaschinen als Brücke zwischen freien Inhalten und Nutzern

1. These: **freie Daten können nur dann wirklich frei sein wenn sie mit freier Software gefunden werden können**



Suchmaschinen als Brücke zwischen freien Inhalten und Nutzern

1. These: **freie Daten können nur dann wirklich frei sein wenn sie mit freier Software gefunden werden können**



2. These: **zwischen freien Informationen und dem Nutzer fehlt eine wesentliche Verbindung: die freie Suche**

Bürgerrechte und Wahrung der Privatsphäre

Das Internet und das World Wide Web sollten nach seiner ursprünglichen Philosophie eine alle-zu-alle Infrastruktur bieten und nicht nur Sender-Empfänger Verbindungen aufbauen.



3. These: **Zensur durch Dritte kann nur durch das Betreiben einer eigenen Suchmaschine verhindert werden.**

&

4. These: **Das Betreiben einer Suchmaschine für eigene Zwecke ermöglicht eine bessere Qualität der Suchergebnisse.**

Datenschutz und Wahrung von Geschäftsgeheimnissen

Unternehmensumfeld



Discussions



Wiki



Code



Bugtracker

Suchmaschine

Das Internet



5. These: Zur Wahrung der Privatsphäre und der Geheimhaltung ist es notwendig, dass der Suchmaschinennutzer eine eigene Suchmaschine betreibt

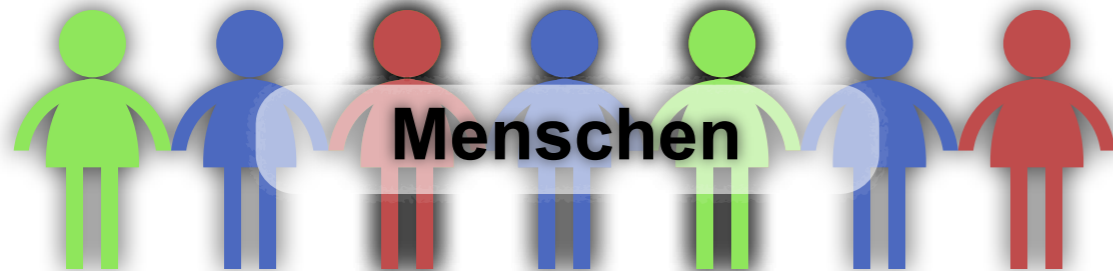
Soziologische Aspekte von Suchmaschinen

Unterscheidung Suchmaschinen/Datenbanken

(beides sind Varianten eines Assoziativspeichers)

unstrukturierte Daten

Suchmaschine



6. These: **Suchmaschinen finden Informationen in unstrukturierten Daten für Menschen...**

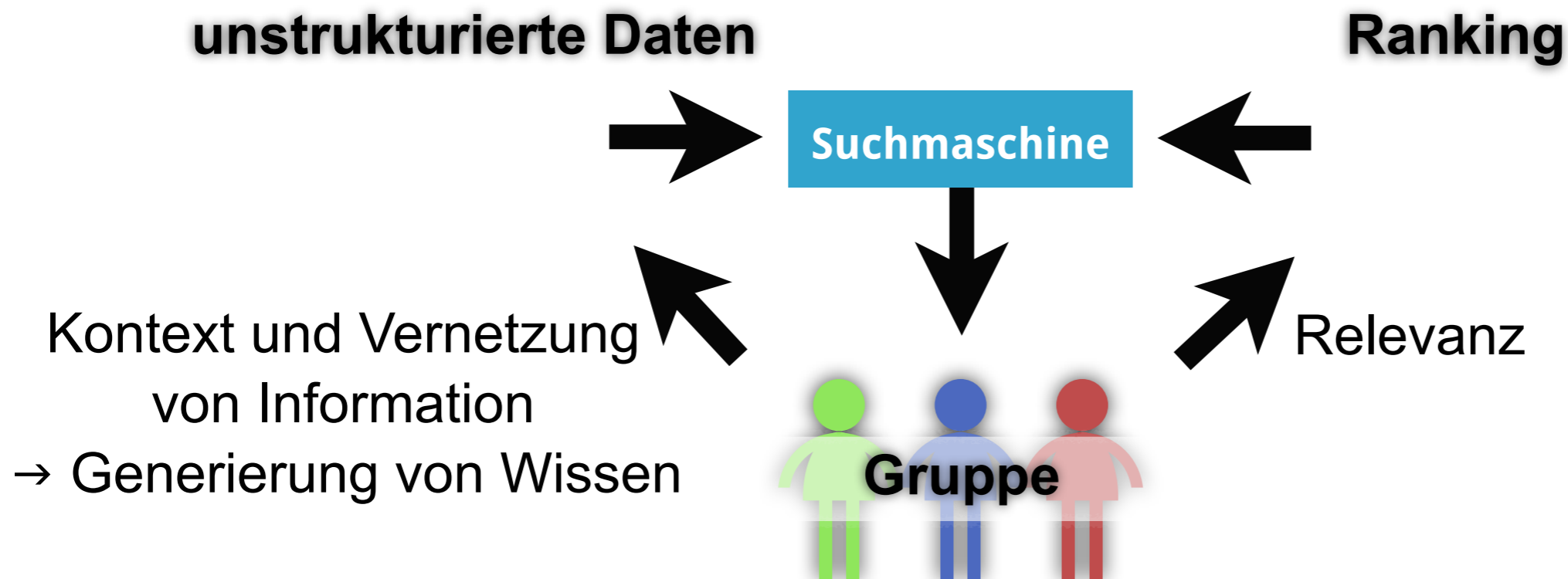
strukturierte Daten

Datenbank



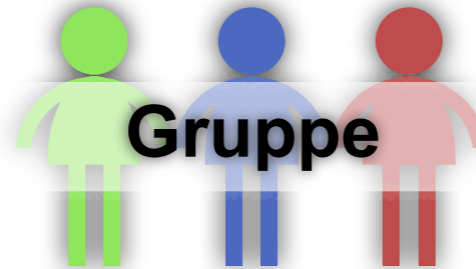
...**Datenbanken finden Informationen in strukturierten Daten für Programme**

Soziologische Aspekte von Suchmaschinen



7. These: Eine Suchmaschine findet relevante Informationen nur für eine bestimmte Gruppe von Menschen, wobei diese Gruppe die gleichen Relevanz-Kriterien teilt

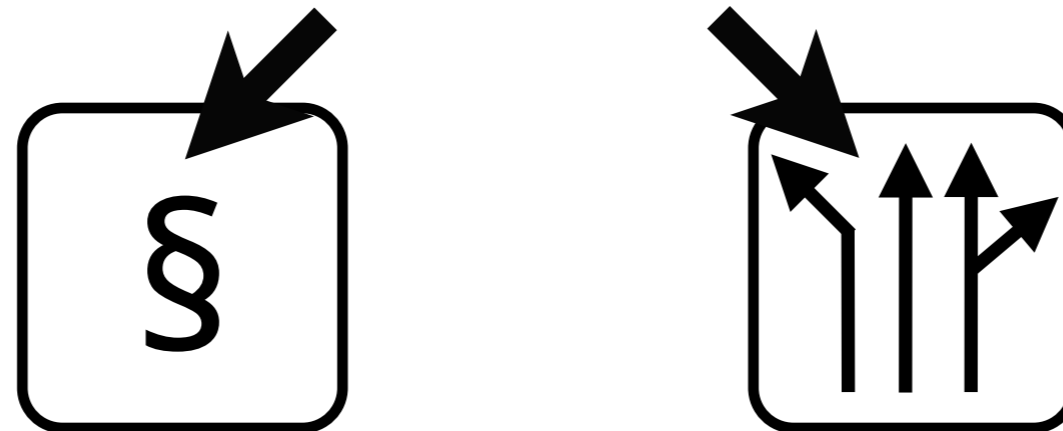
Soziologische Aspekte von Suchmaschinen



8. These: **Gemeinschaften und Gesellschaften brauchen eigene Suchmaschinen um relevante Informationen optimal finden zu können**

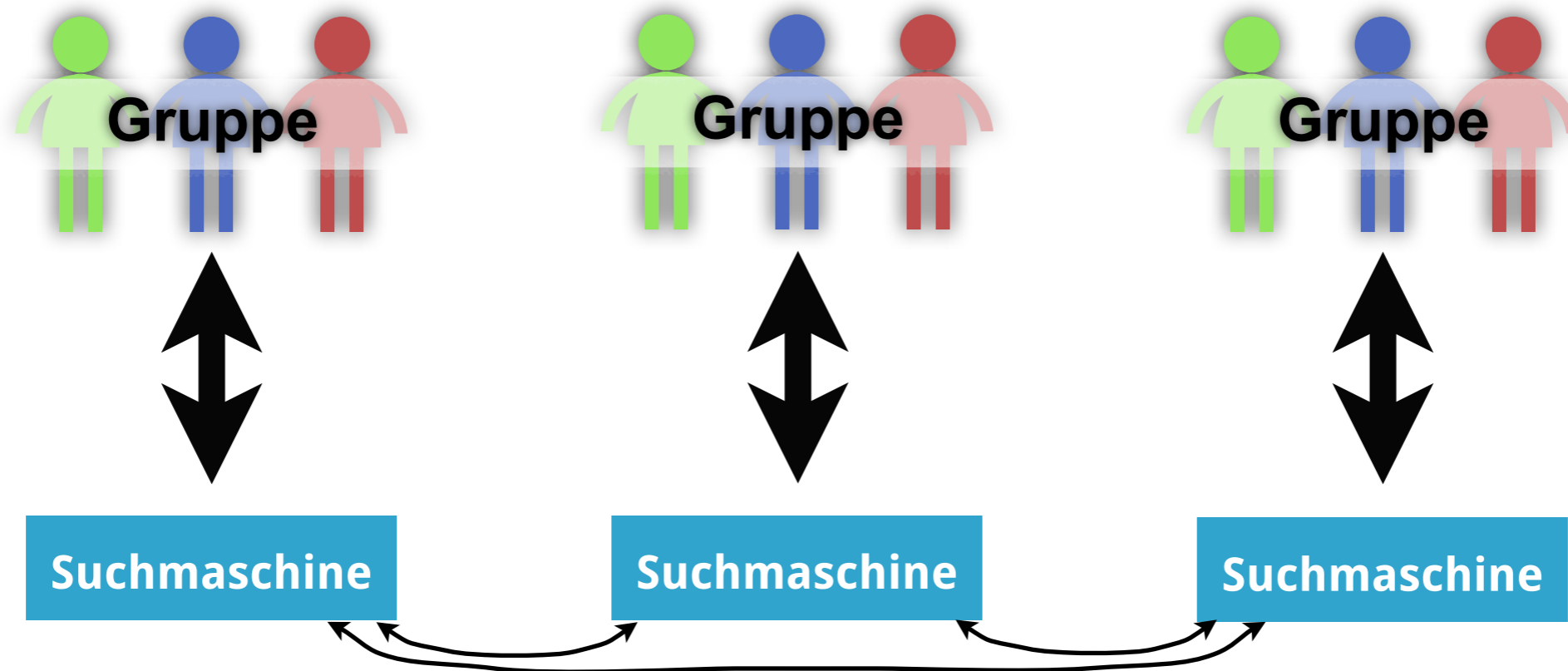


9. These: **Die Suchergebnisse einer Suchmaschine für eine bestimmte Gemeinschaft ist meinungsbildend für diese Gemeinschaft im Bezug zu Normen und Werten**



Soziologische Aspekte von Suchmaschinen

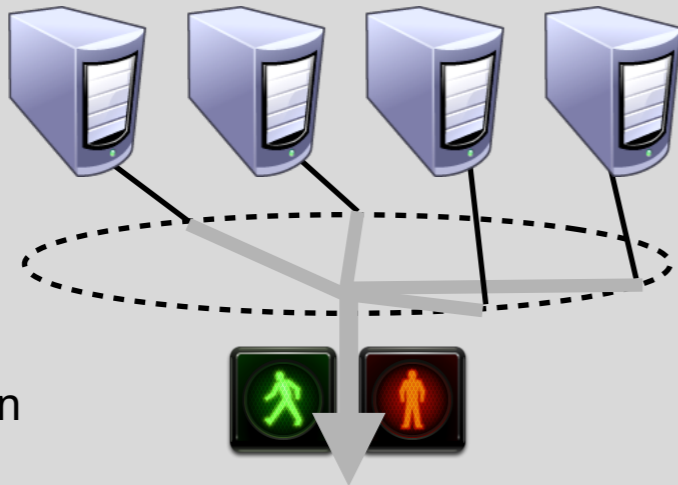
Fazit: **verschiedene Gruppen** → **verschiedene Interessen** → **eigene Suchmaschinen**



10. These: **Die Forderung nach einer nicht-monopolisierbaren und nicht-zensurierbaren Suchmaschine mit gleichberechtigten Rechten zur Moderation des Suchindexes kann nur von einer dezentralen Suchmaschinenteknik erfüllt werden.**

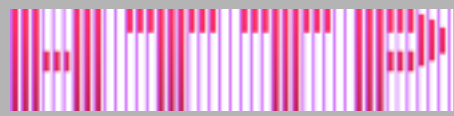
Crawler

Server
im Netz



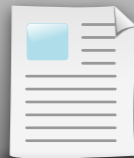
„round-robin“
Zugriff

robots.txt und
Latenz beachten



FTP **SMB**

viele
Dateiformate

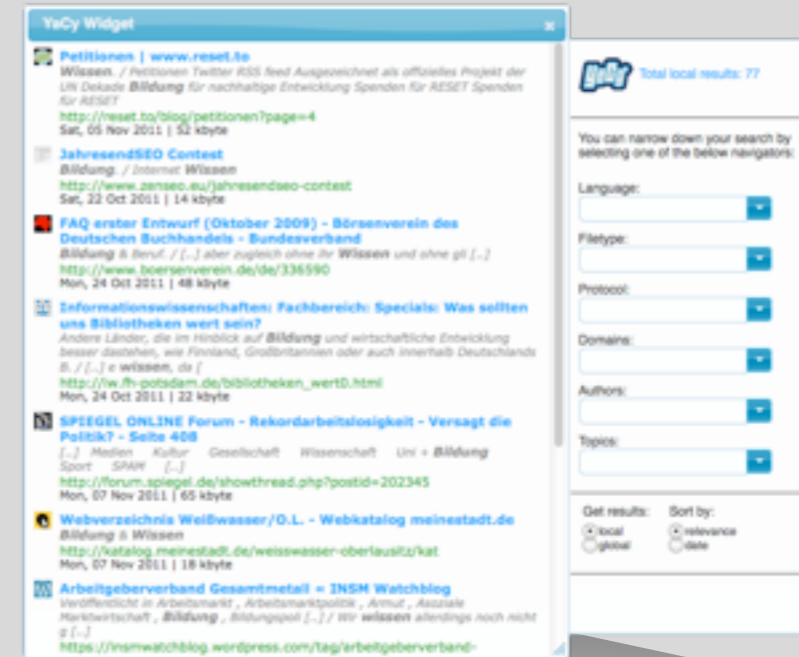


Parser

z.B.: HTML, XHTML, RSS, RDF, XHTML+RDFa, FOAF,
vCard, Flash, PDF, PS, Word, Excel, Visio,
Powerpoint, OpenOffice, RTF, csv, gzip, zip, tar,
rar, bzip2, 7zip, images(EXIF)

Dokumente

Suchseite



Suchen und Ergebnisse mit
Navigatoren darstellen

Snippets

Indexierer

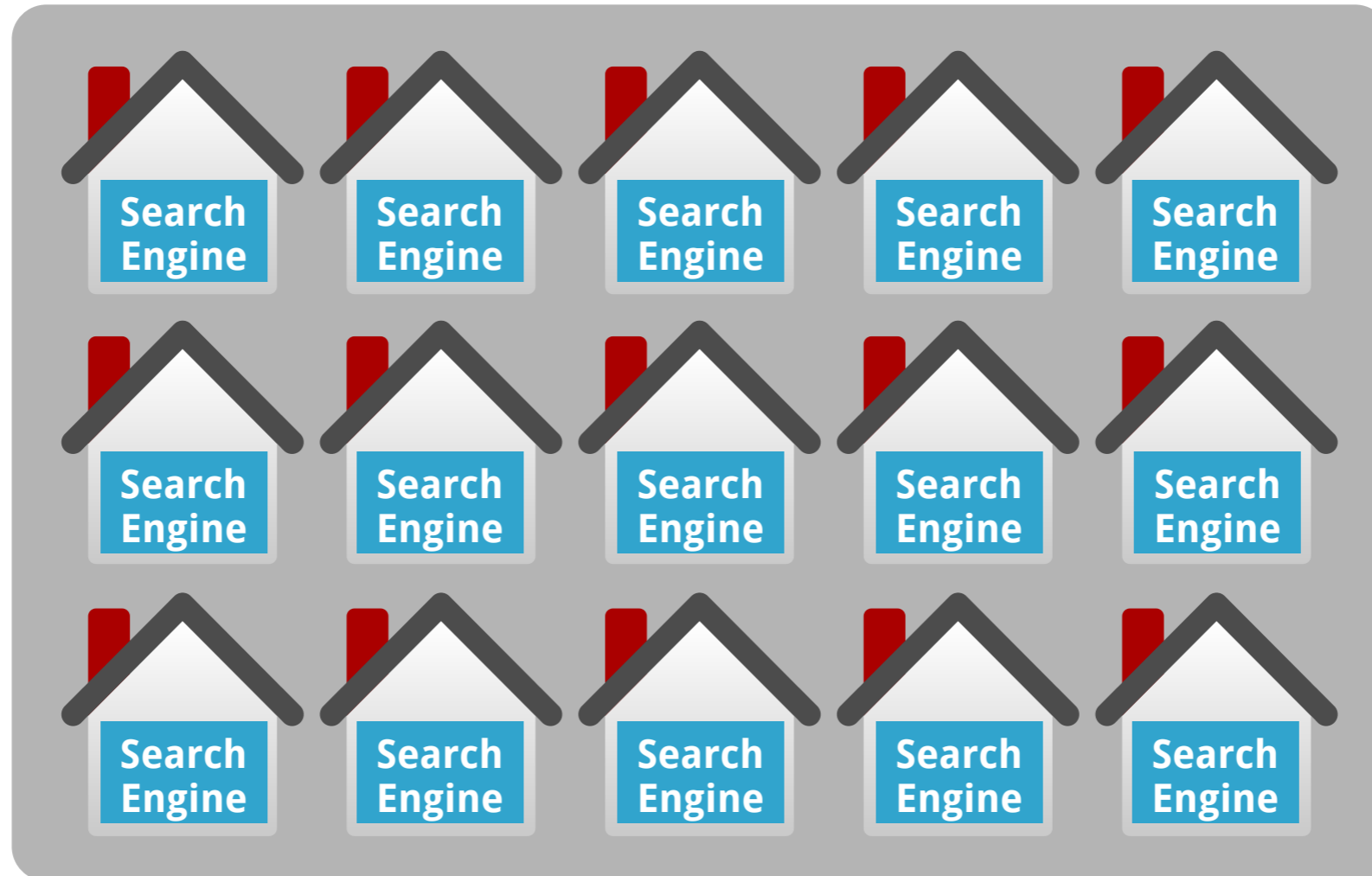
Verschlagwortung ist die Umkehrung der
Dokument: $\text{Referenz} \rightarrow \{\text{Wort}^*\}$
Relation zu einer
Index: $\text{Wort} \rightarrow \{\text{Referenz}^*\}$
Relation

Dokumentencache

Leistungsfähige, skalierbare Suchmaschinen



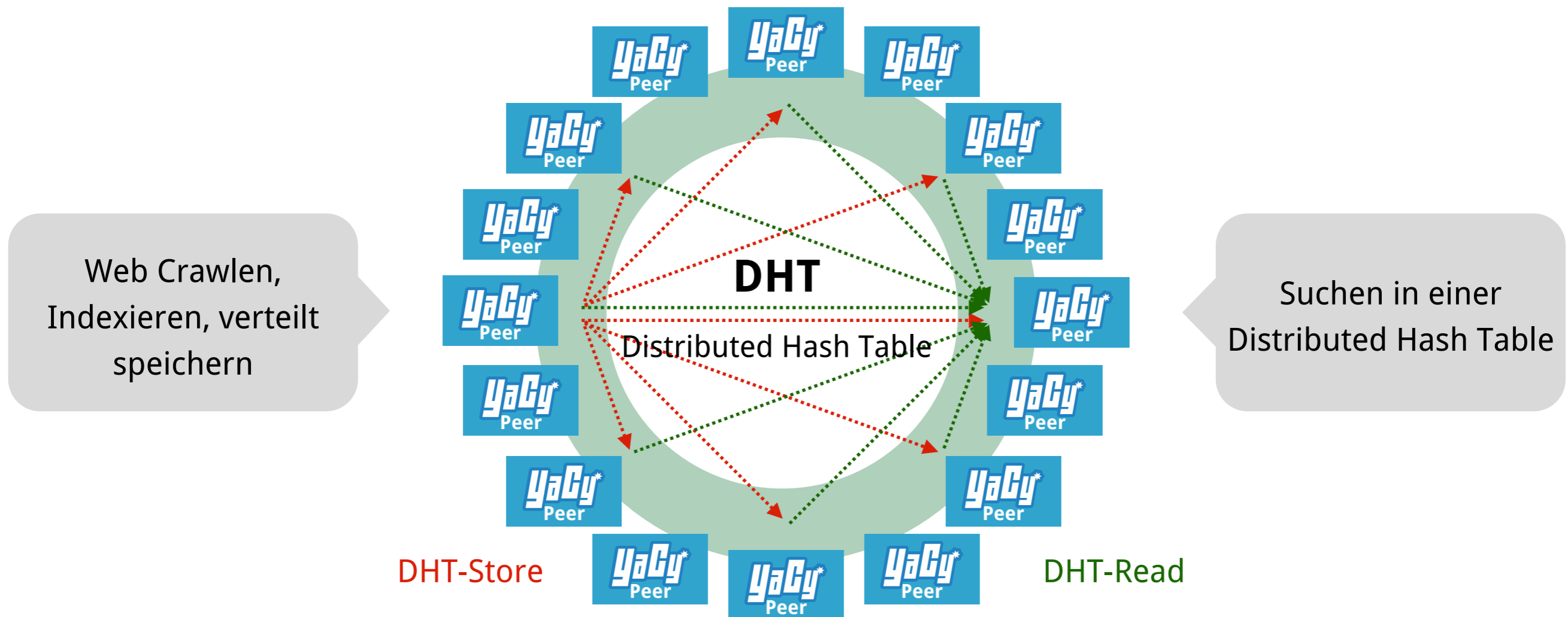
Leistungsfähige, skalierbare Suchmaschinen



Dezentrale Websuche mit Peer-to-Peer Technik



Peer-to-Peer Suchmaschine für das World Wide Web



Das YaCy Suchmaschinennetz besteht aus eigenständigen Suchmaschinen, aber sie sind auch miteinander vernetzt.

Teil 3 - Freier Wissenszugang mit YaCy



Such-Appliance für Portale im Internet oder Intranet



Communities

Your Project



Discussions



Wiki



Code



Bugtracker



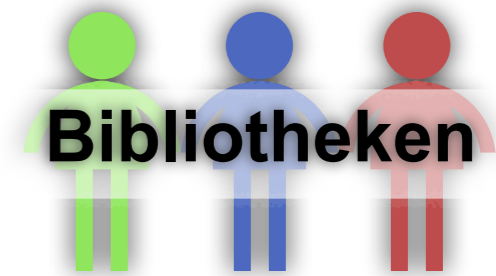
Search Engine



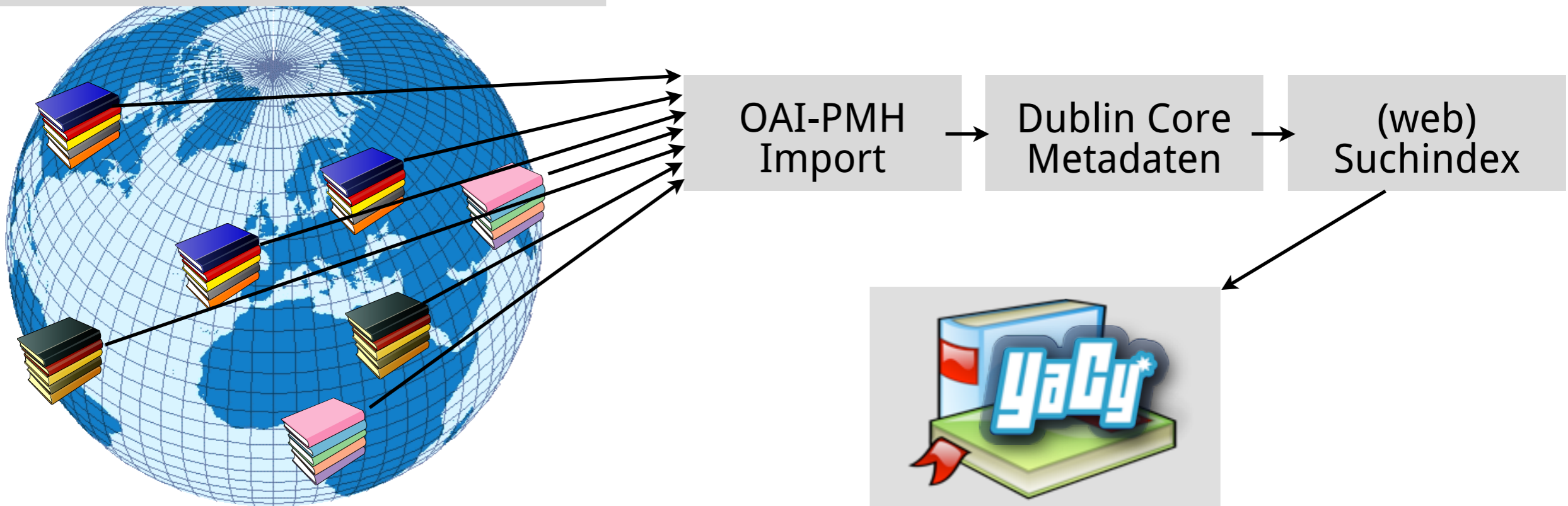
Teil 3 - Freier Wissenszugang mit YaCy



Suchmaschine für virtuelle Fachbibliotheken



>2000 Open Access Repositories



<http://oai.yacy.net>

Fachbibliothek deutsche Universitäten

http://blog.yacy-kochbuch.de/?page_id=23



Teil 3 - Freier Wissenszugang mit YaCy



Suchmaschine für wissenschaftliche Publikationen



Sciencenet:

<http://sciencenet.kit.edu>

Suchmaschine des
Karlsruhe Institut für
Technologie

35 Peers in eigenem Netz,

> 200 Mio. Dokumente



Teil 3 - Freier Wissenszugang mit YaCy



freie dezentrale Suchmaschine für alle



freeworld:

<http://search.yacy.net>

Suchmaschine, bestehend
aus den Peers freiwilliger
600 Peers im Monat,
weltweit verteilt

> 1 Milliarde Dokumente



Verwendung von Suchmaschinentechnik

- technologische und soziotechnische Ebene



Werden die Forderung der Charta der Bürgerrechte erfüllt?

(a) **Wissen** ist Erbe und Besitz der Menschheit und damit **frei**.

(b) Der **Zugriff** auf Wissen muss **frei** sein.

(c) **Alle Menschen** haben das Recht auf Kommunikation und Informationsfreiheit.

(d) Das Recht auf Achtung der **Privatheit ist ein Menschenrecht** und ist unabdingbar für die freie und selbstbestimmte Entfaltung von Menschen in der Wissensgesellschaft.

Freie Suchmaschinen sind die Verbindungen zwischen freien Inhalten und dem Nutzer

Suchmaschinentechnik muss für alle erreichbar und einfach bedienbar sein.

Die Suchmaschine ist nur dann privat, wenn sie nicht aus einer zentralen Instanz heraus genutzt werden muss, sie muss dezentral sein.

*YaCy ist freie Software,
download von <http://yacy.net>*

Die Bedingungen werden von YaCy erfüllt!



Vielen Dank für Ihre Aufmerksamkeit!

Suchmaschinen, Wissensmanagement,
Crowdsourcing, TeamComputing:
Wir beraten Sie!



QR-Code: vCard

YaCy Suchmaschinenprojekt

Dipl. Inf. Michael Christen,
mc@yacy.net
<http://yacy.net>

Beratung TeamComputing

Dipl. Kaufmann Georg von Nessler,
Nessler@IP-ShareMedia.de
Dipl. Inf. Michael Christen,
Michael.Christen@IP-ShareMedia.de
<http://ip-sharemedia.de>

