

Eine offene SuMa-eV Plattform für Open-Crawl Suchmaschinen

3. SuMa-eV Forum, 28.9.2006, Berlin
Michael Nebel

September 2004:

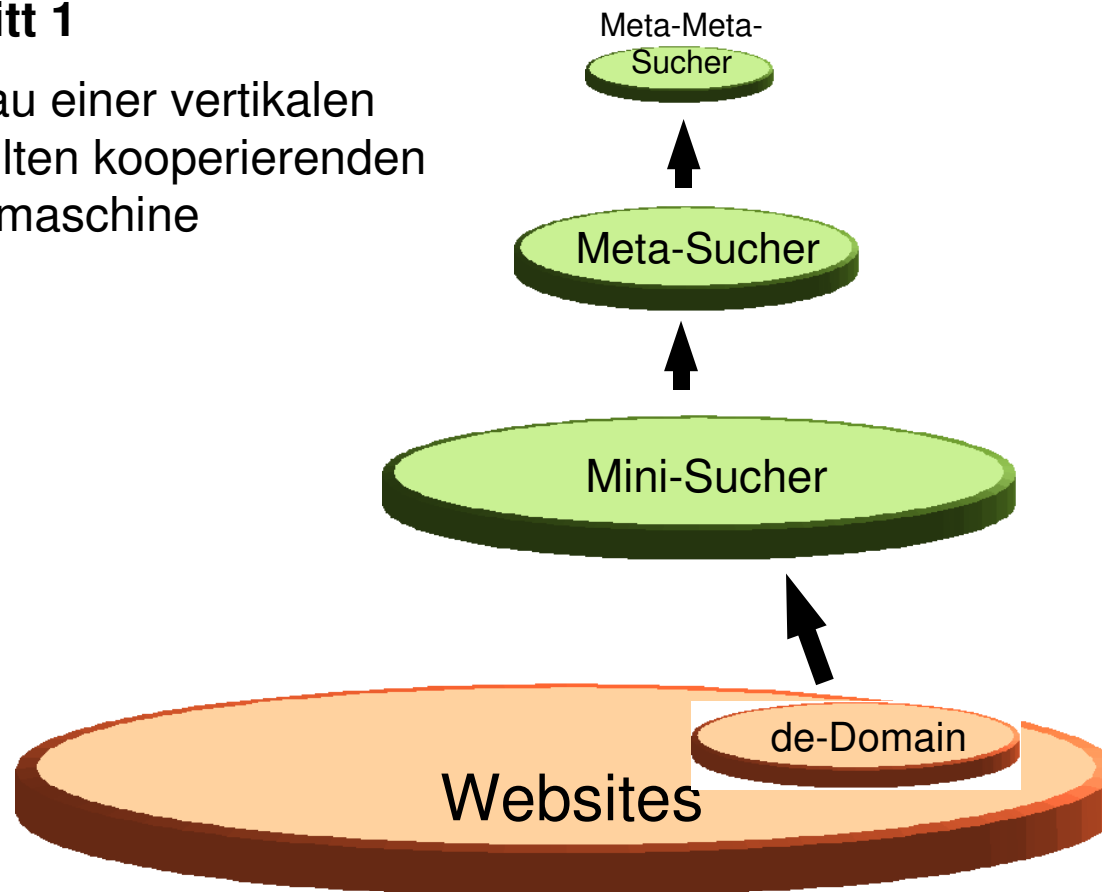
Konzept zum Aufbau einer vertikalen verteilten kooperierenden Suchmaschinen-Struktur in drei Schritten

- 1.) Aufbau von thematischen Minisuchern
- 2.) Überführung der Hierarchien in einen Verbund
- 3.) Aufbau eines "Suchmaschinen Mesh"

Quelle: "Suchmaschinen: Vielfaltsicherung und innovative Konzepte", September 2004,
Univ.-Prof. a.D. Dr. H. Pralle, Dr. W. Sander-Beuermann, Universität Hannover, Dipl.-Ing. M. Nebel

Schritt 1

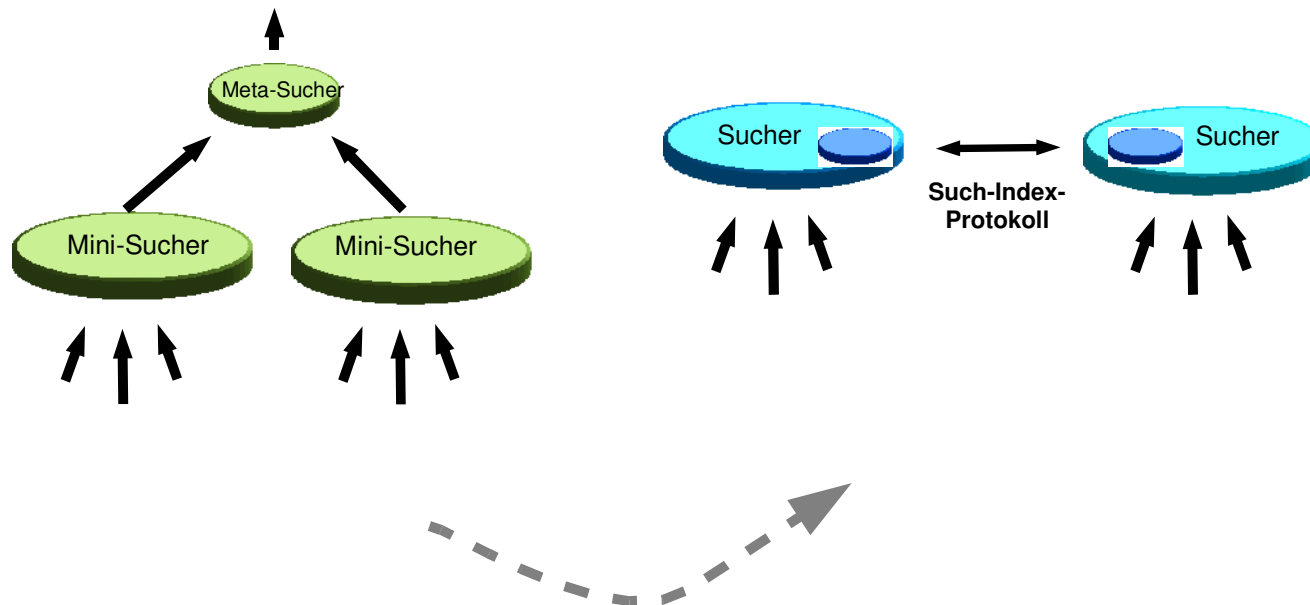
Aufbau einer vertikalen
verteilten kooperierenden
Suchmaschine



Quelle: "Suchmaschinen: Vielfaltsicherung und innovative Konzepte", 09/2004, Univ.-Prof. a.D. Dr. H. Pralle, Dr. W. Sander-Beuermann, Universität Hannover, Dipl.-Ing. M. Nebel

Schritt 2

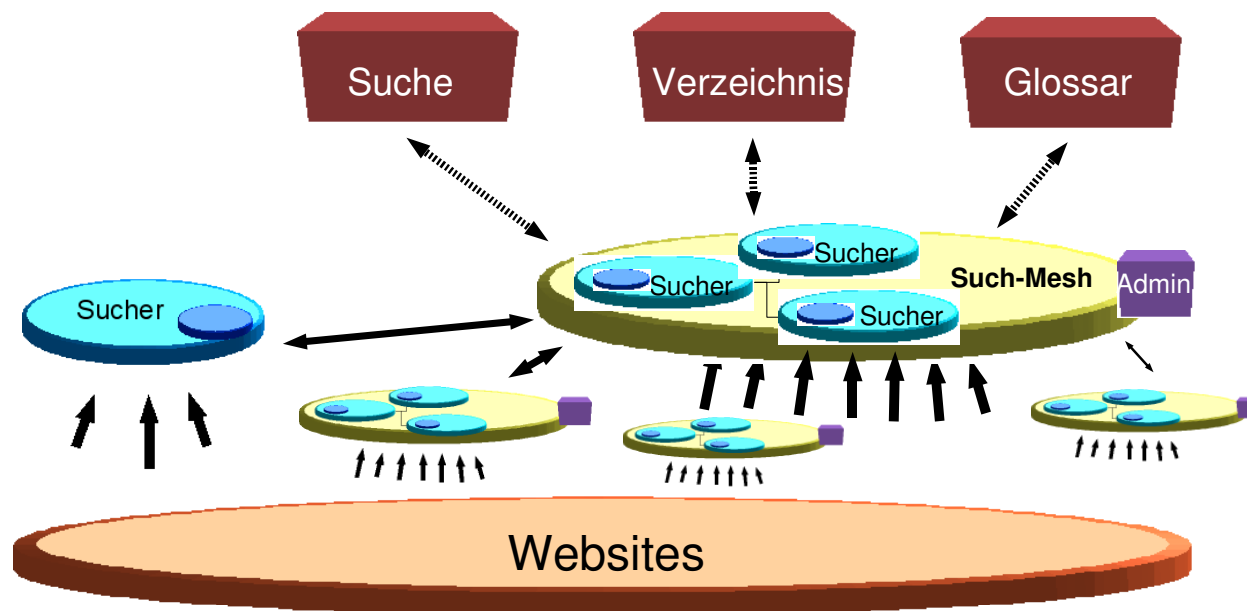
Überführung der vertikalen Hierarchie in einen Suchmaschinenverbund



Quelle: "Suchmaschinen: Vielfaltsicherung und innovative Konzepte", 09/2004, Univ.-Prof. a.D. Dr. H. Pralle, Dr. W. Sander-Beuermann, Universität Hannover, Dipl.-Ing. M. Nebel

Schritt 3

Erweiterung der monolithischen Sucher zu einem Such-Mesh
Einführung neuer Dienste.



Quelle: "Suchmaschinen: Vielfaltsicherung und innovative Konzepte", 09/2004, Univ.-Prof. a.D. Dr. H. Pralle, Dr. W. Sander-Beuermann, Universität Hannover, Dipl.-Ing. M. Nebel

- Gezieltes Erfassen einzelner Themengebiete und Einbindung von hochwertigen „Nischeninhalten“
- Einbindung von Personen/Organisationen, die im Thema des Minisuchers aktiv sind.
- Zentrale Anlaufstellen zur Erhöhung der Reichweite der einzelnen Minisucher
- Bildung von Such-Hierarchien und -Netzen

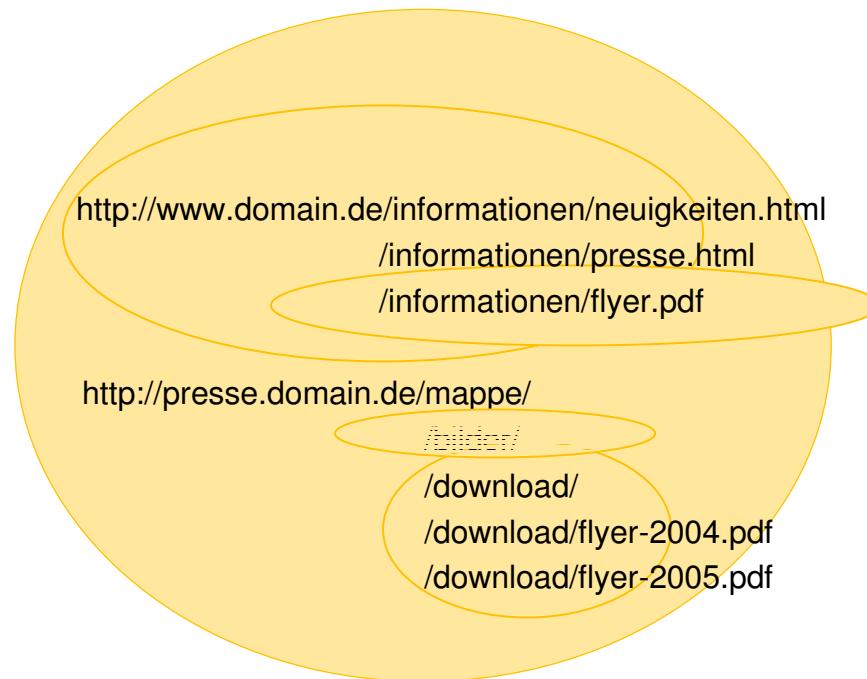
- Unterstützung des Konzepts außerhalb des Vereins im Wesentlichen nur durch gute Worte und durch Einzelpersonen
- Frei verfügbare Software ist nur für Techniker geeignet
- Peer-To-Peer-Suchmaschine: yacy
- Betrieb verschiedener kleiner Suchmaschinen
- Metasucher:
 - metager2.de
 - meta4media.de

- Stabilisierung des Regelbetriebs
- Ausbau der Infrastruktur
- Refinanzierung der Betriebskosten
- Entwicklung neuer Algorithmen zur Verdichtung der Daten
- Steigerung des Mehrwerts für die Benutzer der Suchmaschine
- *Weiterer Aufbau fokussierter Minisucher*

- Bereitstellung einer Plattform für den Aufbau von Minisuchern

<http://www.opencrawl.de/>

- Administration per WWW und E-Mail
- Option zur Einbindung des eigenen Minisuchers auf der eigenen Website (HTML-Include, Opensearch-XML).
- Einfache Suchmaske auf [opencrawl.de](http://www.opencrawl.de)



1. Eingabe der URL

2. Auswahl der Reichweite

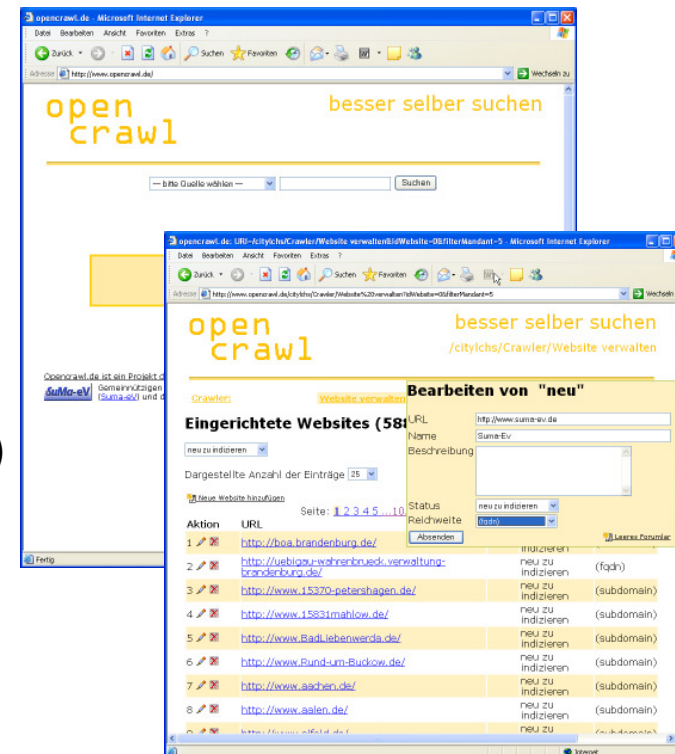
- diese Domain
- alle Unterdomains
- einen Unterast
- exakt ein Dokument
- Ausschluss von Dokumenten

3. Prüfung der Ergebnisse

- Basis: Nutch-/Hadoop-Cluster von netluchs.de
- Limitierung nach "best effort"
- Nutzung des aufgebauten Index der Minisucher in den Metasuchern
- Persönliche Anmeldung



- Anmeldung per E-Mail an info@opencrawl.de
- Informationen:
 - Wer sind Sie?
 - Was wollen Sie indizieren?
 - Wo wollen Sie die Daten nutzen?
(welche Schnittstellen brauchen Sie?)



Kontakt

Dipl.-Ing. Michael Nebel

eMail: michael@nebel.de

Telefon: +49 40 851 581 45