

Semantisches Routing in dezentralen Netzwerkstrukturen

Diana Tsakadze

Seminar: Semantic Grid

Dozenten: Prof. Steffen Staab
Bernhard Tausch

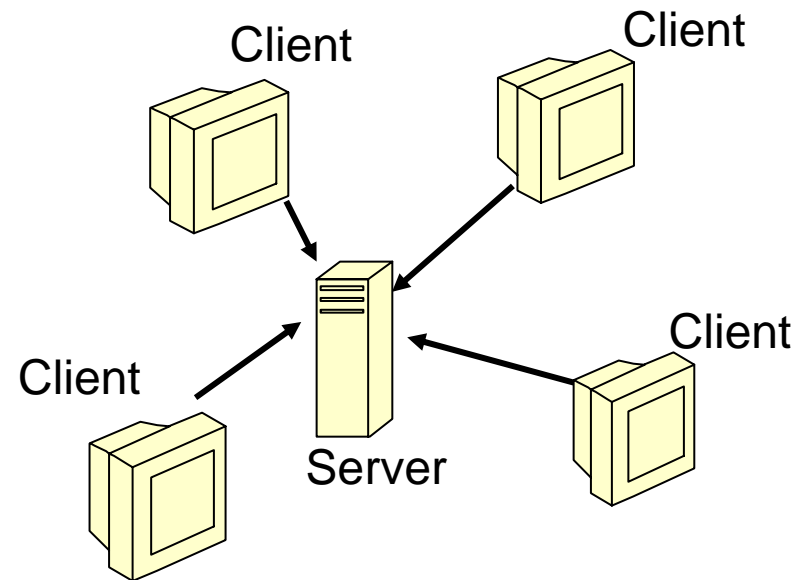
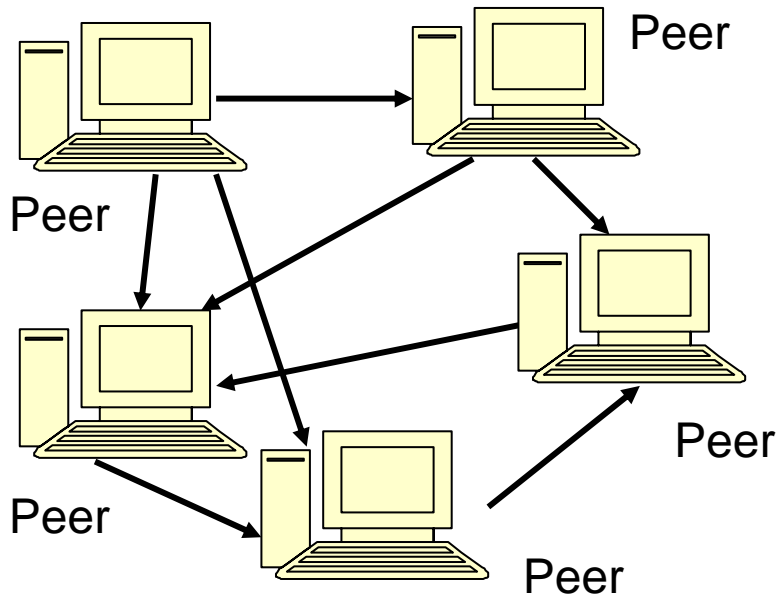
26. Januar 2005

1. Dezentrale Netzwerkstrukturen
 - 1.1 Peer-to-Peer Computing
 - 1.2 Peer-to-Peer und Grid Computing
2. INGA – semantisches Peer-to-Peer System
 - 2.1 Semantische Shortcut Netzwerke
 - 2.2 Basis Infrastruktur
 - 2.3 Queryabhängige und Queryunabhängige Shortcuts
 - 2.4 Peer Selektions-Algorithmus
3. Bibster – Assistent zur Suche und Managing der Bibliographischen Metadaten
4. Zusammenfassung und Ausblick

Dezentrale Netzwerkstrukturen

Peer-to-Peer Computing

- Peer (eng. Gleichgestellter)
- Peer-to-Peer vs. Client/server Prinzip



Peer-to-Peer Computing - Vorteile

- File-Sharing System
- Vermeidet ‚Bottleneck‘ für Rechenperformance und Informationsupdate
- Robust gegen Ausfall einzelner Komponente (Peers)
- Skalierbar in Datenvolumen und in verbundene Teilen

Problem

- ‚chaotische‘ Natur – Mangel vom kohärenten Schema

Grid Computing

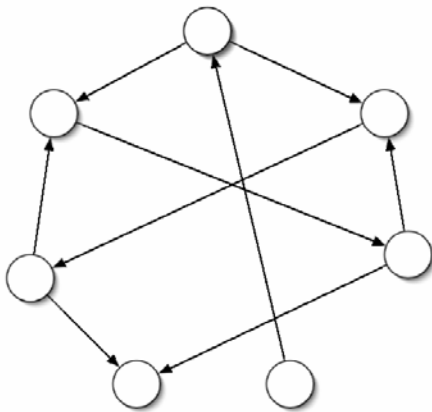
- erlaubt noch das Teilen von anderen Ressourcen z.B. Rechenleistung
- Die Grid Architektur kann man konzeptuell in zwei Layer teilen:
Daten- und Rechen Grid
- Rechen Grid – eine virtuelle Maschine wird erzeugt durch das Zusammenbringen einer größeren Anzahl unabhängiger Hosts
- Daten Grid – massive zusammengebundene Speicher Infrastruktur, verbunden mit einem Rechen Grid durch High-Speed Netzwerkverbindung

Dezentrale Netzwerkstrukturen

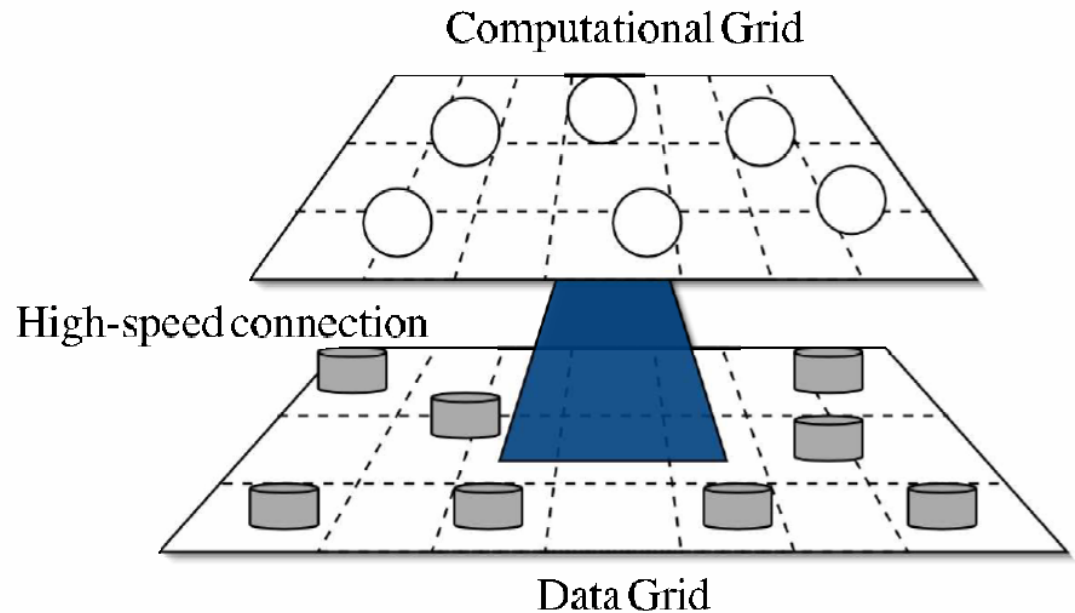
Peer-to-Peer und Grid Computing

Große Herausforderung- Verbinden von zwei Architekturen
- kombinieren der Chaotische Natur des P2P mit strukturiertem Ansatz des Grid

Peer-to-Peer System



Rechen- und Daten Grid



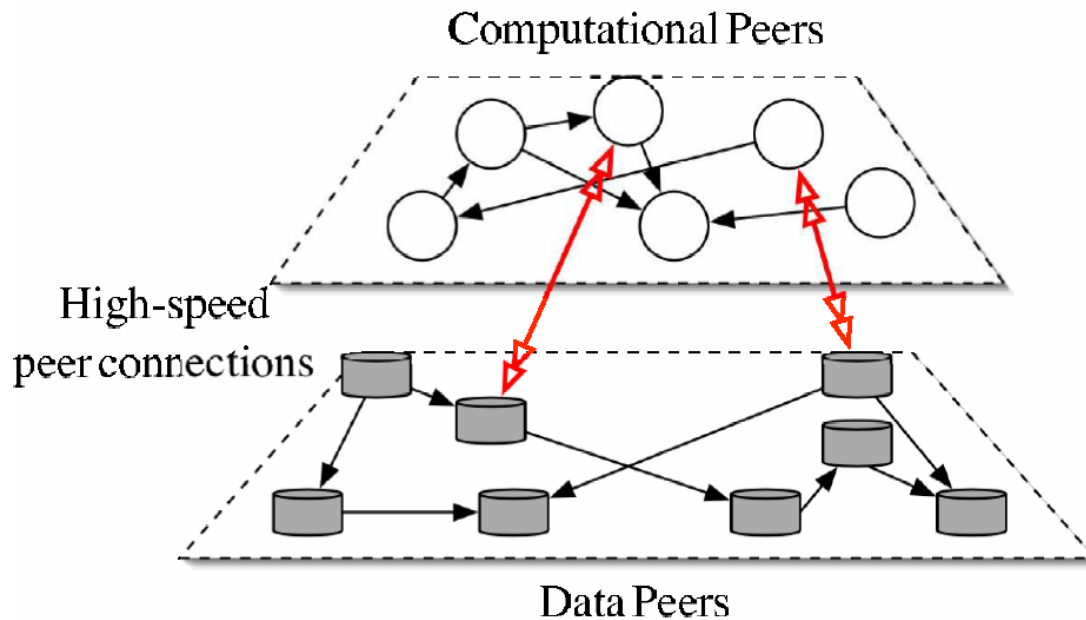
Dezentrale Netzwerkstrukturen

Peer-to-Peer und Grid Computing

Fazit

- **Peer-Peer System kann wie ein Data Grid mit breiterer Grid Architektur handeln**

Pyragrid Architektur:



Dezentrale Netzwerkstrukturen

- Probleme und Herausforderung zu P2P
 - keine einheitliche kohärente Schema
 - unkoordinierte Peers
 - duplizieren der Informationen
 - duplizierte Antworten auf eine Anfrage
- Semantisches Peer-to-Peer Netzwerk – Peers kooperieren für effizientes Routing der Anfrage
- INGA – interest-based Node Grouping Architecture
- Im Vergleich zu Gnutella werden die Anzahl der Messages halbiert, daher wird die Zugriffszeit vermindert

- Problem: finden des besten Peers in einem Peer-to-Peer Netzwerk bei einem bestimmten Query
- Effizientes Routing von Query anhand des shortcutbasierten Overlay mit ausschließlich lokalem Wissen
- Wichtigkeit von Shortcuts für den Informationsfluß im menschlichen Netzwerk
- Menschen mit lokalem Wissen können Nachrichten effektiv steuern und übertragen indem sie in ihrer Umgebung algorithmische Prozeduren ausführen

Semantische Shortcut Netzwerke

- Hauptaufgabe von INGA - finden des richtigen Peers, der eine gute Antwort auf die Frage gibt oder diese zum richtigen Peer weiterleitet
- Menschen, die ein Antwort auf ihre Frage suchen können zwischen drei Typen von Personen innerhalb eines sozialen Netzwerkes wählen

Content Provider - hat früher auf diese Frage mit Erfolg geantwortet

Recommender – hat sich früher mit eine ähnliche Frage beschäftigt. Es wird angenommen, dass er einen geeigneten Content Provider kennt

Bootstrapper - hat guten sozialen Netzwerk in mehreren Bereichen

- Information über einen anderen Peer wird lokal in einem Shortcut gespeichert

Basis Infrastruktur

- Unterschied zur anderen p2p Systemen:
 - Kooperation unter Peers
 - Informationsaustausch zur Selektion eines „guten“ Peers für eine Anfrage
 - Jeder Peer speichert **Qeries x Peers** Paare in einem **Shortcut Index** nach einer erfolgreichen Abfrage
- Ein INGA Peer speichert „egoistisch“ - er behält nur Shortcuts, die für ihn selbst interessant sind

Basis Infrastruktur- lokale Datenbanken

Hauptaufgaben

- **Inhalte des Benutzers veröffentlichen -**

Der lokale Knoten speichert und veröffentlicht Information.
In der Implementation von INGA sind es RDF(S) Statements.

Ein Statement beschreibt
Daten - z.B. (*WWW2005 InstanceOf WWWConf*) oder
Konzeptionelle Information - z.B. (*WWWConf subclassOf Conferences*)

- **Durchführen der Abfrage und liefern der Ergebnisse**

Basis Infrastruktur - Shortcut Management Komponente

Hauptaufgaben

- **Lernen und Speichern von Shortcuts**
- **Empfehlen von Top-k Peers für die Abfrage**
 - **Thema der Anfrage**
 - **Ähnlichkeit zwischen der Anfrage und lokal gespeicherte Shortcuts**

Basis Infrastruktur – Query Messages und Result Messages

Query Message besteht aus

- **Query selbst (Topic)**
- **Bootstrapping Fähigkeit des Peers**
- **Message Path – eindeutige ID's von Peers, die diese Abfrage schon bekommen haben**
- **QueryID**

Result Message enthält zusätzlich

- **Antwort auf der Abfrage**

Basis Infrastruktur – semantische Ähnlichkeits-Funktion

Ähnlichkeits-Funktion zwischen Query $q \in Q$ und Shortcut $\phi \in \Phi$

$$\text{sim} : Q \times \Phi \mapsto [0;1]$$

$$\text{sim}(q, \phi) = \begin{cases} e^{-\alpha l} \cdot \frac{e^{\beta h} - e^{-\beta h}}{e^{\beta h} + e^{-\beta h}} & q \neq \phi \\ 1 & \text{sonst} \end{cases}$$

l - Länge des kürzesten Pfades zwischen q und ϕ

h - minimale Tiefe von q oder ϕ im Baum

$\alpha \geq 0 (\alpha = 0.2)$ und $\beta \geq 0 (\beta = 0.6)$

Queryabhängige Shortcuts

Shortcut Management Komponente pflegt INGA-Peers queryabhängige

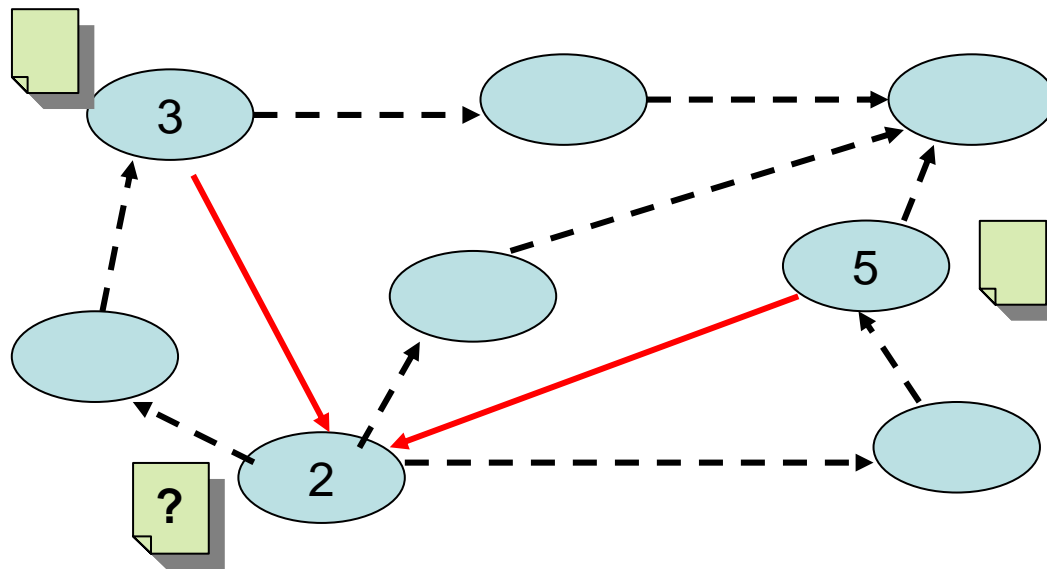
Shorcut Indizes

- 1. Content Provider Shortcuts – Peer bietet einen angemessenen Inhalt für eine Anfrage**
- 2. Recommender Peer Shortcuts – Peer hat früher eine ähnlich Frage gesendet**

Queryabhängige Shortcuts - Content Provider Shortcut

- Interessen-Basierte Lokalisierungsprinzip
- Nach der erfolgreichen Abfrage wird Shortcut ϕ eines neuen remote Peers zu Content Provider Shortcut Liste zugefügt in der Form:
$$\phi(\textit{Topic}, \textit{PID}, \textit{QueryHits}, \textit{C})$$
- Festgelegte Maximallänge der Content Provider Shortcut Liste

Erzeugung von Content Provider Shortcut

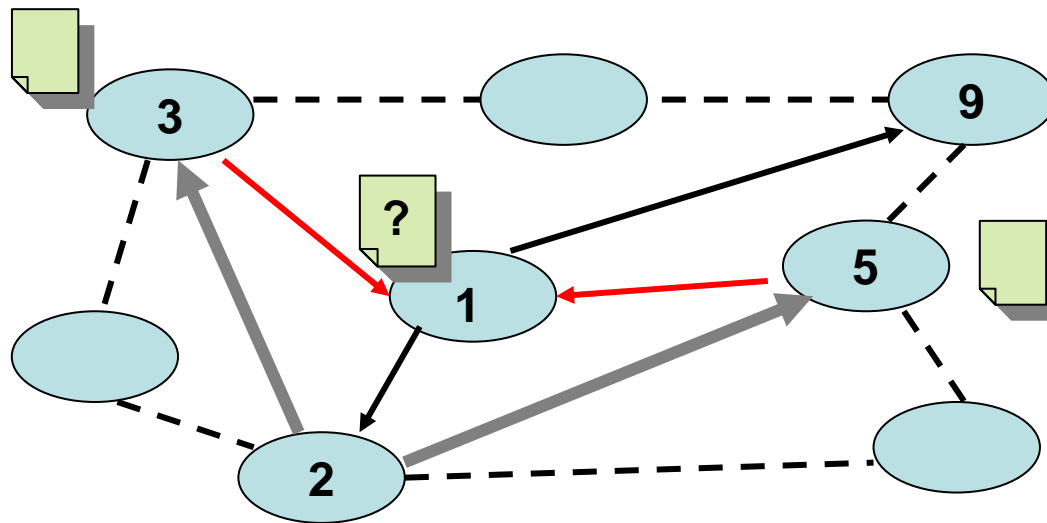


Topic	PID	QueryHits	Type	Utility
/Education/UML	3	23	C	1
/Education/UML	5	20	C	1

Erzeugung von Recommender Shortcut

- Sehr aktive Peers erstellen viele erfolgreiche Abfragen
- Recommender Peer – ein aktiver Peer, mit hoher Überlappung der Anfragen des lokalen Peer
- Wenn kein Content Provider Peer bestimmt werden kann, dann wird in den lokalen Recommender Shortcuts gesucht und die Abfrage zum ‚besten‘ Recommender geschickt
- Hohe Trefferwahrscheinlichkeit

Erzeugung von Recommender Shortcut – aktive Entdeckung



Topic	PID	QueryHits	Type	Utility
/Education/UML	3	23	C	1
/Education/UML	5	20	C	1
/Education/UML	2	43	R	1

Erzeugung von Recommender Shortcuts - passives lernen

- Beobachten von eingegangenen Abfragen
- Speichern von Topic und PID als Recommender Shortcut, wenn die Abfrage das Interesse des lokalen Peers trifft
- Abfrage trifft das Interesse des lokalen Peers, wenn die Ähnlichkeit $sim(q, r)$ zwischen veröffentlichter Resource r und Abfrage q höher ist als der Grenzwert t_{listen}
- QueryHit = 1, da die Information über die Ergebnisse für diese Abfrage nicht bekannt ist

Index Replacement Policy und Indexgröße

- Begrenzen der Größe des Index
- Ersetzen der Shortcuts durch Index Replacement Policy

$$utility(\phi) = \frac{|successfulusages|}{|totalusages|}$$

- Nutzwertbasierte Policy – Shortcut mit niedrigsten Nutzwert wird entfernt

Queryunabhängige Shortcuts

- Lokales Wissen über Content Provider oder Recommender ist nicht vorhanden
- Bootstrapping Knoten – Peer mit vielen Shortcuts zu vielen remote Peers
- Höhere Trefferwahrscheinlichkeit im Vergleich zu default Routing Strategie z.B. Flooding (Überflutung)
- Jedes Query beinhaltet bootstrapping Information von abfragendem Peer
- während der Peer online ist, wird Bootstrapping Index aktualisiert
- Peer sendet seine Anfrage zu den „besten“ bootstrapping Knoten und fängt das Wissen über Content Provider Peers zu sammeln

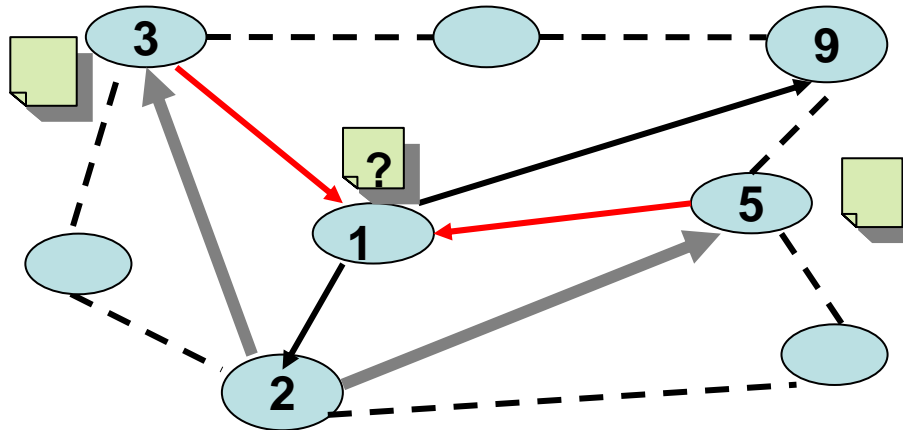
Bootstrapping Fähigkeit eines Peers wird berechnet durch:

$$P.Bo = |Shortcuts| \times |Peers|$$

P.Bo drückt die Fähigkeit des Peers aus, das breite und diverse Wissen über andere Peers zu sammeln

Queryunabhängige
Shortcut Index von Peer 2:

PID	Shortcuts	Peers	P.Bo
1	3	3	9



Peer Selektions-Algorithmus

- Abfrage wird evaluiert :
 - Lokal – in Abhängigkeit von lokalen Datenbanken. Die Antwort ist präsent
 - Durch das Netzwerk
 - Empfehlen
 - Antworten auf Query
- Empfangen von Antwort – lokale Content Provider oder Recommender wird kreiert oder aktualisiert

Shortcut Index Aktualisierungs-Algorithmus

Nach dem der Peer ein Query Message bekommt, wird erst :

- bootstrapping Information über das abfragende Query extrahiert und
- bootstrapping- (Zeile: 2-5) und Recommender- Indizes (Zeile: 6-13) aktualisiert
- erhöht die Anzahl der Sprünge (Zeile 15) und fügt die eigene PID zum Message Pfad (Zeile 16)

Require: QueryMessage qm, LocalShortCutIndex 0

- 1: Peer p=getQueryingPeer(qm)
- 2: if getBootstrappingInfo(qm)>getBootstrappingInfo(local peer)
then
- 3: /* Add querying peer to bootstrapping index if it outperforms*/
/* local value*/
- 4: addBootstrappingPeer(o,p,getBootstrappingInfo(qm))
- 5: end if

Shortcut Index Aktualisierungs-Algorithmus

```
6: Query q=getQuery(qm)
7: for all c  ContentProviderShortcuts do
8:   if Sim(q,c)>  $t_{listen}$  then
9:     /* Add new recommender shortcut with one query hit if incoming
       query matches ones interests*/
10:    addRecommenderPeer(o,q,p,l)
11:    break
12:  end if
13: end for
14: /* Update Query*/
15: incTTL(qm)
16: addToMessagePath(qm,CurrentPeerID).
```

Weiterleitungs-Algorithmus

Aufgabe: Bestimmen zu welchen k Peers die gegebene Abfrage weitergeleitet werden soll

Queryabhängige Weiterleitungs-Strategie

1. Exakte Treffer - Grenzwert für Ähnlichkeitsmaß 1
2. Shortcuts mit dem höchsten Ähnlichkeitsrang - Grenzwert 0.15

Queryunabhängige Selektion

3. Peers mit höchsten bootstrapping Rang
4. Default Suchstrategie z.B flooding-basierte Strategie - die Nachbarn bis zu k -Peers zugefügt

- Semantik-basierte Peer-to-Peer System
- Assistent zur Suche und Managing der Bibliographischen Metadaten (z.B. von BibTex-File)
- System erlaubt dem Benutzer
 - Import eigener Bibliographische Metadaten
 - Formulieren der Anfrage – als Beschreibung wie z.B. Autor, Publikations-Typ etc. oder festlegen des Themas in (ACM Themenhierarchie)
- Routing folgt nach dem Prinzip semantischer Rangordnung
- Duplizieren von Antworten wird vermieden durch das Messen der semantische Ähnlichkeit unterschiedlicher Antworten

Bibster – semantische Suche bibliographischer Metadaten

The screenshot displays the Bibster application window, titled "Bibster - Diana". The interface is divided into several sections:

- Scope:** Includes radio buttons for "Local Peer", "Automatic Search", and "Selected Peers" (which is selected). A "Change..." button is next to "Selected Peers".
- Search:** A search box containing the word "Peer", with "Search Now", "Stop Search", and "Clear" buttons below it.
- Search Details:** A tree view showing "Publication" expanded with sub-items: Article, Book, Booklet, InBook, InCollection, and InProceedings. Below this are input fields for title, abstract, keywords, year, author, journal, and url.
- Search Results Table:** A table with four columns: swrc:title, swrc:author, rdf:type, and bibster:peer. The table lists various research papers, with the last entry highlighted in blue.
- BibTeXView:** A text area showing the BibTeX entry for the selected article, including fields like title, author, journal, type, year, volume, pages, isAbout, key, label, and issn.
- Bottom Panel:** Contains buttons for "Save", "Delete", "Copy All as BibTeX", and "Copy All as HTML".

swrc:title	swrc:author	rdf:type	bibster:peer
Technologies for Personal and Peer-to-Peer (P2P) Knowledge Management	Eric Tsui	TechnicalReport	Diana
Open Problems in Data-Sharing Peer-to-Peer Systems	Neil Daswani and Hector Garcia...	InProceedings	bibster_NTNU
Designing a Super-peer Network	Beverly Yang and Hector Garci...	InProceedings	bibster_NTNU
Chold: A Scalable Peer-to-Peer Lookup Service for Internet Applications	Ion Stoica, Robert Morris, Davi...	InProceedings	bibster_NTNU
Improving Efficiency of Peer-to-Peer Search	Beverly Yang and Hector Garci...	InProceedings	bibster_NTNU
Peer-To-Peer Data Management	Hector Garcia-Molina	Misc	bibster_NTNU
Emerging Capabilities in Intelligent Agents	Thorsten Scholz and Ingo. J. Ti...	InProceedings	de.tzi.mlo
Technologies for Personal and Peer-to-Peer (P2P) Knowledge Management	Eric Tsui	TechnicalReport	de.tzi.mlo
Peer-Mediated Distributed Knowledge Management	Matteo Bonifacio and Paolo Bo...	TechnicalReport	de.tzi.mlo
Technologies for Personal and Peer-to-Peer (P2P) Knowledge Management	Eric Tsui	TechnicalReport	de.tzi.mlo
Socialisation in Peer-to-Peer Knowledge Management	Christoph Schmitz and Steffen ...	InProceedings	de.tzi.mlo
Peer-Mediated Distributed Knowledge Management	Matteo Bonifacio and Paolo Bo...	TechnicalReport	de.tzi.mlo
A distributed algorithm for robust data sharing and updates in P2P database net...	Enrico Franconi and Gabriel Ku...	InProceedings	de.tzi.mlo
A Robust and Computational Characterisation of Peer-to-Peer Datab...	Enrico Franconi and Gabriel Ku...	InProceedings	de.tzi.mlo
Semantic Data Integration in P2P Systems. Databases, Information Systems, ...	Diego Calvanese and Elio Dam...	Article	de.tzi.mlo

```
@Article{CDD+04,  
  title="Semantic Data Integration in P2P Systems. Databases,  
  Information Systems, and Peer-to-Peer Computing: First InternationalWorkshop,  
  DBISP2P 2003 Berlin, Germany, September",  
  author="Diego Calvanese and Elio Damaggio and Giuseppe De Giacomo and Maurizio Lenzerini and Riccardo Rosati",  
  journal="Lecture Notes in Computer Science",  
  type="Article",  
  year="2004",  
  volume="2944",  
  pages="77-90",  
  isAbout="Data and ACMTopic and Systems and Database_Management and Information_Systems",  
  key="CDD+04",  
  label="Semantic Data Integration in P2P Systems. Databases,  
  Information Systems, and Peer-to-Peer Computing: First InternationalWorkshop,  
  DBISP2P 2003 Berlin, Germany, September",  
  issn="0302-9743",  
  ...
```

- Zusammenführen von Peer-to-Peer und Grid
- Semantik-basierte Peer-to-Peer Systeme