State of art

With ever-increasing information overload, Web information retrieval systems (Web-based IR systems) are facing new challenges for helping people not only locating relevant information precisely but also accessing and aggregating a variety of information from different resources automatically. Currently, new technologies for enabling precise and automatic machine processing such as semantic Web [4] and Web services are emerging and have attracted more and more attentions in recent years.

Research for possible decentralization of search engines is a significant problem of contemporary researchers. Centralized systems have a bottle neck effect and do not have effective ways to parallelize tasks of information processing.

Issues description

The term "Peer-to-Peer" (P2P) stands for architecture and a design philosophy that addresses the problems on centralized applications [2]. From an architectural point of view, in P2P network nodes operate mostly autonomously and share resources with other nodes without central control. The design philosophy of P2P systems is to provide users with a greater flexibility to cooperate with other users to form and participate in sharing.

P2P-based Information Retrieval systems (P2P-based IR systems) provide alternatives to Web-based IR systems, especially in cases when it is infeasible to transfer shared data to each node to a central repository. A typical P2P-based informational retrieval system consists of a set of nodes connected in a P2P fashion.

In comparison with client/server architecture, P2P-based systems provide a more open architecture by decentralizing the control from servers [5], allowing nodes to be loosely coupled. Also, as a centralized system has the bottle neck of accessing and its maintenance cost is expensive, scalable P2P systems are receiving more and more attention especially in the research and product development for the open and dynamic Web-environment.

However, now a lot of P2P architectures are basing on existing systems that have been developed. Very often these P2P systems are supported by centralized servers and don’t offer full-decentralization. Also they are susceptible to malicious attacks and single point of failure. [5] Moreover, the centralized server will become a bottleneck for a large number of peers, potentially degrading performance dramatically. Finally, these systems lack scalability and robustness.

Solution trends

We can use for full-decentralization new topologies of P2P systems instead of structured networks: semi-structured networks and unstructured networks [6].

In the semi-structured networks, the location of agents fellow a set of rules but they are of considerable flexibility. Examples of semi-structured networks include two-level “super-node” structure (with a privileged node, which works as local server for group of nodes) and multi-level hierarchical systems (with a query routing between few levels of nodes).

Unstructured networks resemble social networks and exhibit a number of properties including scale-free property, following the power-law distribution. In an unstructured network, each node is responsible for

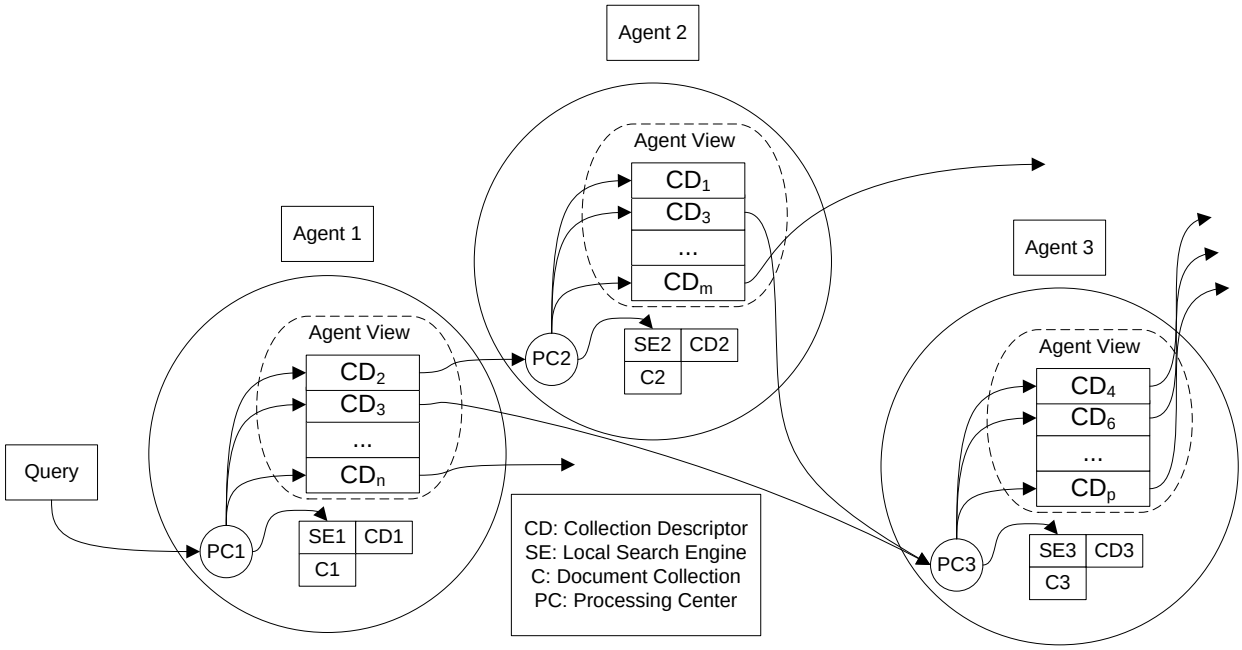
its own data, and keeps track of a set of neighbors that it may forward queries to. [3] There is no strict mapping between the identifiers of objects and those of nodes. Until the mid-2000s unstructured networks were based on the Gnutella project [1], although the search tools of this project are sharply limited. Recently, increasing research opportunities for multi-agent to build a P2P-based IR systems. [7]

My approach

In my PhD project I want to research opportunity to build P2P system with unstructured network topology as multi-agent system (MAS)which composed of multiple interacting intelligent agents. In this P2P system agents play the same role and the system has no centralized decision makers. They must cooperate to forward queries among themselves.

Each agent is composed of five components: a collection, a collection descriptor, a search engine, an agent-view structure, and a control center.

- The collection is a set of documents to share with nodes.
- The collection descriptor can be considered as the “signature” of a collection. By distributing collection descriptors to other nodes in the system, agents can know more about how content is distributed in the agent society.
- The agent control center is the unit that accepts user queries and also performs the distributed search algorithm.
- The local search engine allows each agent to conduct a local search on its document collection and return relevant documents.
- The agent-view structure, also called the local view of each agent, contains information about the existence and structure of other agents in the network.



The query language and protocol communicated between agents need to be defined. Since the semantic Web information is commonly based on RDF which is a recommendation of W3C, a standard interface

for querying and accessing RDF data is ideal for the interoperability between semantic Web information environments.

Steps of project solving

The main tasks in my doctoral are:

- Definition of agent interaction topology in semantic network approach
- Research of the possibility of using RDF for agent interaction, because it's official World Wide Web Consortium (W3C) Semantic Web specification for metadata models
- Definition of search algorithm (presumably from Random Breadth-First-Search, k-Random Walk or Generic Adaptive Probabilistic Search [7] as the most suitable for unstructured P2P systems)
- Developing of application for the simulation of the P2P Semantic Web search system
- Appraisal of system performance

References

[1] Haizheng Zhang, Victor Lesser: Toward Peer-to-Peer Based Semantic Search Engines: An Organizational Approach. 2008

[2] Quang Hieu Vu, Mihai Lupu, Beng Chin Ooi: Peer-to-Peer Computing: Principles and Applications. 2009

[3] Domenico Beneventano, Sonia Bergamaschi: Semantic Search Engines Based on Data Integration Systems. Distributed Agent-Based Retrieval Tools: Proceedings of the 1st International Workshop (Paperback). Milan, Italy. 2006

[4] Carlos Eduardo Pires, Damires Souza, Thiago Pacheco, Ana Carolina Salgado: A Semantic-Based Ontology Matching Process for PDMS (Peer Data Management System). Data Management in Grid and Peer-to-Peer Systems: Second International Conference, Globe 2009 Linz, Austria. 2009

[5] Diego Torres, Hala Skaf-Molli, Alicia Diaz, Pascal Molli: Supporting Personal Semantic Annotations in P2P Semantic Wikis. Database and Expert Systems Applications: 20th International Conference, DEXA 2009, Linz, Austria. 2009

[6] Steffen Staab, Heiner Stuckenschmidt: Semantic Web and Peer-to-Peer: Decentralized Management and Exchange of Knowledge and Information (Paperback). 2006

[7] George H.L. Fletcher, Hardik A. Sheth, Katy Borner: Unstructured Peer-to-Peer Networks: Topological Properties and Search Performance. Indiana University, Bloomington, USA. 2005