

Rámcové téma dizertační práce uchazeče o doktorské studium na FEL ČVUT v oboru Informatika a výpočetní technika

Uchazeč o studium:
Ilya Rudomilov

Navrhovaný školitel:
Doc. Ing. Ivan Jelínek, CSc.

Předběžný název dizertační práce:
Semantic P2P search engine (P2P sémantické vyhledávání na webu)

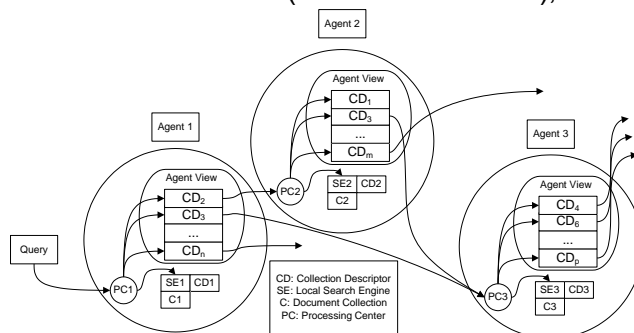
Charakteristika zaměření a cílů práce:

Moderní technologie pro vyhledávání na webu (Web Information Retrieval systems, Web-based IR systems) používají různé technologie pro automatické zpracování současného obrovského objemu informací z různých zdrojů informací nba webu, nejen pro rychlé a přesné vyhledávání. Jedním z takových přístupů, který je předmětem intenzivního výzkumu je oblast sémantického webu [1] [8]

Vyhledávání informací s využitím principů sémantického webu však naráží na řadu technických a technologických překážek. Jednou z nich je problém konstrukce a údržby vhodných ontologií, jinou je omezení vyplývající ze základního komunikačního paradigmatu komunikaci s webem klient-server. Klasické systémy klient-server se potýkají se zpracováním „paralelních“ dotazů, které vedou k enormní zátěži serverů. Tyto problémy lze vyřešit použitím technologií Peer-to-Peer (P2P) [2], která je založena na použití množství uzlů v jedné síti, se společnými pravidly práce a maximálním úrovní nezávislé výměny dat mezi sebou. P2P-based Information Retrieval systems (P2P-based IR systems) je efektivní alternativa ke klasickým Web-based IR systems.

P2P systémy jsou stále častější se schopností vytvářet autonomní, spolehlivý systém s širokými možnostmi pro rozšíření výkonu systémy zvýšením počtu uzlů. P2P sítě zpracovávají jednotné požadavky pro uzly, a umožňují rychle distribuovat zatížení mezi uzly. Ukazuje se, že použití P2P sítí pro sémantické vyhledávání je velice perspektivní směr, o čemž svědčí počet výzkumných pracovníků [3]. Bohužel, většina moderních P2P sítí funguje na základě již existujících sítí, a proto je v její architektuře centrální uzel [6] na vyvažování zátěže mezi uzly a směrování požadavků. Jelikož počet uzlů P2P systému je úzkým hrdlem řešení a jeho produktivita se prudce snižuje v takových centralizovaných P2P sítích[7].

Ve své doktorské práci bych se chtěl věnovat možnosti využití sémantického vyhledávání decentralizovaného systému P2P, který nebude obsahovat hlavní řídicí uzel. Můj návrh spočívá v modelu P2P sítě je zcela nezávislé na množství inteligentních agentů a proto tvoří multi-agentový system (MAS). Každý agent zpracovává příchozí požadavky na něj a bez dalších informací bude komunikovat se sousedními agenty [4]. Každý agent má 5 částí: „*collection*“ (soubor dokumentů), „*collection descriptor*“ (sémantická struktura interakce dokumentů na této stránce s dokumenty ostatních uzlů), „*search engine*“ (vyhledávač pro nalezení dokumentů na tomto uzlu), „*agent-view structure*“ (informace o dalším uzlu sítě P2P) a „*control center*“ (zpracování uživatelských požadavků a odeslání výsledku). Výměna dat mezi nimi bude probíhat přes standard RDF (Resource Description Framework), který byl schválen W3C (World Wide Web Consortium) jako standard pro výměnu dat mezi metamodelů v koncepci sémantického webu.



Tato téma je předmětem intenzivního výzkumu v akademickém prostředí, a má vysoký potenciál vzhledem k rostoucímu šíření P2P technologie a W3C podporovaném konceptu sémantického webu Webu 2.0.

Datum a podpis školitele:

Datum a podpis uchazeče o studium

Literatura

- [1] 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
- [2] Quang Hieu Vu, Mihai Lupu, Beng Chin Ooi: Peer-to-Peer Computing: Principles and Applications. 2009
- [3] George H.L. Fletcher, Hardik A. Sheth, Katy Borner: Unstructured Peer-to-Peer Networks: Topological Properties and Search Performance. Indiana University, Bloomington, USA. 2005
- [4] 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
- [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] Haizheng Zhang, Victor Lesser: Toward Peer-to-Peer Based Semantic Search Engines: An Organizational Approach. 2008
- [8] Jorge Cardoso, Martin Hepp, and Miltiadis D. Lytras: The Semantic Web: Real-World Applications from Industry. 2007