Procedia Computer Science

Volume 88, 2016, Pages 493–498

7th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2016

Thematic Thesauruses in Agent Technologies for Scientific and Technical Information Search

Anastasia Ananieva[*], Boris Onykiy[†], Alexey Artamonov[‡], Kristina Ionkina[§], Ilya Galin[**], Dmitry Kshnyakov [††]

National Research Nuclear University MEPhI, Moscow, Russian Federation

a_ananieva@bk.ru, bnonykij@mephi.ru, aartamonov@mephi.ru, ionkinakristina@gmail.com, ilia.galin@gmail.com, kshnyakov.do@yandex.ru,

Abstract

Diffusion process of professionally relevant information in the Internet thus increasing of access time in the dialogue mode lead to the necessity to solve problem of real-time pertinent information delivery. The article presents the solution to the problem of choosing the most important concepts for selecting texts that are relevant to the thematic direction from the general flow of the texts delivered to the system by web crawler. To solve this problem, the author introduced the concept of "generality index" of the thematic thesaurus' term. It is proposed to measure the terms value as keywords for agent search prescriptions.

Keywords: Information analysis systems, agent-based technologies, databases, table and graph models, thesaurus, laser technologies

1 Introduction

With the development of the Internet, the number of scientific and technical information sources available to the user has grown considerably. Conventional sources – libraries, journals, conference proceedings – are now supplemented by websites of universities, scientific centers, laboratories, academic subdivisions, informal working groups and, finally, specialists. These new sources of information turned out to be attractive to specialists in different subject areas, because news items on these websites appear several months earlier than in journals. Direct dialog with the authors is also

[*] Automative retrieval of subordinate infromation from news messages algorithm development
[†] The automation of agent-based search for information on thematic area in the Internet task statement
[‡] Agents' functions classification within the system and development of agent search algorithms
[§] Development of concept "term generality index"
[**] Development of Control path for automated agent thematic search
[††] Development of the Russian and English thematic thesauruses for agent-based information retrieval

Selection and peer-review under responsibility of the Scientific Programme Committee of BICA 2016

doi:10.1016/j.procs.2016.07.470

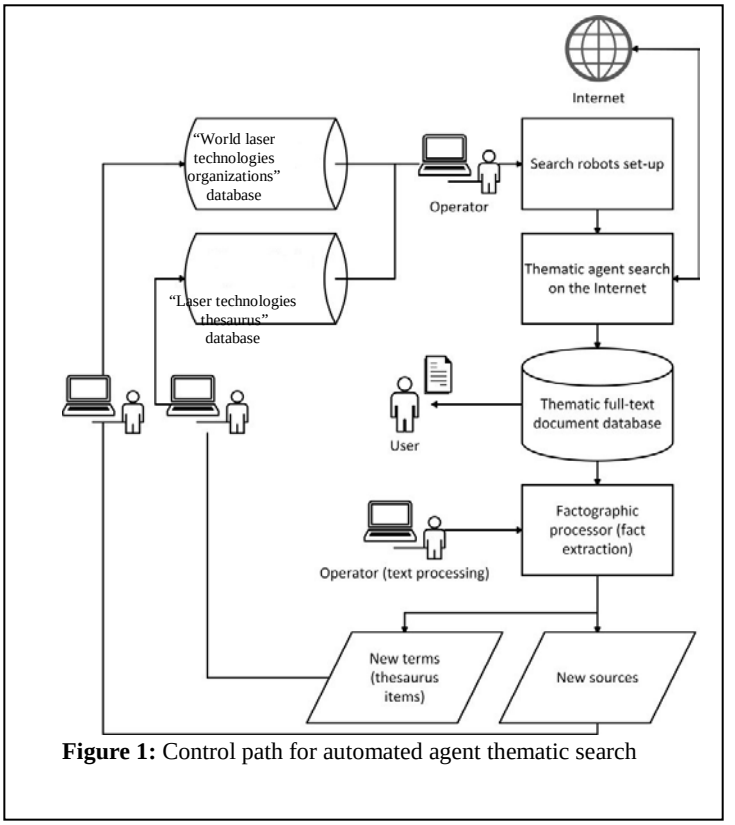

Figure 1: Control path for automated agent thematic search

available. However, along with the growth of potentially accessible data sources, information scattering takes place. As a consequence, time and labor input for information search also increase (Berners-Lee, Hendler, & Lassila, 2001).

Agent-based technologies are supposed to solve these problems (Ananieva, Artamonov, Galin, Tretyakov, & Kshnyakov, 2015). At the present day agent systems are widely used in such areas as distributed complex problem solving, collaborative product engineering, software development, etc.

This paper considers a different class of agent systems – Multi-agent information analysis systems in scientific and technical areas (MIAS). Conducting and controlling agent search in systems of this class has some essential features described in this paper.

An agent-based technology refers to a regular target-oriented automated search of information on a preset cluster of Internet websites. An agent is a network search program driven by the following data: agent owner's name and address for regular delivery of theme-related news items and applying for instructions in case of emergencies; search prescriptions – keywords from the thesaurus of the subject area in preset national languages; websites in the cluster for scanning; thematic cluster scanning timetable.

These definitions of a search agent and agent-based technologies match the definition accepted by the Foundation for Intelligent Physical Agents Standards Committee (FIPA).

Control path for agent thematic search in MIAS is represented in Figure 1.

Problems of creating control path databases were considered by authors in "Visualization of Semantic Relations in Multi-Agent Systems" (Artamonov, et al., 2014). The main issue of this paper is the basic problems of creating thematic thesauruses as databases with a specific structure. Both databases provide control over agent search: the first one sets where to search, the second one – what to search.