

MULTIAGENT INFORMATION TECHNOLOGIES IN SYSTEM ANALYSIS

**V.A. Inkina, E.V. Antonov, A.A. Artamonov, K.V. Ionkina,
E.S. Tretyakov, A.I. Cherkasskiy**

National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), Moscow

E-mail: vainkina@mephi.ru eantonov@kaf65.ru aaartamonov@mephi.ru ionkinakristina@gmail.com
estretyakov@mephi.ru aicherkasskiy@mephi.ru

Agent technologies currently play an increasingly important role in the information technology industry given its ability to learn and evolve, to solve information management problems, to employ data visualization and many other benefits. As a computer program, an agent deals with a challenge Internet users face every single day: to obtain reliable and effective data in the specific thematic field. Multiagent system consists of two or more autonomous agents and aimed at solving complex problems, such as Big Data, Data mining, primary structured and unstructured information processing (including text, numbers and multimedia types of data).

Keywords: agent, multiagent technology, agent system, Big Data, intellectual agent

Vera Inkina, Evgenii Antonov, Alexey Artamonov, Kristina Ionkina,
Evgeny Tretyakov, Andrey Cherkasskiy

Copyright © 2019 for this paper by its authors. Use
permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1. Introduction

Science and technology is one of the most significant development directions within the scope of which technologies are created, solutions are implemented that effectively respond to the grand challenges. Problem solving in this area requires specific complex knowledge on a great number of research areas in conditions of incomplete information, limited resources and lack of time.

Big Data and Data mining class of problems describe knowledge acquisition problem from a numerous multidirectional information resources.

The emergence of information retrieval systems (IRS) is a natural reaction of mankind to solve the Big Data problem [1]. But IRS are not responsible for the data reliability and relevance. Its development led to the inability of a user to master the found amount of data.

IRS filling is conducted by agent technologies - technologies using special autonomous programs for solving various problems.

2. Urgency

Nowadays in the development of intelligent information agent technologies, multi-agent technologies have been established to solve the problems of regular automatic collection and processing of structured and unstructured information in a certain thematic direction from various network sources.

There are three defining properties of Big Data – volume, variety and velocity. Referring to the amount of data, types of data and to the speed of data processing respectively. According to this 3Vs model, the challenges of big data management result from the expansion of all three properties, rather than just the volume alone - the sheer amount of data to be managed.

An agent system is a system created by an intelligent agent and is one of the Big Data problem solving [2, 3]. An agent itself in information technology or a software agent is a computer program that performs a task entrusted by a user and acts on behalf of that user.

3. Modern agent technologies

There are invested additional properties in the modern definition of an intellectual agent: autonomy, sociality, reactivity and proactivity.

Autonomy refers to the ability of an agent to act purposefully to achieve a result without external control by systems. He has control over his actions and the state of internal variables.

The sociability of an agent lies in the interaction with other agents to accomplish a common task, and for consistency in resolving conflicts to reach a consensus.

Reactivity is a reaction to receiving an event from the external environment and adjusting its behavior according to the knowledge gained.

Proactivity means a desire to achieve the goal, improving the characteristics of the internal state non-stop.

Multiagent system (MAS) is the system of autonomous intelligent agents united in a common network and their activity is also aimed at solving the tasks set by the system [4]. The use of MAS is one of the key approaches to the tasks of reducing the time for processing information and reducing the cost of cloud services, which allows autonomously, make decisions and allocate decentralized resources.

Due to its multitasking and structure, multi-agent technologies are used in many different areas: management of distributed and network enterprises, the educational process in systems of distance learning, complex and multifunctional logistics, etc.

A demonstrative example is the social network LinkedIn, which uses intelligent agent technologies in recruiting to search for people based on their career information and research activities [5].

Projects in the field of mass online education, like Coursera, use agent technologies to inform users about events within a certain course.

4. Agent acquisition of information specifics

When working in a global network, agents have to interact with diverse types of information sources [6]. The formation of knowledge in the global network is disordered; therefore, the type of data contained can classify information sources.

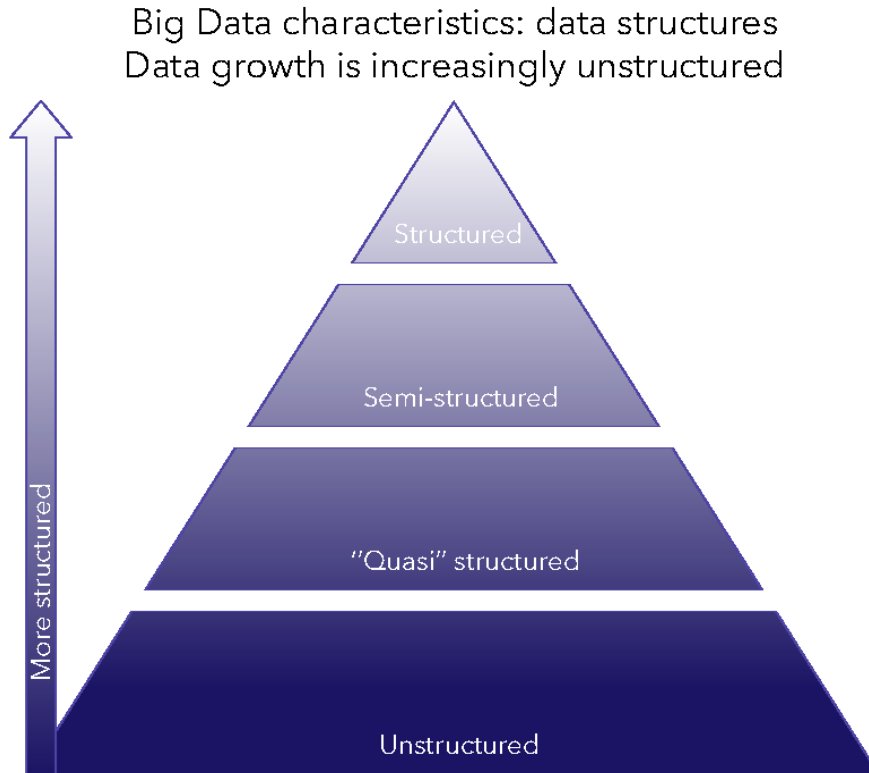


Figure 1. An infographic of data types; data growth is increasingly unstructured

Structured data: Data containing a defined data type, format, and structure (that is, transaction data, online analytical processing data cubes, traditional RDBMS, and even simple spreadsheets).

Semi-structured data: Textual data files with a discernible pattern that enables parsing (such as XML, data files that are self-describing and defined by an XML schema).

Quasi-structured data: Textual data with erratic data formats that can be formatted with effort, tools, and time (for instance, web clickstream data that may contain inconsistencies in data values and formats).

Unstructured data has no inherent structure, may include text documents, PDFs, images, and video.

The degree of structuredness of the collected data is directly determined by the variety and intellectual abilities of the agents. When collecting a news article from an information portal, an agent can collect the whole source code along with ad units and other irrelevant information, and can highlight the components of a publication and collect data in its pure form, e.g.

The development of an agent for each information resource has some difficulties because of the lack of strictly regulated standards for constructing information resources and because of high flexibility of presenting information using html code. Therefore, different algorithms and standards are created for the placement of information for its correct identification: HTML 5.2, Selectors Level 3, ITS Version 2.0, etc.

Time passed by, and a problem arose in the use of agent technologies because the collection of data may occur against the desire of the internet resources owners, thereby the agent, without realizing it, could infringe copyrights.

In U.S. court practice; there have been cases where information portal owners have brought cases against collecting agents. For example, in case №17-cv-03301-EMC of United States District Court Northern District of California hiQ Labs inc. indicted LinkedIn for the fact that the social

network has built protection against the collection of publicly available profile information in reference from United States Antitrust law. This case was initiated in response to the hiQ company's early accusation of illegitimate automated collection of LinkedIn profile information in reference from Computer Fraud and Abuse Act. Subsequently, a case judge Edward M. Chen obliged LinkedIn to remove the previously established protection on the collection of publicly available information by autonomous intelligent agents.

Afterwards the use of agents to collect information led to the practice of implementing the Robots Exclusion Standard (robots.txt, 1994). This is a file with a list of access restrictions for robots to content of an http server. The file should have a path relative to the site name /robots.txt. It reflects the availability of information to be collected by an agent.

5. IT environment research

To build multi-agent systems, it is necessary, among other things, to conduct a study of the information environment[7]. It allows user to determine the structure of the information source, the object of interest and the list of necessary software. Accordingly, three stages can be distinguished.

5.1 Reconnaissance

At this stage, it is necessary to determine the available information about the object of interest and the legitimacy of its collection. It is also important to identify the available attributes of the object of interest in order to build the corresponding data model and relationships with other objects of the information source.

For example, in social networks where the person is the object of interest, communities, audio and video recordings can play the role of objects associated with it. In such cases, it would be necessary not only to collect information about objects of interest, but also to take into account the relationship with other objects, since in the future they can act as characteristics during intellectual processing.

5.2 Crawling

The stage includes the study of mechanisms for placing information about objects of interest into an information source and the definition of an agent behavior algorithm for collecting information with maximum completeness. Any site is a set of interconnected web pages and its structure can be represented in the form of a hierarchical model tree, where the top of the tree is the initial page of the site and pages with the necessary information is the basis of the tree.

5.3 Configuration

The purpose of the configuration stage is to list third-party software solutions to expand the functions of the agent. In particular, the Competitive Systems Analysis Department at MEPHI uses Python 3.7 as the programming language: on which the agent is implemented.

The essence of this method of collecting information is to obtain hypertext documents in html format using the http / https protocol and further identify the necessary data from the received documents. To develop such an agent modules are required, that allow you to access the information resource using the http / https protocol and a module for recognizing the contents of an html document. At this stage, the source of information considered fully explored.

References

- [1] Garcia, A.J., Toril, M., Oliver, P., Luna-Ramirez S., Garcia R. Big Data Analytics for Automated QoE Management in Mobile Networks // IEEE Communications Magazine Volume 57, Issue 8, August 2019, Pages 91-97
- [2] Xiao Y., Liu Q. Application of Big Data Processing Method in intelligent Manufacturing // Proceedings of 2019 IEEE International Conference on Mechatronics and Automation, ICMA 2019, Pages 1895-1900 16th IEEE International Conference on Mechatronics and Automation, ICMA 2019; Tianjin; China; 4 August 2019 - 7 August 2019

- [3] Kulik S.D. Model for evaluating the effectiveness of search operations // Journal of ICT Research and Applications Volume 9, Issue 2, 2015, Pages 177-196
- [4] Ananieva, A.G., Artamonov, A.A., Galin, I.U., Tretyakov, E.S., Kshnyakov, D.O. Algorithmization of search operations in multiagent information-analytical systems // Journal of Theoretical and Applied Information Technology Volume 81, Issue 1, 10 November 2015, Pages 11-17
- [5] Artamonov A.A., Ionkina K.V., Kirichenko A.V., Lopatina E.O., Tretyakov E.S., Cherkasskiy A.I. Agent-based search in social networks // International Journal of Civil Engineering and Technology Volume 9, Issue 13, December 2018, Pages 28-35
- [6] Onykiy B.N., Artamonov A.A., Tretyakov E.S., Ionkina K.V. Visualization of large samples of unstructured information on the basis of specialized thesauruses // Scientific Visualization Volume 9, Issue 5, 2017, Pages 54-58
- [7] Kulik S.D. Factographic information retrieval for semiconductor physics, micro - And nanosystems // IOP Conference Series: Materials Science and Engineering Volume 498, Issue 1, 16 April 2019