International Journal of Soft Computing 10 (3): 220-225, 2015
ISSN: 1816-9503

Conceptual Model of Monitoring Information on the Internet

[1]Vladimir Averchenkov, [1]Sergey Roschin, [1]Mikhail Rytov,
[1]Andrey Averchenkov and [2]Valery Kamaev
[1]Bryansk State Technical University, Boulevard 50 years October 7,
241035 Bryansk, Russia
[2]Volgograd State Technical University, Lenin Avenue 28, 400005 Volgograd, Russia

Abstract: This study consider approach for developing fundamentally new and better monitoring systems. In detail, this research consider the general conceptual model of the monitoring systems and its main functional blocks, processes, consider the mathematical model and the subject area model. We will also pay attention to weak and poorly developed places of such systems and the possible ways of solving the problems. The main approach to the solution these problems in the study, the knowledge forming system that should provide the following functionality: a description of the subject area; the monitoring of the problem-oriented information; pre-processing the collected information; indexing information; structuring information; the storage and management of information in the data warehouse; the selection of information relevant to user's query; provision of information to a user in the form, suitable for reading and analysis. The principle of the knowledge forming system is as follows. Periodically using the control module of the system launches the new modules for monitoring and analytics. The purpose of these modules is to find new and updated information resources on the internet relating to the subject of the subject area which is oriented system and adjust the data warehouse system and knowledge base in accordance with the current state of information resources. In parallel with modules for monitoring and analytics the representation module is executed. The purpose of this module is to use the data warehouse and knowledge base to select the data corresponding to the needs of the users and to present information in a form which is easy for both perception and analysis.

Key words: Knowledge forming system, monitoring information on the internet, subject area model, mathematical model, indexing process, clustering process, classification process, knowledge base, data warehouse

INTRODUCTION

In order to analyze the continuous considerable volume of information flow on the internet a lot of different intelligent software systems are being used. A large number and variety of information meant for monitoring operation on the target system (the selection and analysis of text, numerical data, images, sounds, information on events or objects, etc.) as well as for internal structure, based on a wide variety of methods for sampling and analyzing information, including many of the methods of artificial intelligence (analysis of natural language, pattern recognition, data mining, multi-agent systems, etc.) complicates the analysis of such systems which is needed for developing fundamentally new and better monitoring systems. Because of the high relevance of the development of new systems for monitoring internet information in this study, we consider the general conceptual model of the monitoring systems and its main functional blocks, processes. We will also pay attention to weak and poorly developed places of such systems and the possible ways of solving the problems.

KNOWLEDGE FORMING

The process of knowledge forming (the acquisition of knowledge-based models, methods and algorithms for data analysis) in existing systems of intelligent processing of information is either missing or not fully represented, primarily due to the high complexity of such an approach.

On the basis of the above-stated, the conclusion has been reached about the relevance of the development of

Corresponding Author: Vladimir Averchenkov, Bryansk State Technical University, Boulevard 50 years October 7, 241035 Bryansk, Russia

a new type of systems knowledge forming systems (Fig. 1) which will allow the use of much wider potential of the internet: based on models, methods and algorithms to analyze information in automatic mode to gain knowledge on a particular subject.

Knowledge forming system is an intelligent information system which performs a specially organized, systematic monitoring of the problem-oriented data and mining knowledge used on models, methods and algorithms, based on an integrated approach and consideration of relationships between data. The most significant source of automated execution of these processes is now the global computer network the internet. The knowledge forming system should provide the following functionality:

- A description of the subject area
- The monitoring of the problem-oriented information

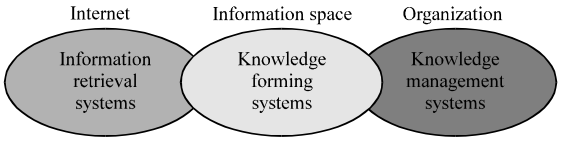

Fig. 1: Communication knowledge forming systems with intelligent information processing systems

- Pre-processing the collected information (cleansing, matching, aggregation, etc.)
- Indexing information
- Structuring (clustering, classification) information to build on the basis of the data warehouse
- The storage and management of information in the data warehouse
- An understanding of the request formulated by the user but the user does not have to know the principles of information organization in the system
- The selection of information relevant to user's query, represented by multimedia documents
- Provision of information to a user in the form, suitable for reading and analysis

The structure and the generalized algorithm of the knowledge forming system: The structure of the knowledge forming system consists of the following basic parts: monitoring module, analytics module, representation module, control module, knowledge base and data warehouse (Fig. 2).

Consider the purposes of the blocks and the basic processes underlying the functioning of the intellectual analytical information system. One of the most important processes in the system is the information monitoring

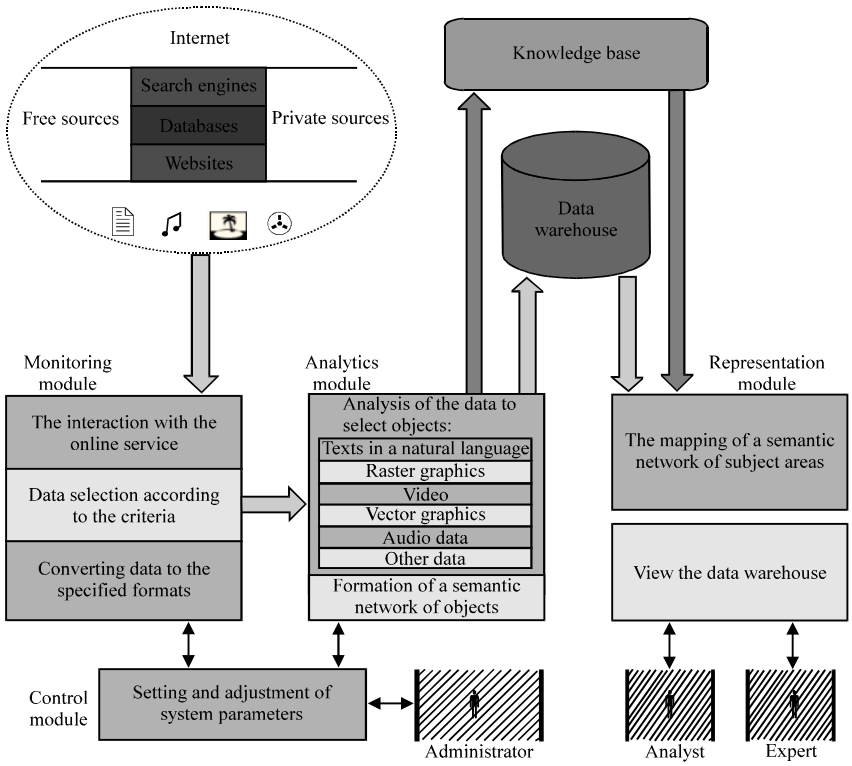

Fig. 2: The structure of the knowledge forming system

which is implemented by monitoring module. The purpose of this module is to select from both public and private (in which access is accomplished through authentication) sources on the internet all the information corresponding to the subject area (Leonov *et al*., 2014; Averchenkov, 2012). The key tasks at this stage are: interaction with the online service; data selection according to appropriate criteria; converting the data into formats of presentation adopted in the system. In order to ensure quality monitoring of problem-oriented information on the internet, the theory of algorithms and methods of artificial intelligence is used: the multi-agent systems and traversal algorithms. Comprehensiveness of information resources in the system is achieved by the implementation of the two approaches, the essence of which is as follows: on a periodic basis the system, sequentially interacts with all the registered (external to it) means of information retrieval and thus, selects the problem-oriented information. Additionally, the system provides the ability to specify the addresses of specific information resources relevant to subject area which are also investigated by the system.

The second important and most complex module of knowledge forming system is analytics module. The data in this module comes from the monitoring module. At the core of the module are several data mining processes, based on the models, methods and algorithms. The main processes of the module analytics are indexing, clustering and classification of the information (Buettcher *et al*., 2010; Baeza-Yates and Ribeiro-Neto, 2011).

The indexing process main goal is to produce search images of documents which are necessary to the knowledge forming system for clustering and classification of the information and the selection of documents matching the user's query through the search for information in the system. The inputs to this process are incoming documents, comprised of different types of information (text, pictures, sounds, etc.). Algorithms implement this process to highlight key features (objects) in the information (including the assignment of a weight to a feature/an object, reflecting the expected value) which form the basis of search images to documents. The main difficulty of this process is to produce search images to multimedia documents (images, video, drawings, sounds, etc.), due to insufficient knowledge of the issue and often a lack of effective (time and quality) algorithms (methods) to process this kind of information.

In the clustering process, the problem-oriented information, based on the available search images of documents is automatically analyzed. It's done in order to identify groups of similar documents and the relationships between them. As a result of this process, a semantic network of objects (groups of documents in which these objects are available) of the subject area or fragments of a semantic network is created and runs the correlation established network (or fragments) with basic semantic network of knowledge forming system if it has already been previously created.

The classification process of information allows to correlate the system selected problem-oriented documents from the existing model of subject area described by the expert or received for example, on the basis of standards (classifications). This model is represented in the knowledge forming system in the form of a separate semantic network. This approach allows organizing selected information in a familiar understanding of the structure which in some cases makes data analysis easier.

At the final stage of the analytics module, the problem-oriented documents are put in a data warehouse analytics and their corresponding search images (assigning objects) are entered into the knowledge base. The resulting semantic network, describing the objects of subject area and their relationships is also put into the knowledge base.

Using the knowledge forming system which accumulates knowledge on a particular subject area and classifies documents, is carried by the end users via the representation module. Using this module analyst formulates the request which is performed based on the selection of relevant information. The system's responses to the query are fragments of constructed semantic network which can be considered as an information model studied by the user task. The system allows users to navigate through the extended fragments of the network and view the documents to which they correspond.

The effectiveness of the knowledge forming system is largely determined by the composition of its formalized knowledge. The knowledge base in the system is the link between the data warehouse and representation module with which the user interacts. Filling the knowledge base is in automatic mode using the analytics module based on the models, methods and algorithms for data analysis.

The main component of a knowledge based system is the semantic network. This unit is used to describe the subject area and for formalization of knowledge about groups of data warehouse documents. The contents of the block a semantic network whose nodes correspond to the sets of terms (objects/features) and groups of documents related to a particular topic and the arc the relationship between them. The semantic network can

improve the process of information retrieval. It classifies Internet documents, relating them to specific nodes and provides a significant narrowing of the search range of the required documents which positively affects the accuracy of search results.

Also, the knowledge base should include a block of metadata or data about data. This is information about what data in data warehouse is how it relates, what the structure of the data is, where the data is stored, how to access it, etc. Modules of the system maintaining and exploiting the data warehouse have the access to metadata. This unit is the basis for all processes in the data warehouse and it provides end users the access to the information in the system.

Data warehouse is the most important internal element of the system, containing information resources associated with a specific subject area (Kamaev, 2014; Rajaraman and Ullman, 2011; Witten *et al*., 2011).

Data warehouse is a complex integrated system. In addition to the database, data warehouse of system includes a sophisticated infrastructure (Russell, 2013; Han *et al*., 2011):

- Means of changing and expanding the database
- Technology of regular replenishment of data
- Checking, cleaning and reconciliation tools
- Technology of aggregation and consolidation of data
- Querying, reporting and analysis in real time tools
- Means of access rights, etc.

Staff using and servicing the system are analyst, expert and administrator. Analyst is an employee of the company which introduced a knowledge forming system. They are mostly leaders of various levels: the chief of the department, division, deputy director and director as well as a specialist in the field of information analysis. On the basis of information obtained from the system the user can make decisions or prepare information for decision making at a higher level.

Expert, this is the person, who can clearly express his thoughts and enjoys a reputation for knowing how to find the right solutions to problems in a specific subject area. The expert uses his techniques and tricks to make search solutions more effective. The main function of the expert in the system is to control the correctness of the construction of a semantic network and analysis of information gathered by the system at a better level with a goal of deeper understanding the issue and consulting analysts in case the latter need help with decisions.

Administrator is a person who performs several important functions in a centralized way. The main ones are: the installation and configuration of the system, controlling the system, control over its operation, ensuring the safety of the knowledge forming system and the optimal functioning of the entire system. In order to help the administrator to perform some of the functions described herein the control module was. An administrator should be a highly skilled professional who has to possess an excellent knowledge of the functionality of the system; the technology of the user experience with the system; the technology of system maintaining. The generalized principle of the knowledge forming system is as follows:

Periodically using the control module of the system launches the new modules for monitoring and analytics. The purpose of these modules is to find new and updated information resources on the internet relating to the subject of the subject area which is oriented system and adjust the data warehouse system and knowledge base in accordance with the current state of information resources.

In parallel with modules for monitoring and analytics (and independently of them) the representation module is executed. The purpose of this module is to use the data warehouse and knowledge base to select the data corresponding to the needs of the users (analyst and expert) and to present information in a form which is easy for both perception and analysis. Meanwhile in response to the query, only those documents are given information on which is already known to the system at the time of the request that is this data is indexed, recorded in a data warehouse and knowledge of them are in a semantic network system.

The mathematical model structure of the knowledge forming system: The generalized structure of the mathematical model of the knowledge forming system includes a subject area model and the model of the processes of monitoring, indexing, clustering, classification and presentation of information (Fig. 3).

The functioning of the knowledge forming system requires a formal presentation and comparison of documents and queries as a means of describing documents and queries and the process of searching.

The subject area model in the knowledge forming system is specified using a semantic network. Knowledge representation in the system through a semantic network allows to store conceptual knowledge about objects of the subject area and the relationships among them as well as to modify knowledge of the system.

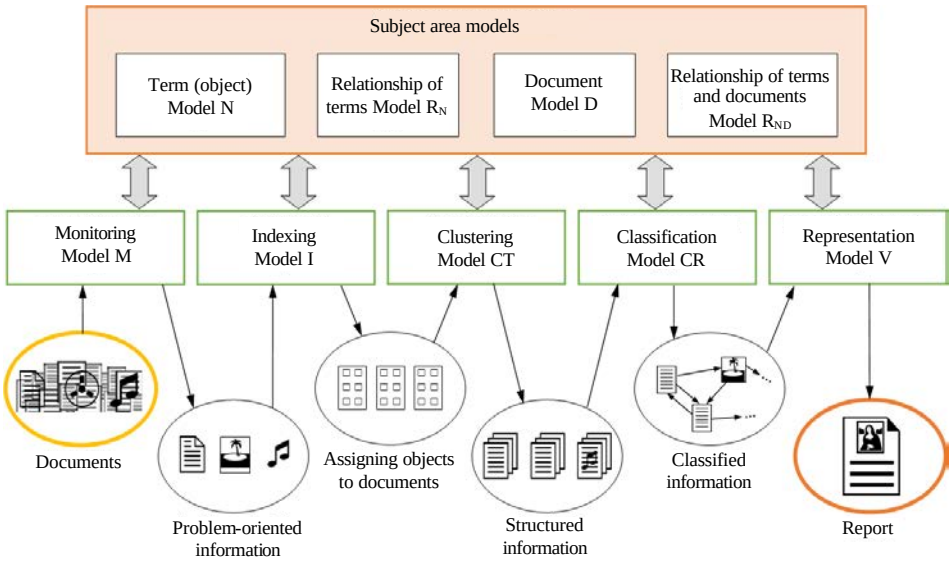

Fig. 3: The structure of the mathematical model of the knowledge forming system

A semantic network is a directed graph of two components: nodes and arcs connecting them. Each node represents a term (object) and the arc the ratio between a pair of terms (objects):

$$S = <N, R_N> \quad (1)$$

Where:

S = The semantic network

N = $\{N_1, N_2, \ldots, N_K\}$ is the set of elements that reflect the terms (objects) subject area

R_N = $R_N(N_i, N_j)$ is the ratio that specifies the type and strength of the interaction between terms (objects)

To determine the relationship of documents and terms in a formal subject area model, we introduce additional components:

$$S = <N, R_N, D, R_{ND}> \quad (2)$$

Where:

D = $\{D_1, D_2, \ldots, D_L\}$ is the number of documents related to a specific subject area

R_{ND} = $R_{ND}(N_i, D_j)$ is the ratio of associates subject are a terms and documents

Subject area terms which are elements of the semantic web are defined by cortege:

$$N = <C_N, R_S, R_I, R_A> \quad (3)$$

Where:

C_N = The name (the image) of the element

R_S, R_I, R_A = Different relationship types between model elements

Each relationship that defines the relationship between the terms is determined by the model:

$$R_N = <C_R, V_R, N_{RS}, N_{RD}, S_R> \quad (4)$$

Where:

C_R = The name relationship name

V_R = The type of the relationship: synonymous, genus-specific, associative

N_{RS}, N_{RD} = The terms that there is a relationship between directed from the element N_{RS} to N_{RD}

S_R = The strength of the relationship, depends on the type of relation ship V_R

Documents in the knowledge forming system are the objects of search. The model of the document in system is:

$$D = <F_D, V_D, T_D> \quad (5)$$

Where:

F_D = FN, FP, FD, FS is the information about the document file: file name FN, the path to FP, created date FD, size FS

V_D = The type of document: text, image (picture, photograph, sketch), the animated image (animation, video), sound (song, speech, melody), drawing, multimedia document

T_D = $\{t_{Di} | i = 1, \ldots, M\}$ is the set of lexical items of the document depends on the type of document V_D

The model that defines the relationship between documents and terms:

$$R_{ND} = <N_R, D_R> \quad (6)$$

where, N_R, D_R are, respectively the term of subject area and document, between which a relationship. Semantic network of the subject area has a simple mechanism output which positively affects the efficiency of the system.

CONCLUSION

The knowledge forming system is universal in the sense that the approaches used in it do not depend on a particular subject area. The system can be configured to research with information from a wide range of subject areas. The implementation of the proposed approach allows for efficient access to large arrays of distributed unstructured information, including the internet.

The conceptual model of the system monitoring problem-oriented information on the internet presented in this study allows to look for this type of systems with common positions and to identify key areas for development, the most important of which is the indexing of textual and multimedia information, research data clustering methods, the analysis of models of efficient data representation and knowledge, the use of cognitive techniques for providing information for users to display the results of the system.

The solution of the designated task set will allow intelligent processing of data on a qualitatively new level. Thus, it will make possible the following: execution of the search text, graphics, audio and other information in accordance with laid down in it with meaning, recognition of objects in images, conducting quality clustering and classification of information as well as many other tasks of data processing.

REFERENCES

Averchenkov, V.I., 2012. Smart search on the internet for the formation of domain-specific knowledge. Proceedings of the 7th International Scientific Conference on Information Technologies in Industry, October 30-31, 2012, UIIP, NASB, Minsk, pp: 109-110.

Baeza-Yates, R. and B. Ribeiro-Neto, 2011. Modern Information Retrieval: The Concepts and Technology Behind Search. 2nd Edn., Addison-Wesley Professional, New York, ISBN-13: 978-0321416919, Pages: 944.

Buettcher, S., C.L.A. Clarke and G.V. Cormack, 2010. Information Retrieval: Implementing and Evaluating Search Engines. MIT Press, USA., Pages: 632.

Han, J., M. Kamber and J. Pei, 2011. Data Mining: Concepts and Techniques. 3rd Edn., Morgan Kaufmann Publishers, USA., ISBN-13: 9780123814791, Pages: 744.

Kamaev, V., 2014. Knowledge discovery in the SCADA databases used for the municipal power supply system. Proceedings of the 11th Joint Conference on Knowledge-Based Software Engineering, September 17-20, 2014, Volgograd, pp: 1-14.

Leonov, E.A., V.I. Averchenkov, A.V. Averchenkov, Y.M. Kazakov and Y.A. Leonov, 2014. Architecture and Self-learning concept of knowledge-based systems by use monitoring of internet network. Proceedings of the 11th Joint Conference on Knowledge-Based Software Engineering, September 17-20, 2014, Volgograd, pp: 15-26.

Rajaraman, A. and J.D. Ullman, 2011. Mining of Massive Datasets. Cambridge University Press, UK., ISBN-13: 978-1107015357, Pages: 326.

Russell, M.A., 2013. Mining the Social Web: Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub and More. 2nd Edn., O'Reilly Media, USA., ISBN-13: 978-1449367619, Pages: 448.

Witten, I.H., E. Frank and M.A. Hall, 2011. Data Mining: Practical Machine Learning Tools and Techniques. 3rd Edn., Elsevier, New York, ISBN-13: 9780080890364, Pages: 664.