# Информатика и вычислительная техника

## INFORMATIONAL SUPPORT OF DECISION-MAKING BASED ON INTELLECTUAL ADAPTIVE FORUM

*A.A. Kozko*, *alkozko@mail.ru*
*Chelyabinsk State University, Chelyabinsk, Russian Federation*

This paper focuses on the issue of online forum information overload. Web forum technology are considered, the disadvantages of existing implementations which leads to the problem of information overload are defined. Also, the paper described the existing methods of solutions to this problem. The algorithm of structural and semantic analysis on the web forum that allows combine messages into logical units (subtopics) was offered. Structural web forum adaptation methods based on this algorithm, to automatically structure the web forum posts, according to their semantic content, was proposed. The prospects of using this approach to deal with information overload were shown.
*Keywords: semantic analysis, online forums, information overload.*

### Introduction

Modern information technologies offer a variety of methods for mass online communication and collaboration – forums, question-answer sites, blogs, comment systems and others. With the development of communications, the total number of Internet users and users of Internet forums in particular has increased. If at the beginning of development of the Internet the number of users of forums comprised of hundreds and thousands, now it is millions of people. This has naturally led to an increase in large-scale involvement of online discussions and increased the amount of information that these users create.

However, the lack of tools to work with large amounts of information on the forum makes it difficult to use it as a means of communication and cooperation, as well as a source of information to support decision-making. This is due to the so-called information overload effect [1], which manifests itself in the fact that people can't quickly orient in a large number of messages and find the necessary information.

At that, forums are still actively used for interaction on various issues – from entertaining conversations on abstract topics to collaboration in the fields of education, medicine, science, and technology. Forums and comment systems are also used as part of automated information management systems at enterprises and in universities. For example, taking into account the main directions of universities' automation [2], it may be noted that the online interaction tools are applicable to the control and direction of the educational process, storage and provision of information resources, as well as user support.

However, the effectiveness of the use of forums in their current form for users to interact with each other has been sharply reducing because of the large number of duplicate topics, meaningless messages, questions that remain unanswered, and other problems. This fact speaks about the need for radical changes in instruments of processing and presenting the information contained on the forums, which require the development of new intellectual methods of its analysis and transformation, tailored to the specifics of Internet forums.

### 1. Forums Structure

As noted above, there are four major means of mass online interaction: forums, question-answer sites, blogs, comment systems. All of these tools have a similar structure, and differ only in focus on various aspects of online discussions [3]. This allows us to consider all of these means on an example of forums, as the most popular tool for mass collaboration.

When considering forum the following terms and concepts are usually used:

• post – message on the forum created by one of the users, usually text, but can also include images, links, etc.;

• topic (or theme) – a set of posts;

• section (or part) – a set of topicsunited on some basis.

Structurally, the forum is presented as a tree, the top of which is the root page, where the sections are placed. Sections, in turn, contain topics. Each topic consists of a root post and reply posts, which are arranged in a tree and, accordingly, form a branch of the discussion. At that, reply-post tree can be represented in an explicit or explicit form. In the case of the explicit representation, a tree structure is displayed in the user interface "as is", and one can easily visually trace the question-answer communications between the posts. In the case of explicit tree representation posts in the interface (and usually in the database) are arranged linearly one behind the other in chronological order, and the question-answer tree structure is possible to be restored only by analyzing the message texts and basing on specific attributes contained in the reports, such as quote or reference by user's name. Restoration of the tree structure of postsfrom the linear one can be seen in the following work [4]. More clearly examples of the tree and linear posts structures are shown in Fig. 1.

The described structure is defined by administrators and forum moderators in the part relating to the sections, as well as by the users themselves in the part relating to the creation and placement of topics and posts.

Such a structure is quite rigid, is difficult to change and rebuild, and can't adapt, therefore does not always correctly reflect the semantic content of the forum and its topics, due to the fact that each topic may contain meaningless, junk messages, and subtopics – a whole branch of discussions on the topic other than specified in the root post or topic, more specific than the specified in the root. This is clearly shown in Fig. 2.
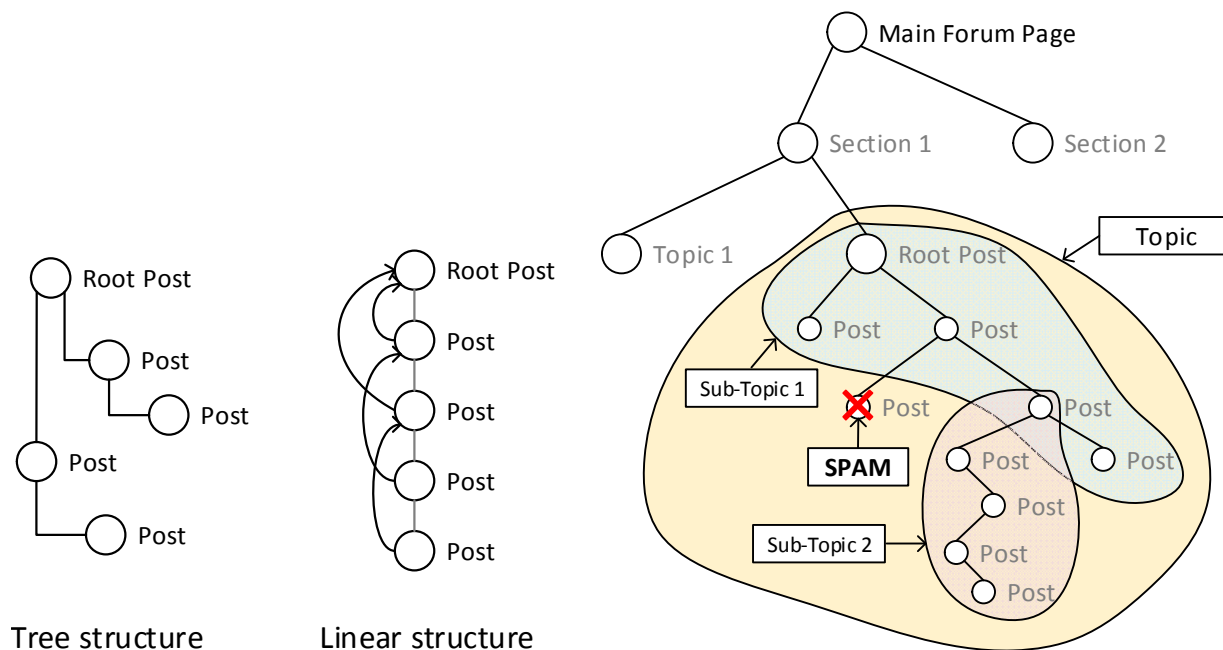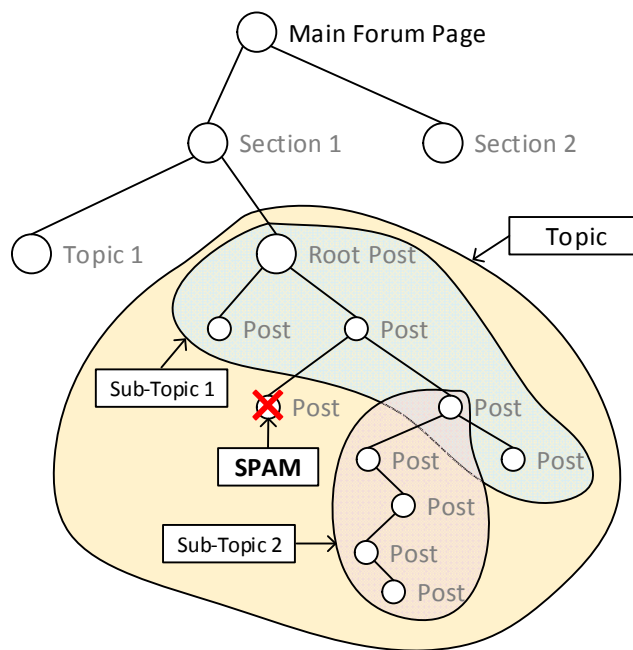


Fig. 1. Examples of thread structures



Fig. 2. Subtopics into Topic

All this leads to the duplication of the content on the forum and "smearing" of discussions into several topics. Because of this there are serious difficulties in the use and analysis of information on the forum.

### 1.1. Classification of Posts on the Semantic Content

Each position can be assigned to one of three classes:

● meaningless post, not bearing semantic load (in forums such messages are usually called "Flood");

● undefined post, by the content of which is impossible to determine its informational value, for example, short answers like "Yes" or "No";

● informative post, this post contains enough information to determine its thematic focus, such posts can be divided into two types:

– with relevant content – such posts are semantically linked with others from the same branch of the discussion and are part of the general discussion;

– with foreign content – such posts are semantically close to the adjacent message in the thread, their meaning is quite different from the general subject of discussion, and they can be called junk (in terms of forums – "Flame").

## 2. Existing Approaches

The problem of information overload on the online forums has been long time existing, but in most cases it can be solved by manual intervention of administrators and moderators of the forum in order to maintain the integrity of the discussions, however, their resource is very limited. The research of such an approach is presented in the work [5]. Similarly, there are attempts to solve this problem at the level of different forum tools, such as, for example, labels (tags), such opportunities are now added to some forum engines,and the effectiveness of using such meansis studied in the work [6].

There are several works on methods of topics analysis, each of these methods can be used to simplify the work with the information on the forum. For example, the works [7] are devoted to various algorithms of drawing up a brief summary of forum topics, and work [8] considers the principles of operation of such algorithms in general. Works [9, 10] examine questions of search of duplicate posts and topics in the forum, as well as the questions in question-answer systems. Solutions of the general problem of evaluating the quality of information in a topic for ranking are proposed in [11]. Another group of works is devoted of topic recommendation to users based on the content and model of the user's interests. Finally, the study [12] considers methods of filtering unwanted for user posts.

However, a comprehensive solution that would help reduce information overload and increase the efficiency of access to information on the forum was not so far suggested.

## 3. Information Management System Based on an Intellectual Forum

To solve the problem of information overload it is necessary to make the transition from traditional forums to the development of intellectual processing and presentation of the information contained in the forum, with the aim of using it for the tasks of information support of decision-making. An important feature of the proposed system is the ability to automatically restructure the information depending on the topic and its ranking in accordance with specific user information requests.

To achieve these goals it is necessary to solve several tasks:

1. Introduction of metrics and development of basic algorithm for determining the semantic relatedness of posts.

2. Development of forum restructuring methods on the basis of data of semantic similarity of individual messages and discussions.

3. Software implementation of a system based on these algorithms.

## 4. Determination of Semantic Relatedness of Forum Posts

To assess the relatedness of forum posts among themselves by semantic features it is required to insert a metric value that would show how the two messages are related to each other within the meaning.

To determine this metric we propose the following algorithm:

1) to determine key terms;

2) to determine the proximity between the key terms;

3) to estimate the proximity of posts based on the proximity between the key terms in the message.

When selecting a class of methods for determination of key terms and definition of the connection between them it is necessary to take into account specific forums. Forum posts are usually short (a few sentences) or even over short (a few words), and they may contain enough specific terminology characteristic of completely different subject areas, which depends on the thematic focus of the forum or its section. Also, it is important to take into account the fact that the terminology is constantly changing, expanding and updating over time.

Taking into account these facts, it is necessary to abandon the text-based methods, because search and maintaining text document collections up to date, which will fully enough describe a wide range of subject areas, is very resource-intensive task. These methods include methods of TF-IDF [13] and their

more recent modifications, for example, TF-ICF [14], used to extract key terms, as well as the semantic similarity evaluation methods, for example, LSA [15]. For the same reason, the use of ontology-based semantic similarity assessment methods are complicated, which include, for example, Resnikmeasure, Lin measure and others.

For efficient analysis of forums methods based on full enough and thus resupplied data source are required. Such source is the open encyclopedia Wikipedia. Its collection of articles, of course, can be used as body for the already mentioned above TF-IDF and the LSA methods, however, important features of Wikipedia are article markings and presence of the tree category by which the articles are structured. Accounting these features can create algorithms more effective compared to those based on only untagged bodies.

By algorithms extracting key terms, based on Wikipedia one can mean keyphraseness algorithm [16], and to evaluate the semantic proximity and connectivity between the terms one can use WLM techniques [17], WikiRelate [18], ESA [19] and others, but it's adviseable to use the ESA method (Explicit Semantic Analysis), because it is rather simple to implement, does not require disambiguation, and also shows greater accuracy on the test patterns.

### 4.1. Semantic Distance Between Messages

On the basis of information about the key terms which are contained in the posts, and the evaluation of their semantic relatedness we can estimate the semantic distance between messages as the product of the normalized distance between keywords that are contained in each of the posts:

$$dist(M_1, M_2) = \prod_i dist_{norm}(k_{1i}, k_{2i}),$$

where $M_1$, $M_2$ areposts, $k_{1i}, \ldots, k_{ni}$ – key terms, $dist$ – distance, reciprocal of proximity, is calculated as $dist(k_1, k_2) = 1/sim(k_1, k_2)$.

### 5. Analysis of Posts Structural Features

As we have noted, the forum, as well as each topic, has its own structure. The structure of a topic is tree-like, but this tree relationship between the posts can be internal and are linear, however, the location of the messages in the post tree is not the only structural feature, except that there are others. Thus, besides the textual content, each post also has a set of structural features, each of which may be used in the information forum analysis:

1) the location of the post in the answer tree;
2) the author of the post – including the name and all other features of the user;
3) the time of post creation;
4) the citation of the text of another post – the citation can be plain text and special structural element;
5) references to another user – similarly can be text or a special structural element;
6) inserted hyperlinks, images and videos;
7) the evaluation of the post by other users – available not in all forums.

Thus each position is represented as a tuple of structural and semantic features $\{x_{01}, \ldots, x_{0N}, x_{11}, \ldots, x_{1N}\}$, where $x_{01}, \ldots, x_{0N}$ are semantic features which characterize the text of the post, and $x_{11}, \ldots, x_{1N}$ – structural features describing post.

Each of the features listed above finds its application in the analysis of posts and topics in the forum, part of the features are used to restore the question-answer tree structure, while others – for the communication of topics in subtopics. Information about the participants of the discussion (authors of the posts) may be one of the criteria for assessing the proximity of the two topics, and data on the time of posts creation – the basis for the detection of the most important and relevant current topics.

### 6. Methods of Forum Restructuring

To transform the structure of the forum in order to improve access to information, we offer the following group of methods.

### 6.1. Hiding Posts

Posts that do not carry the meaning ("Flood") or which content is semantically extraneous ("Flame") are necessary to be hidden from the branches of discussions and topics, as such messages only increase information noise without adding utility to discussions.

In addition to the semantic features of each post, one can also take into account the structural features in the analysis – most undesirable posts are more often terminal nodes of the answer tree, or nodes that have only one answer, and most often having negative connotation. In addition, posts from users who have already participated in this discussion will be meaningful more likely than messages from users who have not participated in the discussion.

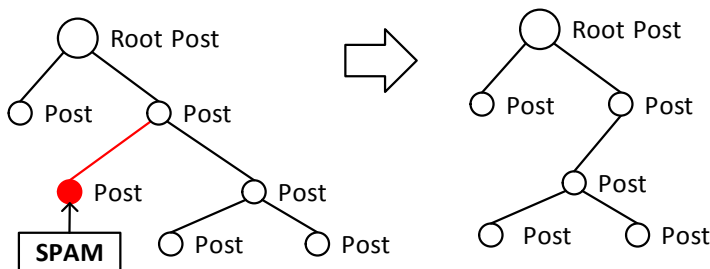Visually, the example of hiding posts from the answer tree is shown in Fig. 3.



**Fig. 3. Hidding junk posts**

### 6.2. Aggregation of Multiple Posts into a Subtopic

Each topic contains a root position from which the discussion began and posts that are replies to a root post or on one of the posts that were created previously. In this part of the topic posts can be a discussion of the topic, other than stated in the initial post, or on the contrary, the more detailed topic than the rest of the discussion. In this case, one can not clearly determine whether such group of posts is undesirable and whether to hide it from the general discussion, however, such post group can be considered as an independent unit of meaning – subtopic. Subtopic may contain several branches of posts, at that the post which is the beginning of subtopic can simultaneously refer to this subtopic and to the main topic.

To highlight the subtopic from the common set of posts it is proposed to use clustering techniques using semantic metric connectivity between the posts and taking into account their structural links and features.

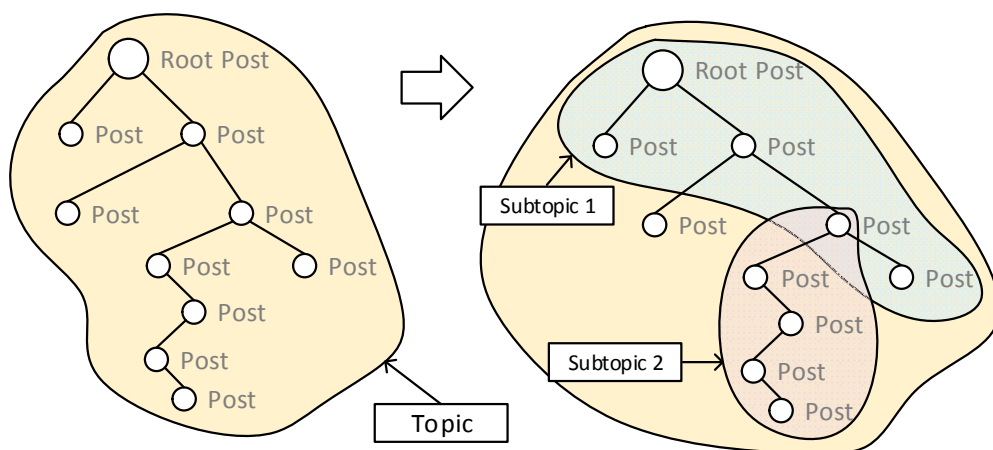Example of post aggregation into a subtopic is shown in Fig. 4.



**Fig. 4. Post aggregation example**

### 6.3. Decomposition of Topic into Several

When defining a topic in one of several subtopics the situation can occur when one of the subtopics is weakly associated in meaning with the main subtopic (the one which root post coincides with the root

of the whole topic). In this case, it makes sense to divide the discussion of different topics into several topics, each of which will contain a discussion only on one subject.

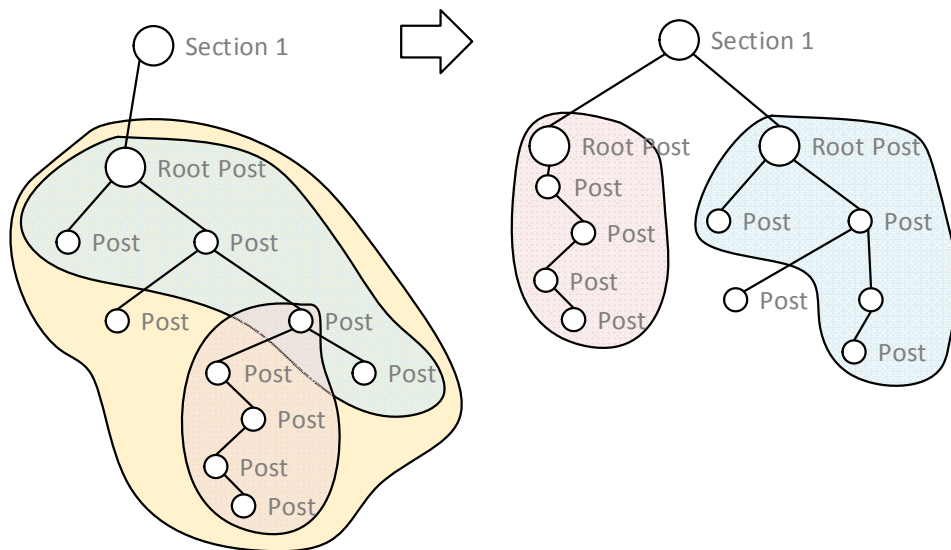Visually decomposition of topic is displayed in Fig. 5.



**Fig. 5. Topic decomposition**

### 6.4. Association of Various Subtopics into a New Topic

Semantically close subtopics of different topics of forum are sensible to present as a single new topic, to group all useful information on the same topic in one place. At that association of different subtopics may occur by a few rules:

● in chronological order – if the discussion started before, it will be higher in the list of comments;

● by relevance – subtopics more relevant to the root post will be added higher than the other;

● by semantic significance – the more in-depth discussion is conducted in a subtopic, the higher it appears in the list;

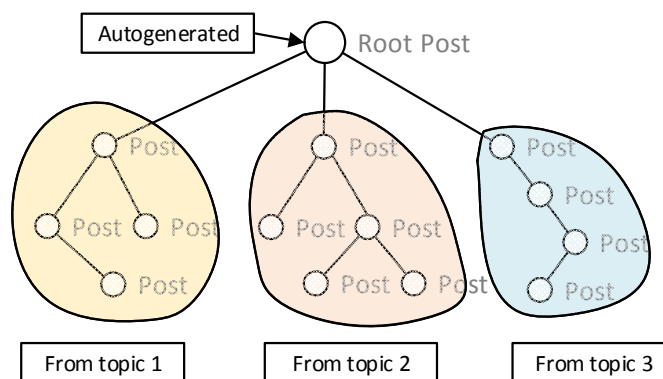● combined approach – in different forms combines all the rules described above.



**Fig. 6. Topic association**

Fig. 6 shows the association of subtopics into one, from left to right shows the order of inclusion of subtopics into a new topic (leftmost subtopic will be displayed in the user interface the first). The root of the new topic can be one of the root posts of subtopics or specially generated summary message.

### 6.5. Transfer of the Topic to Another Section

The semantic discrepancy may occur not only at the level of posts and branches of discussions in topics, but also more globally – at the level of topic accommodation in the section. In this case, it is logical to carry out an assessment of a topic proximity to the other in the current and in the other sections,

if it turns out that the topic is more relevant to the other sections, it is necessary to carry out its transfer, along with all its contents, in a more appropriate section. Transfer example is shown in Fig. 7.
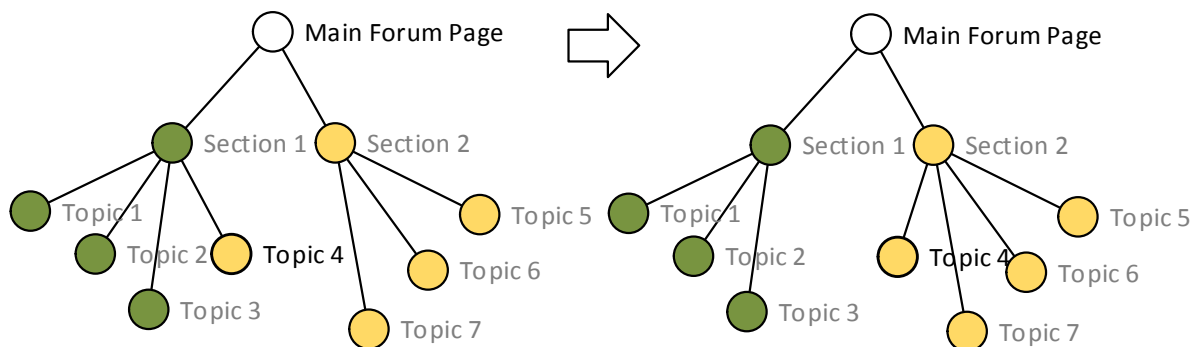
**Fig. 7. Topic transfer to another section**

### 6.6. Creation of a New Section

In extreme cases, it may occur that the initial breakdown into sections is not sufficient for the provision of forum information. Creation of a new section is necessary if there is a group of topics on one subject, while these topics are part of the section and the semantic distance between the group and the other topics of this section is sufficiently large. Then the creation of the section and the transfer of this group of topics to this section happens. Section creation example is shown in Fig. 8.
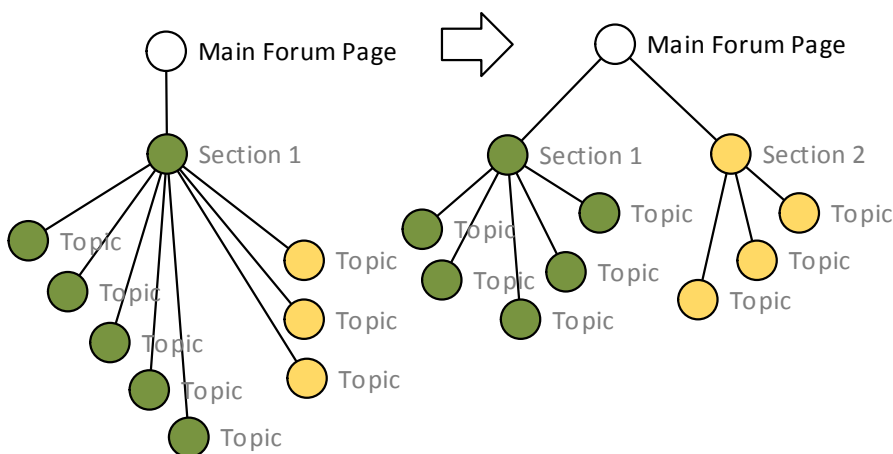
**Fig. 8. Section creation decomposition**

### Conclusions

Users of mass asynchronous online interaction means, such as forums, today face a huge amount of information that is generated daily, which leads to the effect of information overload and the inability to efficiently access information on the forum.

In this paper we propose methods and algorithms for processing information from online forums and intellectual change in the forum structure according to their content. We expect that the proposed approach to change the principle of data representation on the forum will provide the tools that will help increase the efficiency of access to information, reduce the level of information overload and lead to an increase in the efficiency of the web forums as instruments of cooperation and communication, as well as will form the basis for the creation of system of information decision-making using the information from the forums.

### References

1. Qiu M. and McDougall D. Influence of Group Configuration on Online Discourse Reading. *Computers & Education*, 2015, 87, pp. 151–165. DOI: 10.1016/j.compedu.2015.04.006

2. Loginovskiy O.V., Nesterov M.I., Shestakov A.L. [Corporate Information-Analytical System of a Major University as an Effective Tool for Improving Quality of Management]. *Higher Educational Establishment News*, 2013, no. 1, pp. 40–52. (in Russ.)

3. Koz'ko A.A., Construction of Adaptive Educational Forums Based on Intellectual Analysis of Structural and Semantics Features of Messages. *Supplementary Proceedings of the 4th International Conference on Analysis of Images, Social Networks and Texts (AIST'2015)*, Yekaterinburg, Russia, April 9–11, 2015, pp. 46–51.

4. Wang Y.C., Joshi M., Cohen W.W. and Rosé C.P., Recovering Implicit Thread Structure in Newsgroup Style Conversations. *ICWSM*, 2008, March.

5. Lampe C., Zube P., Lee J., Park C.H., Johnston E. Crowdsourcing Civility: A Natural Experiment Examining the Effects of Distributed Moderation in Online Forums. *Government Information Quarterly*, 2014, 31(2), pp. 317–326. DOI: 10.1016/j.giq.2013.11.005

6. Simarpreet G. Exploring the Benefits of Tagging Forum Posts Based on a Hierarchical Domain Model of the Course Content in Online Forums. PhD Diss., 2014.

7. Ren Z., Ma J., Wang S., Liu Y. Summarizing Web Forum Threads Based on a Latent Topic Propagation Process. *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, 2011, October, ACM, pp. 879–884. DOI: 10.1145/2063576.2063703

8. Almahy I., Salim N. Web Discussion Summarization: Study Review. *Proc. of the First International Conference on Advanced Data and Information Engineering (DaEng-2013)*. Springer Singapore, 2014, pp. 649–656.

9. Muthmann K., Löser A. Detecting Near-duplicate Relations in User Generated Forum Content. *Proc. of OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*. Springer Berlin Heidelberg, 2010, October, pp. 698–707. DOI: 10.1007/978-3-642-16961-8_96

10. Muthmann K., Barczyński W.M., Brauer F., Löser A. Near-Duplicate Detection for Web-Forums. *Proc. of the 2009 International Database Engineering & Applications Symposium*, 2009, September, ACM, pp. 142–151. DOI: 10.1145/1620432.1620447

11. Lee J.T., Yang M.C., Rim H.C. Discovering High-Quality Threaded Discussions in Online Forums. *Journal of Computer Science and Technology*, 2014, 29(3), pp. 519–531. DOI: 10.1007/s11390-014-1446-5

12. Bitton E. A Spatial Model for Collaborative Filtering of Comments in an Online Discussion Forum. *Proc. of the third ACM Conference on Recommender Systems*, 2009, October. ACM, pp. 393–396. DOI: 10.1145/1639714.1639797

13. Wu H.C., Luk R.W.P., Wong K.F., Kwok K.L. Interpreting TF-IDF Term Weights as Making Relevance Decisions. *ACM Transactions on Information Systems (TOIS)*, 2008, 26(3), p. 13. DOI: 10.1145/1361684.1361686

14. Reed J.W., Jiao Y., Potok T.E., Klump B.A., Elmore M.T., Hurson A.R. TF-ICF: A New Term Weighting Scheme for Clustering Dynamic Data Streams. *Proc. of the 5th International Conference on Machine Learning and Applications (ICMLA'06)*, 2006, December, pp. 258–263. DOI: 10.1109/icmla.2006.50

15. Landauer T.K., Foltz P.W., Laham D. An Introduction to Latent Semantic Analysis. *Discourse Processes,* Vancouver, 1998, 25(2–3), pp. 259–284. DOI: 10.1080/01638539809545028

16. Mihalcea R., Csomai A., Wikify!: Linking Documents to Encyclopedic Knowledge. *Proc. of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*, ACM, 2007, November, pp. 233–242. DOI: 10.1145/1321440.1321475

17. Witten I., Milne D. An Effective, Low-cost measure of Semantic Relatedness Obtained from Wikipedia Links. *Proc. of AAAI Workshop on Wikipedia and Artificial Intelligence: an Evolving Synergy*, AAAI Press, Chicago, USA, 2008, July, pp. 25–30.

18. Strube M., Ponzetto S.P. WikiRelate! Computing Semantic Relatedness Using Wikipedia. *AAAI*, 2006, July, vol. 6, pp. 1419–1424.

19. Gabrilovich E., Markovitch S. Computing Semantic Relatedness Using Wikipedia-Based Explicit Semantic Analysis. *IJcAI,* 2007, January, vol. 7, pp. 1606–1611.

# МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО ПЕРЕСТРОЕНИЯ ФОРУМА, ОСНОВАННЫЕ НА СЕМАНТИЧЕСКОМ АНАЛИЗЕ

## А.А. Козько

*Челябинский государственный университет, г. Челябинск*

Статья посвящена вопросу информационной перегрузки пользователей интернет-форумов. Рассмотрена сама технология веб-форума, определены недостатки существующих реализаций, которые приводят к проблеме информационной перегрузки. Рассмотрены существующие методы борьбы с этой проблемой. Предложен алгоритм структурно-семантического анализа сообщений на веб-форуме, позволяющий связывать сообщения в логические единицы – подтопики. Предложены, основанные на этом алгоритме, методы структурной адаптации веб-форума, позволяющие автоматически структурировать сообщения веб-форума в соответствии с их семантическим содержанием. Показаны перспективы использования такого подхода для борьбы с информационной перегрузкой.

*Ключевые слова: семантический анализ, онлайн-форумы, информационная перегрузка.*

### Литература

*1. Qiu, M. Influence of Group Configuration on Online Discourse Reading / M. Qiu, D. McDougall // Computers & Education. – 2015. – 87. – P. 151–165. DOI: 10.1016/j.compedu.2015.04.006*

*2. Логиновский, О.В. Корпоратвиная информационно-аналитическая система крупного вуза как эффективный инструмент повышения качества управления / О.В. Логиновский, М.И. Нестеров, А.Л. Шестаков // Известия вузов. Уральский регион. – 2013. – № 1. – С. 40–52.*

*3. Koz'ko, A.A., Construction of Adaptive Educational Forums Based on Intellectual Analysis of Structural and Semantics Features of Messages / A.A. Koz'ko // Supplementary Proceedings of the 4th International Conference on Analysis of Images, Social Networks and Texts (AIST'2015), Yekaterinburg, Russia, April 9–11, 2015. – P. 46–51.*

*4. Recovering Implicit Thread Structure in Newsgroup Style Conversations / Y.C. Wang, M. Joshi, W.W. Cohen, C.P. Rosé // ICWSM. – 2008, March.*

*5. Crowdsourcing Civility: A Natural Experiment Examining the Effects of Distributed Moderation in Online Forums / C. Lampe, P. Zube, J. Lee et al. // Government Information Quarterly. – 2014. – 31(2). – P. 317–326. DOI: 10.1016/j.giq.2013.11.005*

*6. Simarpreet, G. Exploring the Benefits of Tagging Forum Posts Based on a Hierarchical Domain Model of the Course Content in Online Forums. PhD Diss. / G. Simarpreet. – 2014.*

*7. Summarizing Web Forum Threads Based on a Latent Topic Propagation Process / Z. Ren, J. Ma, S. Wang, Y. Liu // Proceedings of the 20th ACM International Conference on Information and Knowledge Management. – ACM, 2011, October. – P. 879–884. DOI: 10.1145/2063576.2063703*

*8. Almahy, I. Web Discussion Summarization: Study Review / I. Almahy, N. Salim // Proc. of the First International Conference on Advanced Data and Information Engineering (DaEng-2013). – Springer Singapore, 2014. – P. 649–656.*

*9. Muthmann, K. Detecting Near-duplicate Relations in User Generated Forum Content / K. Muthmann, A. Löser // Proc. of OTM Confederated International Conferences "On the Move to Meaningful Internet Systems". – Springer Berlin Heidelberg, 2010, October. – P. 698–707. DOI: 10.1007/978-3-642-16961-8_96*

*10. Near-Duplicate Detection for Web-Forums / K. Muthmann, W.M. Barczyński, F. Brauer, A. Löser // Proc. of the 2009 International Database Engineering & Applications Symposium. – ACM, 2009, September. – P. 142–151. DOI: 10.1145/1620432.1620447*

*11. Lee, J.T. Discovering High-Quality Threaded Discussions in Online Forums / J.T. Lee, M.C. Yang, H.C. Rim // Journal of Computer Science and Technology. – 2014. – 29(3). – P. 519–531. DOI: 10.1007/s11390-014-1446-5*

12. Bitton, E. A Spatial Model for Collaborative Filtering of Comments in an Online Discussion Forum / E. Bitton // Proc. of the third ACM Conference on Recommender Systems. – ACM, 2009, October. – P. 393–396. DOI: 10.1145/1639714.1639797

13. Interpreting TF-IDF Term Weights as Making Relevance Decisions / H.C. Wu, R.W.P. Luk, K.F. Wong, K.L. Kwok // ACM Transactions on Information Systems (TOIS). – 2008. – 26(3). – P. 13. DOI: 10.1145/1361684.1361686

14. TF-ICF: A New Term Weighting Scheme for Clustering Dynamic Data Streams / J.W. Reed, Y. Jiao, T.E. Potok et al. // Proc. of the 5th International Conference on Machine Learning and Applications (ICMLA'06). – 2006, December. – P. 258–263. DOI: 10.1109/icmla.2006.50

15. Landauer, T.K. An Introduction to Latent Semantic Analysis / T.K. Landauer, P.W. Foltz, D. Laham // Discourse Processes. – Vancouver, 1998. – 25(2–3). – P. 259–284. DOI: 10.1080/01638539809545028

16. Mihalcea, R. Wikify!: Linking Documents to Encyclopedic Knowledge / R. Mihalcea, A. Csomai // Proc. of the Sixteenth ACM Conference on Conference on Information and Knowledge Management. – ACM, 2007, November. – P. 233–242. DOI: 10.1145/1321440.1321475

17. Witten, I. An Effective, Low-cost measure of Semantic Relatedness Obtained from Wikipedia Links / I. Witten, D. Milne // Proc. of AAAI Workshop on Wikipedia and Artificial Intelligence: an Evolving Synergy. – AAAI Press, Chicago, USA, 2008, July. – P. 25–30.

18. Strube, M. WikiRelate! Computing Semantic Relatedness Using Wikipedia / M. Strube, S.P. Ponzetto // AAAI. – 2006, July. – Vol. 6. – P. 1419–1424.

19. Gabrilovich, E. Computing Semantic Relatedness Using Wikipedia-Based Explicit Semantic Analysis / E. Gabrilovich, S. Markovitch // IJcAI. – 2007, January. – Vol. 7. – P. 1606–1611.

**Козько Александр Андреевич**, преподаватель кафедры информационных технологий и экономической информатики Института информационных технологий, Челябинский государственный университет; alkozko@mail.ru.