

THE ELECTRONIC DIGEST FORMATION AND CATEGORIZATION

Vasyl Andrunyk¹, Lyubomyr Chyrun²

¹Information Systems and Networks Department, ²Software Department, Lviv Polytechnic National University, S. Bandery Str., 12, Lviv, 79013, UKRAINE, E-mail: ¹phottt@ukr.net, ²chyrunlv@mail.ru

Information support is always needed when solving complex problems in any field of vital activity [1-5]. Meeting the information requirements is a mandatory requirement for innovation realization. At the same time, the complexity of information obtaining affects efficiency and quality of solutions. The Internet can be considered as large-scale mass media. The chaotic nature and the existence, the absence of the clear maintenance and update intervals of the most sites, as well as the problems concerning the effective information search, contribute to the Internet not having become the single reliable mass media so far. Only certain network elements (which are often called network or the Internet media) are considered to be full-fledged mass media. Network mass media are deemed to be news portals with a certain updating frequency, electronic versions of printed periodical publications and newspapers or magazines that exist in online electronic format.

One of the main features of the information production is the constant growth of its rate. Apart from increasing the amounts of information to the scales that render impossible its direct processing, there were a number of specific problems associated with the rapid information technology development. Therefore, on the one hand, there is quite a powerful array of information (the Internet resources) for decision making in various public living environments, society and the individual. On the other hand, there is a lack of information which is necessary for decision-making because of its dynamics, volumes, sources, and unstructured nature. The coverage and the generalization of the large dynamic information flows which are continuously generated in the media entail qualitatively new approaches.

While working in informational and analytical services or enterprises, one has to deal with a great diversity of the information sources, such as electronic newspapers and other Internet resources. This article considers Ukrainian electronic mass media – their disadvantages, benefits, and services. The conducted research studies of electronic mass media found that it is inexpedient to exploit human labor in the processes related to the digests formation and the digests ranking. The main part of this paper is also devoted to the methods elaboration of the formation and the rubricating of electronic digests. The practice of implementing systems in different organizations has proven effectiveness and simplicity of the system adapting due to the developed instrument of the automated digest formation and their rubricating. The universal data acquisition module allows you to automate completely the electronic information introduction from sources with bringing information to a common internal format, i.e. to minimize the routine work while entering the text data.

The digest is an annotated text based on the analysis of several documents. While compiling digests the automated abstracting methods of one document apply to the array with a large number of documents. The majority of the document automated abstracting algorithms consists of three basic stages: the source text analysis, the significant fragments definition (suggestions or whole paragraphs) and the conclusion formation. At the same time, a digest may also be considered as the source of annotated hyperlinks to the underlying documents. While forming out digests, the usage of quasi-abstracting methods makes the task of a coherent text attainment almost impossible. The combination of each document abstract will contain an excessive amount of incoherent information. However, while forming out the auto abstract, which consists of a certain number of source document announcements, and which is divided into subsections, the usage of the above mentioned method is quite acceptable.

Keywords – content, content analysis, content monitoring, content search.