One of the basic problems in supplying professional economists with information, in spite of available robust databases, advanced information systems, and information networks in the area of economics, is the problem of information search. Information search is understood as “a set of logical operations, the ultimate purpose of which is to trace by special signs all documents (annotations, summaries, full texts of articles, books, etc.) containing the required information…” [1, p. 248].

For indexing documents in information-search systems, special dictionary handbooks (thesauruses) are used as a means of control over the uniform usage of words and word combinations of the language of the object domain.

In forming a thesaurus, one should have an idea of what array of documents the information query will be made in; who will use the thesaurus and for what purposes; and the requirements for accuracy, completeness and speed of information search. The key words for the thesaurus are selected on the basis of the frequency with which they occur, the value of particular key words for the information search, and the adequacy of key words as scientific terms.

At present, with arrival of the Internet and full text databases, it may seem unnecessary to form complicated thesauruses. The user has at his disposal full text information systems with many millions of arrays of documents. However, as is marked by many analysts, for example in [2], the accuracy of vocabulary information search in the Internet makes up 10% on average and no single known query machines provides the completeness of search. The huge output in the Internet shows clearly that the main information problem in the 21st century will be the improvement of search accuracy. The removal of ambiguity of terms to make their meanings more exact is the way to optimize the information search. The main bulk of users formulate their queries too broadly, and for this reason do not obtain truly relevant information. The creation of a thesaurus as a part of the information-search system becomes increasingly actual.

In the current understanding (for full text databases), a thesaurus is no longer considered a means of translating texts from the natural language to the descriptive information-search language by the substitution of appropriate descriptors for key words.

The Russian State Trade Economic University is working now on the creation of a thesaurus of electronic commerce.

The thesaurus has been developed in accordance with the GOST 7.25-2001 standards, “Information Search Monolingual Thesaurus.”

In doing this, the experience in creating thesauruses of the Institute of Scientific Information in Social Sciences (INION) of the Russian Academy of Sciences...
was studied. The thesaurus for economics and demography was created stage-wise over 20 years and was intended to be a linguistic means for content processing and the search of documents being introduced into the bibliographic information system. Additionally, we have analyzed WordNet software packages for the creation of computer thesauruses of the Russian language within the framework of the RussNet project [6], experience of RUSSIA information system of the University of the Research Computation Center of the Lomonosov Moscow State University [7], software packages of automatic preparation of the thesaurus of the Medialingua close-end company (Thesaurus SDK) [8], Excalibur of the Vest-MetaTechnology company, thesaurus RCO of the Garant-Park-Internet company [9], etc. As a result, it was concluded that the formation of a thesaurus in the field of economics remains a difficult, theoretical, and practically little-known problem.

In the formation of general concepts of the subject field of electronic commerce, it was taken into account that the basic primary sources of information and the means of its popularization are informal personal contacts among scientists and professionals, unpublished scientific-technical documents (reports, dissertations), scientific-technical literature (newspapers, journals, books, proceedings of scientific conferences, descriptions of inventions, standards and other normative documents), electronic publications.

In the process of the study, we have identified the core zone, i.e., journals that present the subject field of electronic commerce more completely; formed bibliographic and thematic indexes of literature; selected profile periodicals; selected personalities and groups of authors working in the field of electronic commerce, as well as on special WEB sites; and carried out frequency analysis of texts on electronic commerce and semantic analysis of the subject field with the use of software packages freely available in the Internet. A set of such software packages are available at the site of the Russian virtual library (http://www.rvb.ru).

As a means of frequency analysis, we have used the URS software of 1.1 version (PCC source) of the NooLab company, Novosibirsk (http://www.rvb.ru).

According to GOST 7.25-2001 standards, the formation of the thesaurus on electronic commerce included the following stages: creation of general concepts of the subject field of electronic commerce, determination of the requirements of state standards for the formation of a thesaurus; search on the Internet of profile periodicals, books, etc.; identification of profile WEB sites; talks with professionals in this subject field in order to refine the obtained information; work with dictionaries on electronic commerce; selection of core sources of information; accumulation of information arrays; preparation of an information array for frequency and semantic analyses; frequency analysis of texts; determination of the parameters of a statistically authentic array; formation of ordered frequency lists of terms of the subject field according to the frequency of occurrence; semantic analysis of texts; determination of paradigmatic links of the terms of the subject field; and proper thesaurus formation. The permutation index in its classic form was not prepared.

The result of the analysis of the text “Payment Systems of the Internet” by the URS utility with the terms by sorted by frequency of occurrence has shown, for example, that the most frequently occurring words in the given excerpt are

- **PAYMENT** 182
- **SYSTEM** 121
- **BANK** 103
- **ACCOUNT** 66
- **CARD** 60
- **PURCHASE** 57
- **CLIENT** 51
- **ELECTRON** 48
- **AUTORIZ** 36

It is seen that the terms selected as key terms in the text “Payment Systems of the Internet” by the URS utility correspond well to the title mentioned in the heading. Indeed, in the text on payment systems of the Internet, the most frequently occurring words are “payment,” “system,” “bank,” etc. To build a thesaurus, the full list of key words is formed through an analysis of arrays of texts of a certain minimum volume.

With the help of the URS utility, different texts on electronic commerce selected on the sites of the Internet by electronic commerce and from a handbook were studied. As a result, a collection of terms (key words) on electronic commerce based on frequency of occurrence in different sets of texts on this topic was formed.

To build the thesaurus, a necessary condition is the presence of a full list of key words of the given subject field. However, key words can be not only separate words, but also word combinations that adhere to each other by meaning. To identify such word combinations, a semantic analysis of the text is required.

The semantic analysis of texts on electronic commerce was made with the use of the software TextAnalyst of the MicroSystems firm (available on the site of the Russian virtual library, http://www.rvb.ru).

As a result of the analysis of the subject field by this software, a semantic network consisting of concepts taken from the text and the relationships between them is formed. Each concept is characterized by semantic weight, i.e., by its significance in the text and statistical measure, that is, its frequency of occurrence in the text. The identified concepts are arranged by weight, alphabet, and frequency of occurrence.

For example, in the analysis of the term **card** of the subject field of payment systems of electronic commerce by TextAnalyst, the following results were obtained: