

УДК 330.46

Taras HAIDA

ECONOMIC DATA SOURCES FOR ENTERPRISE'S INFORMATION AND ANALYTICAL SUPPORT SYSTEMS

Summary. The article systematizes information sources for enterprises analysts, who receive the data for further processing and analysis with information system to improve management decisions. In this paper seven main sources are selected and analyzed. Internet sources are found to be the most promising and available for small and medium enterprises. The author has also conducted the research of the most affordable software (such as Teleport, Offline Explorer, WebCopier as tools for site-saving, Avalanche as an automatic internet information-mining tool) and also hardware for these purposes. In addition to this, the author revealed that Text-mining method is a highly effective tool for knowledge extraction with competitive nature of the data sets. With this tool it is possible to search and process information on the Internet automatically, allowing for analysts to significantly improve efficiency and reduce costs for this type of work. The features, advantages and disadvantages of purchasing of ready-made information products are also shown. Purchase of ready-to-use information products now greatly simplifies the procedure for obtaining new knowledge about the environment, but is usually expensive and can have a number of shortcomings like a lack of information quality. The author concludes that the choice of sources by management has to be done carefully, basing on the needs and capabilities of the company. As a further way of study in this direction the author can name modern software tool packages for gathering, systematization and storage of competitive information for enterprises decision making units.

Key words: information, sources, supply, enterprise, system, competitiveness, analytics.

Тарас ГАЙДА

ДЖЕРЕЛА ЕКОНОМІЧНИХ ДАНИХ ДЛЯ ІНФОРМАЦІЙНО-АНАЛІТИЧНИХ СИСТЕМ ПІДПРИЄМСТВА

Резюме. Систематизовано джерела інформації для аналітиків підприємства, які збирають дані для їх подальшого опрацювання та аналізу інформаційною системою з метою підвищення ефективності управлінських рішень. Вибрано й проаналізовано сім основних джерел. Інтернет-джерела визначено найперспективнішими й доступними для малих і середніх підприємств. Розглянуто ряд програмного забезпечення (наприклад Teleport, Offline Explorer, WebCopier як інструменти для завантаження та збереження веб-сайтів, Avalanche, як засіб автоматичного збору інтернет-даних), а також апаратного забезпечення для цих цілей. Крім цього, показано, що метод text-mining є ефективним інструментом для видобування знань із загальних відомостей конкурентного характеру. За допомогою цього інструменту можна шукати й опрацювати інформацію в Інтернеті автоматично, дозволяючи аналітикам значно підвищити ефективність і суттєво знизити витрати на цей вид робіт. Також розкрито особливості, переваги та недоліки купівлі готових до використання інформаційних продуктів, що значно спрощує процедуру отримання нових знань про середовище підприємства, але, як правило, є дорогою і може мати ряд недоліків, таких, як можливість отримання неякісної інформації. Зроблено висновок, що вибір джерел керівництвом повинен бути зроблений, виходячи із потреб та можливостей компанії. В якості подальших шляхів дослідження в даному напрямку автор назвав сучасні комплекси програмного забезпечення для збору, систематизації й зберігання інформації про конкурентне середовище, яка необхідна для відділів, що приймають управлінські рішення.

Ключові слова: інформація, джерела, забезпечення, підприємство, система, конкурентоспроможність, аналітика.

Formulation of the problem. Due to recent geopolitical events there is no doubt that Ukrainian companies will have to operate in the global market. This is not only a desire to expand its sales to the world market, but also with the entry of foreign companies into domestic market. Typically, their products have extremely high competitiveness. For successful competing, enterprises of our country, except, of course, improve product quality

according to the world-class standards, have to be provided with competitive information. Such information includes information about existing technologies, state and intentions of competitors, changes in the requirements of international organizations in the countries where it is intended to carry out foreign economic activity, market conditions, possible trends in changing consumer needs, the emergence of substitute products, etc. The described type of information is an objective reflection of the external environment of the company, and a thorough knowledge of the status and trends in this environment is critical for the adoption of certain management decisions, especially in foreign trade. To study the competitive environment practices of leading companies gained experience in creation of information and analytical support system (the IASS) – a set of tools, techniques, resources, technical and staff support for collection, processing, storage of described above type of information. However, operation of such system requires input data, the gathering of which is considerable obstacle to the establishment and effective operation of information units in enterprises, especially small and medium because of their limited resources for such information systems.

Analysis of recent research and publications. The subject of information and analytical support, competitive and business intelligence engaged by such scholars as P.H. Lun, D. Vesset, B. Evelson and others. Especially noteworthy are A. Matvienko, M. Tsyvin, A. Masalovych, A. Tyulin, D. Lunde and others, whose works focused considerable attention on primary sources of data, methods and techniques for obtaining them. Among Ukrainian scientists papers of G. Pocheptsov, O. Lyashenko, V. Galitsyn and others should be mentioned. However, despite considerable scientific achievements, there are still no comprehensive, systematic approaches to this problem in the context of enterprises, of practical possibility to collect primary data systematically, and therefore the problem is not completely covered.

Formulating of article goals. The main objective of this work is the identification and description of possible data sources and methods for their preparation, which are necessary for the operation of information-analytical support system of people responsible for decision-making in enterprises, taking into account national specifics, limited resources of local enterprises for such information systems as well as for the purchase of ready information products.

The main research material. Today there are many sources where you can receive all necessary information. It should be noted that before the mass emergence and spread of computer technology some problems with access to data sources still existed. This is one reason for the negative attitude of many scholars to the term "commercial intelligence". After a century ago to get really valuable information its own insider is needed. Therefore, the etymology of the term "commercial intelligence" is equal to the industrial "espionage." Nowadays, with the rapid development of information systems and technology, enormous spread of the Internet in almost all areas of human life information exchange became more intense. For objective reasons, information about enterprises seeps into various mass media. It should be noted that in this study only those data sources, access to which can be obtained without violation of the laws and legal norms are considered. Of course, much of this information is not yet suitable for the adoption of certain management decisions, but after processing these "raw" data it can be transformed into an extremely valuable knowledge.

First of all it is advisable to specify the basic requirements for data that can be obtained from various sources. We believe the most successful is the approach to classification requirements by A. Matvienko and M. Tsyvin shown in figure 1.

In figure 2 the author shows the general scheme of obtaining input data for the system, where the branches often act as information agents and sources through which these agents distribute it. Let us analyze each element.

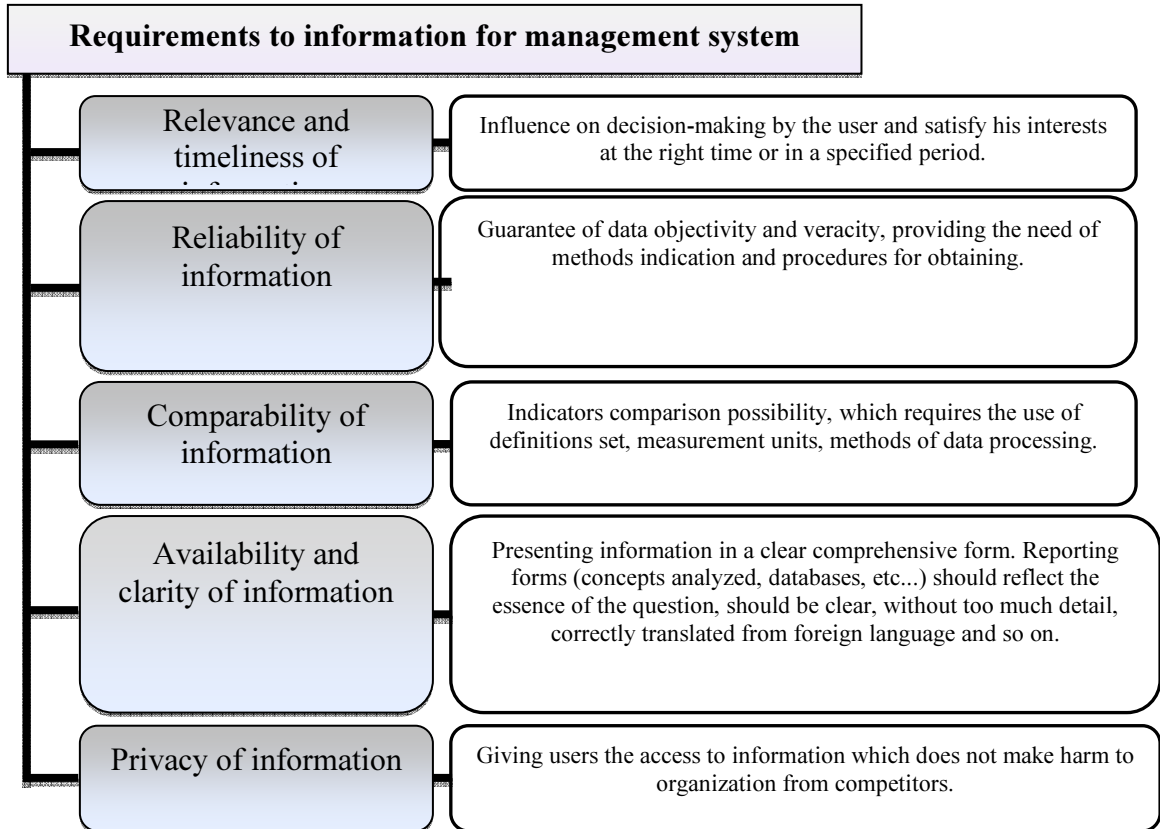


Figure 1. Requirements for management information [1, p.65]

Рисунок 1. Вимоги до управлінської інформації [1, с.65]

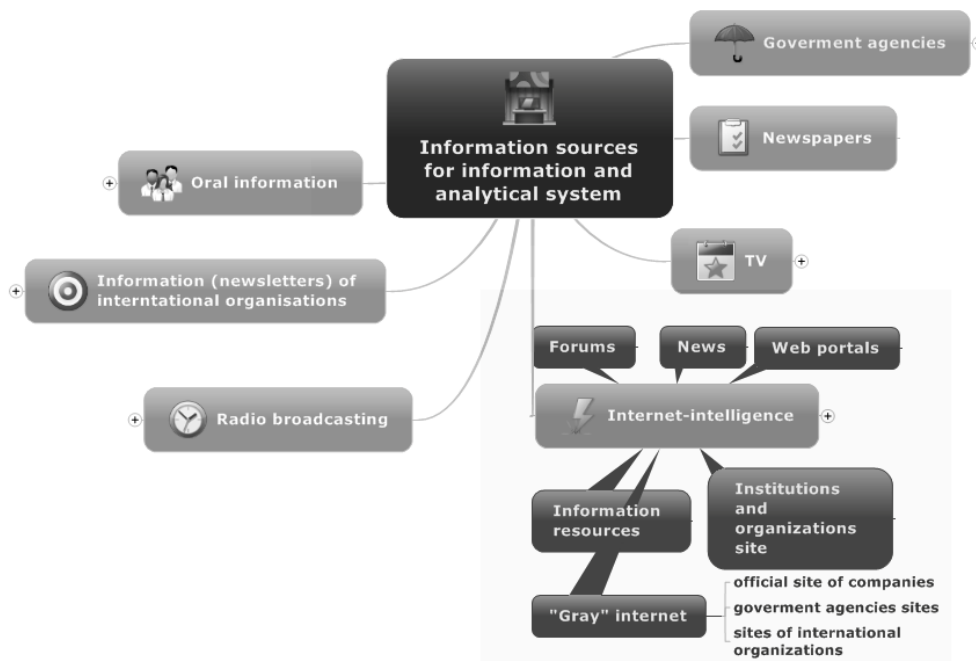


Figure 2. Data sources for informational and analytic support system

Рисунок 2. Джерела даних для системи інформаційно-аналітичного забезпечення

Thus, the authorities are affordable and effective source of data. This group is very large, this includes data from ministries, government agencies, committees, services and others. Information is generally accurate, it is freely available and can be supplied in almost all forms. Also, state agencies maintain the legal regulation of economic activity, including foreign trade, and therefore require monitoring of their information.

Newspapers are also a powerful source of data. These include a variety of professional journals, collections, digests, press releases and domestic newspapers and magazines with news and more. The latter are relatively accessible source of data, but often require confirmation, especially unscientific editions. News-like materials have the greatest lag time for obvious reasons.

Television is rather difficult the media for monitoring, because it requires the individual worker or the company that would either constantly review and analyze television programs on the researched subject or accumulate a considerable amount of archival material, which, in turn, require significant amount of working time. However, it should be noted that with the increasing popularity of the Internet TV materials can be sliced on passages and published on the websites of TV channels. Metadata are assigned to these passages and they are placed in the relevant sections. So the video or TV show can be quite easily found using standard search programs.

Broadcasting is similar to the previous data source. News, officials interviews etc. are often transmitted via radio. Such information is not always reliable and requires verification. Often radio is a backup source that replicates information from newspapers and television, sometimes with a certain time lag.

International organizations are extremely important source of data at the aspect of international business activity. This is due to the fact that international organizations governing relations between actors in the global market, establish certain rules for its operation, impose restrictions or, on the contrary, contribute to the revitalization of certain areas in certain countries (companies that represent them). That is why the company, which carries out foreign economic activity (especially sales), is extremely important to regularly monitor all changes in the regulation of international relations, including those between states.

Verbal information is unreliable source, difficult to assess because of its subjectivity and, sometimes, misleading. That is why the information should be most carefully checked. In practice, analysts do not so often have to deal with verbal information.

A separate group, the value of which can't be overestimated, is Internet. This is facilitated by several factors: Internet network covers the whole world today. It is practically impossible not to leave a mark of enterprise's activity, information available 24/7 from anywhere in the world as information data are almost always free.

Although in some cases there are drawbacks and limitations to obtain data from the Internet. For example, some highly specialized areas of information-analytical work may not have their imprint on the network, or not be indexed because they can't be found even using specialized search tools. Also, important scientific or professional journals, providing valuable information cannot be published on the net in general, published in part or with some time lag. Another fact is that the Internet, because of its accessibility, may have the effect of information noise, being a field for the misinformation. Thus competitor enterprises can protect themselves against the spread of information, the disclosure of which is critical to their business.

The data on the Internet can be placed on representative sites, forums, portals, on news sites, as conventional electronic documents (files) and others. Obtaining such data generally does not require highly skilled analyst in the company and is widely available. However, we should focus on some specific factors that contribute to the effective use of the Internet. We identified 3 groups of software products that significantly improve the efficiency

of the employees involved in the information-analytical work. These include the following programs:

- For integrated sites storage;
- Automatic collectors of text, numerical and other data by predetermined parameters (filters);
- Programs for Text-Mining technology, content analysis, transformation of unstructured data arrays into structured, with well-specified "shape".

The actuality of the first group can be explained with the fact, that web sites on the Internet is unstable for various reasons - closed by owner, blocking by providers, technical and hardware server problems, etc. The site may change; it can be modified over time. If the researcher is interested in its contents in the form as it is at any given time, he can save it in the local computer (server) at his own enterprise. To do this, there are several simple but effective programs such as Teleport, Offline Explorer, WebCopier and others.

The next group of programs is designed to relieve analyst from routine work in terms of manual content analysis of web sites for relevant information. These programs are called search robots. This software is designed to search the Internet for information with predetermined parameter, transition on several levels of hyperlinks and download it to your computer (server) at the enterprise. An example of such a program is Avalanche, which allows you to flexibly configure the search results catalog. Of course, this type of program has its flaws - the relevance of the results is not absolute, but the value of this tool in informational and analytical system of the company is evident.

As for the latter group of software, it should be noted that the "raw" unstructured data make up a large part of the information which users deal with. Finding something valuable in such data became realistic only by using specialized information technology tools. Unceasing development of Internet resources has deepened the problem of information overload.

This category should be considered from the point of obtaining data, on which information and analytical work, the system of indicators and indices that will show a status at a particular market where company operates or plans to operate is based.

The main elements of the Text Mining approach are [2, p.161]:

- Abstracting;
- Identification of phenomena;
- Classification;
- Clustering;
- Answers to questions;
- Content indexing and searching by keyword.

Also, in some cases the described set can be completed with support and a taxonomy and thesauri tools.

The Internet can be divided into two components: a stable and dynamic. The stable part of the network contains long terms; while dynamic include resources which are constantly updated. One part of the dynamic component eventually flows into the stable side, but most of it disappears from the network or segment falls in the "gray" of the Internet and cannot be available for traditional search methods. "Gray" Internet is that part of the network that is not indexed, but open and accessible, and access to it can only be a direct link. This network is also a part of the area of interest of business intelligence at the enterprise, because through it, you can access a number of documents, such as [3]:

- Pages not visible for search robots;
- Documents in various forms (doc, pps, ppt, etc.);
- Open sections;
- Open ftp-servers;

- Massive leaks classified as "for official use only";
- Documents of short-term storage;
- Information about the vulnerability of protected areas that allow you to bypass the protection without attacks and hacking.

Also, a good and practically free method of gathering all necessary data base for a given topic - using computer programs like parsers, spiders, crawlers, etc.

Returning to the general characteristics of data sources, it is worth noting that the same data can be obtained simultaneously from multiple sources. For example the State Statistics Committee publishes information on its website and issues it in the traditional paper version. Also printed publications after a certain period often publish materials online. And seen in Fig.2 "state authorities" actually can submit information via branch "publications" or "TV". All this must be considered, based on the specifics of the data sources. It is important to correlate the required efficiency and costs, as some of them may lose relevance very quickly, but then will be free (free) access. It should also be noted that the obvious advantages and disadvantages of each data source in general are almost impossible to describe. These sources are used by analysts in the enterprise based on specific needs and requests of information support system.

Moreover a large number of obtained data from different sources is ready to substantiate decisions. Main activities of analytical units of large companies are focused on finding this information. It usually characterizes the microenvironment of the company – the environment which the company can influence on or directly interact with. An example is the increase in prices for grain, which leads to an increase in costs for bread products - for the manufacturers supply of grain belongs to microenvironment. Although such information can characterize macro environment as well - the environment of the company it cannot influence and does not interact directly. For example, for manufacturer of bread higher prices for industrial ammonia causes higher prices for nitrogen fertilizers for wheat, which in turn, increases costs, and thus the cost of bakery products. That is quite a long logical chain of events, the first of which is quite remote from the main occupation of the investigated company, does or may cause a significant impact on its production activities.

However, it should be mentioned that there are authors who believe that current trends are talking about some changes in the process of information and information resources. Current experience shows that open source analysis can detect about 60–70% of the information, having regard to the general theoretical, technical issues, and creates original base in the solution of scientific, technical and industrial business problems [4]. This information gives an idea of what kind of problems and what general technical base may be resolved, but usually does not answer the question how one or another competitive problem will be resolved. This information provides the necessary "medium" level of knowledge about the problem, however, does not always reveal the most important and promising developments, methods and working techniques, opponent intents, features and technologies, etc. But despite this, the author notes that it is impossible to ignore public resources, in any case.

As an argument, which partially refutes this statement, we can give the practice of various analytical tools for data mining. They allow receiving from available raw data sources (including "noisy" information) well prepared and ready-to-use data and information.

Also, data, databases or even knowledge banks can be bought in ready-made form from specialized companies that provide information and analytical services or information products. Obviously, the acquisition of ready-to-use information has its advantages. First of all, the information obtained is ready for use, often the data is structured, and it can be quantitatively and qualitatively measured. Such information is gathered and processed by professionals. That's why the value of this information is high enough. In addition, it is

possible to order necessary research in a specific branch of enterprise's activity from these information companies.

However, there are several drawbacks. The first one is the high cost of such services. Also required information may not be available at the information company; an error factor is not excluded, and that's why this data sometimes need validation. And finally, the order can be improperly made, when the received information product will have a lack of quality.

Conclusions. After analyzing the main sources and methods of obtaining information, it can be concluded that at the present stage of information technology development as the possibility of obtaining raw data and ready for decision-making data is quite wide – from traditional analysis of publications, to text-mining technology. Open sources analysis is a powerful and widely available method of obtaining information, making it beneficial for domestic (Ukrainian) enterprises, especially small and medium. But at the same time it requires higher labor, time constraints may have some limitations because of the unavailability of certain information and other shortcomings. Purchase of ready-to-use information products now greatly simplifies the procedure for obtaining knowledge about the environment, but it is usually expensive and can have a number of shortcomings like a lack of quality of information. Therefore, the choice of sources by management has to be done carefully, basing on the needs and capabilities of the firm, according to its size or life cycle. As a further ways of study author can name new software for automated data collection and categorization, structuring and storing information in relational and non-relational databases as a way to maximize optimization, reducing the cost and increasing the efficiency of gathering competitive information for businesses.

Висновки. На сучасному етапі розвитку інформаційних технологій можливості отримання як первинних даних, так і готових для прийняття рішень відомостей є досить широкими – від традиційного аналізу друкованих видань, до інтелектуального аналізу за допомогою технології text-mining. Відкриті джерела є потужним та широкодоступним способом отримання інформації, що робить його вигідним для вітчизняних підприємств, особливо малих та середніх. Але водночас вимагає більш високих трудових, часових затрат, може мати ряд обмежень через недоступність певної інформації та інші недоліки. Придбання готових інформаційних продуктів підприємством значно спрощує процедуру отримання знань про зовнішнє середовище, проте, зазвичай, є дорогим та може мати ряд недоліків через «неякісність» відомостей. Саме тому вибір джерел керівництвом підприємства доцільно здійснювати зважено, виходячи із потреб та можливостей фірми, відповідно до його розміру чи життєвого циклу. Подальшого вивчення потребують новітні програмні засоби для автоматизованого збору та категоризації, структуризації й збереження інформації в реляційних та не реляційних базах даних як спосіб максимальної оптимізації, зниження вартості та підвищення ефективності збору конкурентної інформації для підприємств.

Literature

1. Матвієнко, О.В. Інформаційний менеджмент: опорний конспект лекцій у схемах і таблицях [Текст] / О.В. Матвієнко, М.Н. Цивін. – К., 2006. – 170 с.
2. Ландэ, Д.В. Поиск знаний в Internet. Профессиональная работа [Текст] / Д.В. Ландэ. – М.: Вильямс, 2005. – 272 с.
3. Масалович, А. Обеспечение прогнозирования, выявления, предупреждения и пресечения внешних угроз информационной безопасности в процессе ВТС методами компьютерной конкурентной разведки: доклад на секции НТС ФГУП "Рособоронэкспорт" 04 декабря 2009 г. [Электронный ресурс]. – Режим доступа: <http://iam.ru/world/rosoboronexport.htm>.
4. Тюлин, А. Открытые источники информации – достаточно ли их для конкурентной разведки [Электронный ресурс] / А. Тюлин. – 2012. – Режим доступа : <http://www.amulet-group.ru/page.htm?id=906>.
5. Ляшенко, О.М. Нові інформаційно-аналітичні технології дослідження систем управління підприємствами [Текст] / О.М. Ляшенко // Інноваційна економіка. – 2012. – №3(29). – С.20–24.

6. Davison, L. Measuring Competitive Intelligence Effectiveness: Insights from the Advertising Industry [Text] / L. Davison // Competitive Intelligence Review. – 2001 – Vol.12. – No.4. – P.25–38.
7. Herring, J.P. Measuring the Value of Competitive Intelligence: Accessing & Communication CI's Value to Your Organization [Text] / J.P. Herring. – Alexandria: SCIP, 1996.
8. Evelson, B. The Forrester Wave: Enterprise Business Intelligence Platforms [Text] / B. Evelson // Forrester Research. – 2010.
9. Morris, H. The Financial Impact of Business Analytics: Build vs. Buy [Text] / H. Morris // DM Review. – 2003. – №(13)1. – P. 40–41.