DIFFICULTIES IN THE COLLECTION OF DATA AND INFORMATION IN ROMANIA

Elisabeta Haba
Academy of Economic Studies Bucharest
Sabina Popescu
Academy of Economic Studies Bucharest
Victor Velter
Valahia University of Targoviste

Abstract

The purpose of this paper is to highlight the difficulties encountered by any person or company when searching for valuable information on the official sites of the ministries or state institutions, using the available search keys. At the same time, the paper states the necessity to design a new, comprehensive database, with another two keys (one primary and one secondary), in order to better assist the search effort. The data and information contained in this database should be accurate, supplied in a timely manner and against small or no fees. At the same time, the level of the information supplied as result of the search should strictly observe the present regulation regarding the privacy of data and information in the European Union.

Key-words: *CAEN, search, database, data privacy, data mining*

JEL Classification: O_{32}

Introduction

The issues stated in this paper surfaced while trying to identify, by accessing at first the public official sites, all the Romanian producers of a certain item. The result of the search, although rich in data, was less than satisfactory, due to the modality in which the databases are constructed and to the way in which they are designed to be accessed. The search uses as primary key the CAEN code. The correspondence of the Romanian coding system is NACE Revision 2, as to Regulation (EC) No 1893/2006 of the European Parliament and of the Council of 20th of December 2006 establishing the statistical classification of economic activities NACE Revision 2 and amending Council Regulation (EEC) No 3037/90 as well as certain EC Regulations on specific statistical domains. NACE coding system has the same structure as CAEN – four levels, the fourth level (the class) consisting of headings identified by a four-digit numerical code.

The first problem associated with using the CAEN code is that each four-digit code consists of not one, but several activities. As such, the result of the search might contain irrelevant data for the searcher, since it will always include activities which are out of the desired range.

The second problem is the accuracy of the code used in the search, since under the Romanian present laws, a company may register several CAEN codes, apart from the main one, and be inactive in regard to them. In other words, the main code might be 6201 (which stands for computer programming activities), while in fact the company's main activity is better described by code 4651 (wholesale of information and communication equipment).

Therefore, the main registration code, at the moment of creating the company, might be completely different from the code of the actual activity carried on by the same company.

Paper content

The present design of the database of the Romanian National Statistical Institute (www.insse.ro) allows the search by using exclusively the CAEN code, with the problems described above – the results of the search are too general and might not be accurate, in terms of the real activity of a company. On the other hand, the database of the Romanian Ministry of Finance (www.mfinante.ro) allows individuals to search for companies either by name and county, or by unique registration code. The external search using another key is not possible. In a real situation, when a person wants to find a set of companies with the CAEN code, the usage of this particular official site is useless. Even more, when interrogating the database of the same institution, using either of the data required in the search field, the activity code does not appear on the first page returned. Searching even deeper, by selecting a year for the financial situation, allows the visitor to see the main CAEN code as a description, not as a number. In addition, the code declared on the official site might be accompanied by several others, not visible to the user. As such, while gathering information from this database, some of the activities a company performs are overlooked.

In the effort to obtain information, an individual might, at first, be drawn away from the official sites and into common international search engines, such as Google, Bing, Yahoo, AltaVista, Excite, AOL or others. The main problem with these engines is the huge amount of information returned as a result, relevant for the search engine, but not necessarily for the searcher.

In common practice, the solution to the search problem generally comes from consulting specialized directories, such as www.yellowpages.eu or www.kompass.com. Their respective database uses codes (over 50.000 for the latter) to pinpoint a product or a service. In addition, other indicators are being used, such as "P" for producer, "D" for distributor, "I" for importer and so on. In this manner, by using detailed, fractioned codes as primary key, and then applying other parameters to the search, the relevant results are narrowed down from a larger pool of results. Other private directories allow search using pre-determined keys, listed either alphabetically or by other criteria – activity domain, location etc.

The main concern in relying exclusively on these directories, even when they are very well designed and contain a lot of data and/or information is that they might not contain it all.

On the other hand, the state institutes, agencies and ministries are fully comprehensive, not leaving any company aside. As such, their respective databases, with an improved design, could offer all the necessary and exact information in a timely manner, and at the same time observe the laws regarding privacy of data and information dissemination.

The difficulties of listing all codes is obvious, but in our opinion, the database table should contain, along with the main CAEN code, also the code of the next most frequent activity performed by the company. In this way, the database will store both the declared code (which shows the activity for which the company was first registered for) and the functional code, which gives an insight of the current activity and is relevant to the search.

For instance, at the present time, the production of homeopathic preparations is included in code 2120, along with the manufacture of: antiserum and other blood fractions, vaccines, chemical contraceptive products for external use and hormonal contraceptive medicaments, medical diagnostic preparations, including pregnancy tests, radioactive in-vivo diagnostic substances, biotech pharmaceuticals, medical impregnated wadding, gauze, bandages, dressings etc., botanical products (grinding, grading, milling) for pharmaceutical use. To pinpoint the homeopathic preparations out of the code 2120 and using the theory of combinations, another field of maximum four digits can be introduced to uniquely determine the exact range of activities carried out by a company. So the code for these products can be, after adding the supplementary code, for instance 2120-0004, which stand for manufacture of pharmaceutical preparations, homeopathic preparations.

Conclusions

The most reliable data and information are stored in official organizations. Therefore, any search should start with their respective databases, the design of which can be improved, in order to provide the searchers exact, accurate, timely data and information.

REFERENCES

Ramez Elmasri and Shamkant Navathe, *Fundamentals of Database Systems*, Fourth Edition, Pearson, 2004, 135-139, 314-320.

Raghu Ramakrishnan and Johannes Gehrke, *Database Management Systems,* Third Edition, McGraw-Hill, 2003, 29-70.

EUR-Lex, Access to European Union Law, Official Journal L 393, http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2006:393:0001:01:EN:HTML,last accessed on September 20th, 2011.