The Polish University Web Sites

Małgorzata Kaliczyńska

mkal@po.opole.pl
Technical University of Opole, Faculty of Electrical Engineering and Automatic Control,
Sosnkowskiego 31, PL-45-272 Opole (Poland)

Abstract
The poster presents the participation of selected Polish universities in the academic web link database project. The researches were conducted in several phases – crawling the web, web link mining and creation of link metrics. The selected results have been presented in a form of charts and graphs and refer to a group of sixty universities. In the nearest future the researches outcome of about 100 Polish academic institutions of the major importance will be presented.

Introduction
The development of the Internet, has been the result of activity of the academic environment. Also in Poland the scientific institutions played a decisive part in joining the global net which resulted in creation of the Scientific Academic Computer Net having a profound impact on shaping the Polish Internet until present days. Polish universities' Internet sites are the source of knowledge to both students and scientists while links between hosts and domains reflect the level of co-operation between the institutions. Creation of the database of Polish academic web links and placing it in the Net will enable an easy access to such data to the scientists from all over the world.

The Academic Web Link Database Project
The project[6] was run by the Statistical Cybermetrics Research Group at the University of Wolverhampton five years ago to respond to the need for research into web links: web link mining and the link metrics creation. Everybody can use all of the resources for non-commercial reasons. Currently, on the project web site there are available databases of the link structure of 23 US universities, a 125 universities from the UK, 38 from Australia, 8 from New Zealand, 64 from Spain, 46 from Taiwan and 76 from Mainland China. The project has been described by its authors in a number of publications (Thelwall, 2002/2003). Shortly, the database of the link structure of Polish universities will be added.

Software tools
The following software is available and used: *SocSciBot Tools* for crawling the Web and fetch the WWW sites and *Cyclist* for word index creation and phrase search. Also, professional software for performing network analysis called *Pajek*[7] is used. Pajek enables drawing a web network from files created by SocSciBot Tools and then computing link structure.

Some Polish university Web Sites
The researches began with a group of sixty Polish higher education institutions also called 'university', 'academy', 'polytechnic', 'technical university'. As only some Polish universities use `edu` domains for their web sites, it does create a great difficulty to identify academic domains. The Polish university web pages contain educational materials, information on scientific research, academic staff achievements, administrative data on university structure and its authorities. The language used is Polish, ca. 30% of pages are in English and even less in German or French. Other languages are in a prevailing minority. The above features make that Internet pages of Polish universities are different from pages of the equivalent universities in other parts of the world. Thus, the following question appears: is the structure of Polish web links so much different from the one described by the authors in their project (Payne & Thelwall, 2004). To find an answer, the methods of statistcal analysis collected in the book *Link Analysis: An Information Science Approach* (Thelwall, 2004) were used. Amongst numerous methods of analysis below are presented their results in a graphical form. More results and a discussion on their outcome for a bigger group of universities will be shown on a poster.

[6] `http://cybermetrics.wlv.ac.uk/database`

[7] `http://vlado.fmf.uni-lj.si/pub/networks/pajek`

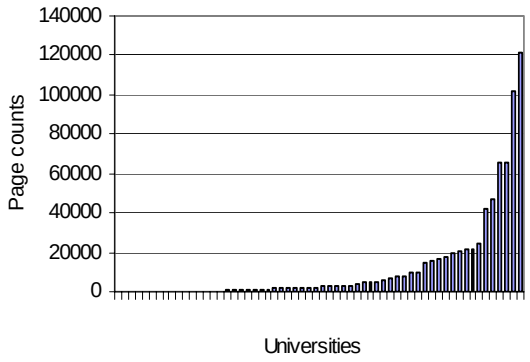

Figure 1: The range of university web sizes for the sixty Polish universities crawled

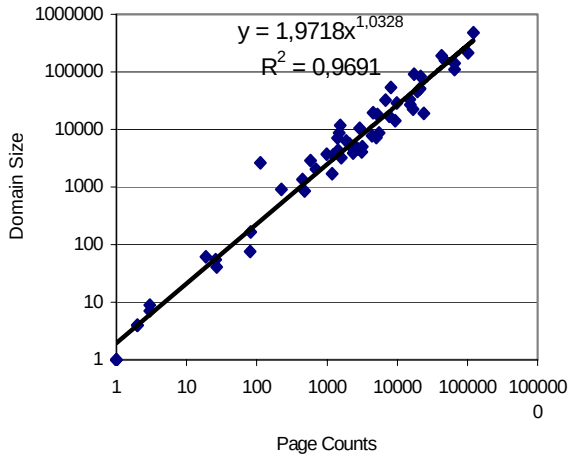

Figure 2: Logarithmic chart of Domain Size vs Page Counts

The researches conducted with option: *don't crawl any URLs with a question mark*. The next researches are intended to also include URLs containing a question mark.

References

Payne, N., Thelwall, M. (2004). A Statistical Analysis UK of Academic Web Links. *Cybermetrics*, 8(1) paper 2. Retrieved from: http://cindoc.csic.es/cybermetrics/articles/v8i1p2.htm.

Thelwall, M. (2002/2003). A Free Database of University Web Links: Data Collection Issues. *Cybermetrics*, 6/7(1) paper 2. Retrieved from: http://cindoc.csic.es/pruebas/v6i1p2.htm.

Thelwall, M. (2004). *Link Analysis: An Information Science Approach*. Elsevier