



# Információkeresés a weben és tanítása

Doktori értekezés tézisei

*Information Seeking on the Web and its Teaching*

Ph. D. Thesis

**Krauszné Princz Mária**

Debreceni Egyetem  
Természettudományi Kar  
Debrecen, 2006

## Tartalomjegyzék, Table of Contents

Információkeresés a weben és tanítása .....	1
Information Seeking on the Web and its Teaching .....	1
1. Bevezetés, a témaválasztás indoklása .....	3
2. A kutatás célja .....	3
2.1. Az információ-visszakeresés elméletének áttekintése.....	4
2.2. A hallgatók információ-visszakeresésének felmérése .....	5
2.3. Egy pedagógiai kísérlet a keresési készség javítására .....	8
3. Összegzés.....	8
1. Introduction, Verification of the Choice of the Topic.....	10
2. Research Goals.....	10
2.1. Review of the Theory of Information Retrieval .....	10
2.2 A Survey on How Students Seek Information on the Internet .....	12
2.3 A Pedagogical Experiment to Improvement Searching Skill .....	14
3. Summary .....	15
Irodalomjegyzék, References.....	16
Publikációs jegyzék, List of Publications .....	20

## 1. Bevezetés, a témaválasztás indoklása

Napjaink tudásközpontú társadalma feltételezi, igényli és ösztönzi tagjainak folyamatos tanulását. A globális gazdaság hosszú távú, dinamikus fejlődésének feltétele, hogy az emberek mind nagyobb többsége hozzájusson a szellemi tőkéhez és felhasználja azt. A már elsajátított szellemi tőke hamar leértékelődik, a szaktudás, az ismeretek elavulnak. Ezért az egyén sikeressége nem a konkrét ismeretektől, a tudás mennyiségétől, hanem a gyarapítására való képességtől, vagyis mindennél inkább a tanulástól függ.

Az Európai Bizottság 2000 októberében *Egész életen át tartó tanulás* (Lifelong learning) [EC, 2000] címmel kiadott memoranduma egy új tanulási kultúra kialakítását, a kompetencia alapú oktatás elterjesztését helyezi a középpontba, és a tanulás iskolán kívüli formáit is felöleli.

Az előtérbe helyezett új alapkészségek: az információs technológiában, az idegen nyelvekben, a technikai kultúrában, a vállalkozásban és a szociális készségekben (úgy mint magabiztosság, önállóság és kockázatvállalás) való jártasság. További alapvető és kiemelt fontosságú készség a tanulás képessége, ennek elsajátítása pedig a változáshoz való alkalmazkodást, és az egyre kiterjedtebbé váló információáramlásban való eligazodás biztosítását jelenti.

Új módszerek, technikák segítik az egyént a tanulásban, s a legjelentősebb ezek közül a számítógép felhasználása, illetve az Internet s a webes felületek biztosította lehetőség. Az Internet használata egyre inkább tért hódít a mindennapi életben, így az oktatás területén is. Az e-learning rendszerek megtervezett, összeválogatott tudástartalmat közvetítenek a hallgatók felé formális, illetve nem formális oktatási keretek között. Természetesen megjelenhetnek a tananyaggal összefüggő kérdések, amelyekre a válaszokat nem tartalmazza a rendszer tudásbázisa, de a mindennapi életben is felbukkanhatnak olyan kérdések, amelyekre a választ önszántából, érdeklődési körének, előképzettségének megfelelően keresi az egyén. A válaszkeresésben a hozzáértő személy (tanár, könyvtáros, szakértő) mellett egyre gyakrabban fordulunk a weben át elérhető információk felé. Ezen óriási mennyiségű információban való eligazodás alapvető követelmény az információs társadalom tagjai számára, s szükséges feltétele annak, hogy a web alapú tanulás sikeres lehessen.

## 2. A kutatás célja

A web hatalmas méretű információs tengernek tekinthető, amelyen való eligazodás korunk elvárása. A weben lévő információ keresésekor a leggyakrabban használt eszköz - a szabadszavas kereső - az információ-visszakereső rendszerek közül a legismertebb.

Kutatási tevékenységünk a weben található információk keresésével kapcsolatban három részből áll:

## 2.1. Az információ-visszakeresés elméletének áttekintése

Az információ-visszakeresés nagyon széles és tág tárgyterület, amely magában foglalja az információ tárolását, megjelenítését és visszakeresését is. Számos kutatási terület kapcsolódik hozzá: például algoritmusok, metaadatok használata, multimédia visszakeresés, mesterséges intelligencia, természetes nyelvi feldolgozás.

Az információ-visszakeresés elméletéhez tartozó kutatási területek eredményeinek nemzetközi irodalma gazdag [Blackwell, Amazon], de a téma összefoglaló, rendszerező áttekintésére – a téma nagyságára való tekintettel – angol nyelven is csak néhány könyv áll rendelkezésre [Baeza-Yates&Ribeiro-Neto,1999], magyarul pedig pedig nem érhető el összefoglaló mű [OSZK]. Értekezésünk első részében arra vállalkozunk, hogy e széles tárgyterületet felölelő, gazdag témát saját szempontjainknak megfelelően, a webről történő információ visszakeresésre fókuszálva áttekintsük, és magyar nyelven is hozzáférhetővé tesszük.

A webes keresések legfőbb eszköze, a szabadszavas keresők kifejlesztése, üzemeltetése napjainkban egyre inkább üzleti jellegű vállalkozás, amely szabadalmazott technológiát is jelent egy erős versenypiacon. Ebből következik, hogy pontos működésükről annyit tudunk, amennyit a fejlesztők biztonságosnak tartanak közzétenni, s ez általában a technikai részletekről kevés publikációt jelent. A tájékozódás legfőbb eszközei a weben át elérhető cikkek, konferencia előadások anyaga, a szakmai webportálokon át elérhető információ, illetve a próbálkozás és a következtetés.

- Dolgozatunk első részében az információ-visszakeresés alapfogalmait, az információ-visszakereső rendszerek működését, az információ-visszakereső modelleket, a dokumentumok jellemzőit ismertetjük. Bemutatjuk a dokumentumok formai és tartalmi leírására szolgáló leíró nyelveket. Áttekintjük a lekérdezések fajtáit, valamint bemutatunk néhány, a lekérdezésnél használt technikát. Példákon keresztül bemutatjuk a szöveges dokumentumok indexelésére és a bennük való keresésre kifejlesztett módszereket. Ismertetjük a multimédia objektumokban való keresés lehetőségeit.
- *Az információ-visszakeresés a webről* fejezet elején a webről történő információ-visszakeresés problémáit, a web megoszlását tekintjük át. Ismeretes, hogy a szabadszavas keresők a webnek csak a kisebb részét térképezik fel, ezért szólni kell a láthatatlan web létezéséről, valamint e területen való információkeresés lehetőségeiről. Ismertetjük a webes keresések három alaptípusát: a kulcsszó, a téma és a hiperhivatkozások alapján végrehajtható kereséseket és eszközeiket. Áttekintjük a szabadszavas keresők, mint a webes keresések leggyakrabban használt eszközének felépítését, az általános és egyedi tulajdonságokat a begyűjtő, indexelő, lekérdező és rangsoroló részben. Ismertetjük az információs igény minél pontosabb megfogalmazását lehetővé tevő összetett és részletes kereséseket. Bemutatjuk a rangsorolások jobbítására létrehozott megoldásokat. Példát adunk a hiperhivatkozások alapján történő keresésekre, ismertetjük a webbányászat taxonómiáját.

- *A weboldalak optimalizálása* című fejezetben megvizsgáljuk, hogy a lekérdezéseknél mi befolyásolja az eredményül kapott dokumentumoknak a rangsorban elfoglalt helyét. Ha tudjuk, hogy a szabadszavas keresők mely tulajdonságok alapján szerepeltetik a keresés eredménylistáján előkelőbb helyen az egyes oldalakat, akkor a webre szánt oldalak optimalizálásával javíthatjuk az eredménylistán weboldalunk helyezését, s a jobb eredménnyel valószínűleg több emberhez jutnak el közlésre szánt információink. Ekképpen a szabadszavas keresők tulajdonságainak ismerete hasznosítható a webre szánt dokumentumok elkészítésénél is.

## 2.2. A hallgatók információ-visszakeresésének felmérése

A weben át elérhető információk közötti eligazodás az információs társadalom tagjai számára egyre inkább elvárás, s különösen érvényes ez a megállapítás a ma felsőfokú tanulmányokat folytató hallgatókra, a jövő értelmiségére.

A felhasználók keresési szokásait, ismereteit számos tanulmány elemzi. [iProspect,2004, Greenspan,2002, Nielsen,2001, Environics Research Group, 2001]. Különösen érdekes volt számunkra egy kanadai felmérés [Gunn&Hepburn,2003], amely végzős középiskolások Internetes keresési ismereteit értékeli.

A Debreceni Egyetem Műszaki Főiskolai Karán az első éves hallgatók körében vizsgáltuk, mennyire képesek tájékozódni a weben. A gyakorlati órák tapasztalatai alapján megfigyelhettük, hogy:

- hallgatóink szinte kizárólagosan néhány keresési kulcsszó megadásával keresnek;
- a többség nem ismeri a szabadszavas keresők által biztosított lehetőségeket a szükséges információ pontosítására, valamint nem tudják az eredményhalmazt szűkíteni, szűrni;
- nem tudatosult bennük, hogy bizonyos információs igények megválaszolása más-más típusú keresőeszköz használatát igényli.

A hiányosságokat látva kíváncsiak lettünk arra is, hogy mennyire általánosítható ez a megfigyelés, azaz vajon hogyan keresnek az egyetem többi karán tanuló hallgatók? Ennek felmérése végett kérdőívet készítettünk, amelynek kitöltésére a Debreceni Egyetem különböző karain tanuló nappali tagozatos hallgatók közül, a karokon tanuló hallgatók arányának megfelelően 200 főt kértünk fel.

A keresési készség fogalma alatt a kereséskor megfigyelhető, alkalmazott ismeretek összességét értjük. A hallgatók webes keresési készségével kapcsolatban a következő kérdésekre kerestük a választ:

1. A hallgatók mely Internet szolgáltatásokat és milyen sűrűn vesznek igénybe?
2. Mely információs forrásokat részesítenek előnyben?
3. Hogyan vélekednek az Interneten található információk közötti keresési készségükről?

4. Milyen Internet hozzáférési lehetőségek állnak a hallgatók rendelkezésére?
5. Honnan szerzik ismereteiket az Interneten található, a tanulmányaikhoz felhasználható információk kereséséhez?
6. Milyen ismeretekkel rendelkeznek a különböző Internetes kereső szolgáltatókról?
7. Milyen stratégiákat és technikákat használnak a weben át elérhető információ-visszakeresésnél?
8. Mi befolyásolja a hallgatók keresési készségét?
9. Milyen felhasználói szokások jellemzik a hallgatókat a kereséseknél?
10. Hogyan vélekednek a szabadszavas keresők hatékonyságáról?

Elvégeztük a kérdésekre adott válaszok feldolgozását, a kapott eredmények elemzését, az eredmények összehasonlítását hasonló témájú felmérések eredményeivel.

#### **Legfontosabb eredményeink:**

- A DE hallgatóinál a web (99%) és az e-mail (86%) a leggyakrabban használt Internet szolgáltatás. Ezen szolgáltatások használatának aránya megegyezik a nemzetközi felmérések adataival [Greenspan,2002, Environics Research Group, 2001] Hallgatóink személyes célból gyakrabban keresnek a weben információt, mint tanulmányaikhoz.
- A hallgatók többsége információforrásként egyformán preferálja a webet és a könyveket, de összességében mégis többen választják a tankönyveken kívül is a nyomtatott forrásokat. Ez a tény a felsőoktatásban kissé meglepő, hiszen a tankönyveknek tartalmazniuk kell a törzsanyagot, míg a tananyaghoz kapcsolódó kutatások legfrissebb eredményei leginkább a webről érhetőek el: A nyomtatott források létrejöttéhez sokkal hosszabb idő szükséges, mint a weben publikálni egy anyagot, s a beszerzési források szűkössége miatt számos nyomtatott forrás nem is jut el az intézmények könyvtáráig.
- Kevés diák használja a szabadszavas keresők nyújtotta lehetőségeket az információs igény pontos megfogalmazására, az eredmények szűkítésére, szűrésére, ennek ellenére keresési képességét a legtöbbször átlagosnak vagy jónak tartja. A pozitív önértékelés a kanadai diákoknál még erősebb.
- A hallgatók döntő többsége elégedett az Internet elérési lehetőségeivel az oktatási intézményén belül. A hallgatók felének otthonában is biztosított a Világháló elérése.
- A diákok eléggé magukra (60%), illetve társaikra (44%) hagyatkozva tanulják a webről az információ-visszakeresést, a tanároknak ebben a tevékenységben kisebb szerep jut (21%).
- A hallgatók körében a legismertebb keresők között szerepel az USA népszerűségi listájának [SearchEngineWatch] jelenlegi első 3 helyezettje (Google, Yahoo, MSN). Az amerikai, kanadai középiskolások között népszerű Ask Jeeves a debreceni egyetemisták között alig ismert. A legismertebb magyar nyelvű kereső a Vizsla24.

- Felmérésünk szerint a hallgatók átlagosan 2,35 keresőt használnak alkalmanként és 1,13 keresőt rendszeresen, ami 3,48 kereső használatát jelenti valamilyen rendszerességgel. Ez a szám magasabb, mint az amerikai, kanadai felmérésekben, ahol átlag 1-2 keresőt használnak a felhasználók. [Greenspan,2002, Gunn&Hepburn,2003]
- Az információs igény megfogalmazásakor a DE hallgatóinak 76%-a nem használja a logikai operátorokat, 54% a kifejezések keresését, 66% a helyettesítő karaktereket, 64%-uk a részletes keresést. Hasonlóan rossz arányok jöttek ki a kanadai diákok eredményeinél.
- Az eredmény szűkítésére vonatkozó technikákat (pl. szűrés fájlformátum, dátum, előfordulások, tartomány, nyelv szerint) a diákok átlag 90%-a nem ismeri.
- A kérdőív feldolgozása során szeretnénk volna választ kapni arra a kérdésre, mitől függ a hallgatók keresési készsége, pontosabban befolyásolja-e azt a hallgató neme, az informatika iránti érdeklődése, az Internethez való hozzáférése, illetve van-e különbség a különböző karok hallgatóinak keresési készségében? Statisztikailag igazoltuk, hogy a hallgatók keresési készsége függ az informatika iránti érdeklődésüktől, valamint az Informatikai Kar hallgatóinak keresési készsége szignifikánsan különbözik némely kar hallgatóinak keresési készségétől. Statisztikailag igazoltuk azt az állítást is, hogy a felhasználó neme nem determinálja a felhasználó keresési készségét. Bár kimutatható, hogy az Internethez otthon is hozzáférő hallgatókra a keresések tanulásánál nagyobb önállóság és nagyobb magabiztosság jellemző, de statisztikailag nem tudtunk szignifikáns összefüggést kimutatni a keresési készség és az Internethez való hozzáférés között.
- A hallgatók felhasználói szokásaira jellemző, hogy az első keresés sikertelensége esetén a döntő többség (85%) a kereső kérdést finomítja tovább, és csak 12% próbálkozik másik keresővel. Hallgatóink az eredménylisták nagyobb százalékát tekintik át, mint az átlag felhasználók.
- Különböző becslések találhatóak a weben a láthatatlan web nagyságáról. [Bergman, 2001, BrightPlanet], amelyek szerint mérete a látható webnek sokszorosa. A láthatatlan web kereshetővé tételére különböző technikai megoldások születtek [Princz, Rutkovszky, 2004]. A láthatatlan web nagyon gyors mértékben növekszik, így tartalmának lekérdezhetővé tétele sürgető feladat. A felhasználók a szabadszavas keresők alkalmazásával szinte minden témáról kimerítő mennyiségű tudásanyagot találnak a weben, éppen ezért hajlamosak azt gondolni, hogy e szoftverek segítségével minden információt fellelnek, ami a weben át elérhető. Ezt támasztja alá felmérésünk [Princz, 2006], amely szerint a legjobb keresők által indexelt web méretére vonatkozó karonkénti becslések átlaga 48% és 72% között mozog, ami 64%-os egyetemi átlagot eredményez. Ezek nagyon magas számok, s mindenképpen a szabadszavas keresők teljesítményének túlértékelését tükrözik

### 2.3. Egy pedagógiai kísérlet a keresési készség javítására

Felmérésünk egyik eredménye szerint a kereséseket a hallgatók legtöbbször magukra, illetve társaikra hagyatkozva tanulják, s csak a megkérdezett hallgatók ötöde nyilatkozta, hogy keresési készségének kialakításánál, azaz a keresések tanulásánál tanárának is szerepe volt.

Hipotézisünk szerint tanári irányítás mellett, alkalmas keresési feladatok elvégzése, az alkalmazott stratégiák és technikák megbeszélése által a hallgatók keresési készsége jelentősen növelhető.

A kísérlet független változói:

- az általunk kidolgozott, a keresési stratégiák és technikák megismertetésére alkalmas feladatsor;
- a kísérlet céljának és a didaktikai követelményeknek megfelelő tanítási stratégia.

A kísérlet függő változója a különböző információs igények megválaszolásában nyújtott teljesítmény.

Kísérletünk egycsoportos kísérlet [Falus et al.,2000], ahol a kiindulási állapotot a tanítást megelőzően, a különböző keresési stratégiák és technikák alkalmazását igénylő, információs igények megválaszolásában nyújtott teljesítmény alapján mérjük.

A keresési készség tanítására – mint ahogy az alapvető készségek és az elemi ismeretek elsajátítására – a direkt oktatás módszere a legmegfelelőbb [Falus et al.,2003]. A direkt oktatás alapelve az, hogy a tanulók akkor sajátítják el leginkább az alapvető ismereteket és készségeket, ha a kitűzött célokat elemeire bontjuk, s határozott tanári irányítás mellett végigvezetjük a tanulókat az elsajátítás menetén. E módszernél a tanulók egyéni gyakorlásainak van jelentős szerepe. A kísérlet zárásakor az eredményt a modulban tanult ismereteket számonkérő zárthelyi dolgozatban mutatott teljesítmény adja.

A kísérlet eredménye alátámasztja, hogy a webről történő információ-visszakeresés tanítása után jelentősen javult az információs igény megválaszolását lehetővé tevő kereső eszközök és technikák ismerete és használata.

### 3. Összegzés

A globalizált világ elvárása, hogy a tanulók ne konkrét ismereteket, hanem tudásuk gyarapításának képességét tanulják meg. Az információ elérése kiemelt fontosságú szerepet kap. Az Internet az emberi tudás határtalan tárházaként tekinthető, amelyben az információ keresése alapvető készsége kell legyen korunk minden polgárának.

Általános szabály, hogy egy gyakorlati feladat sikeres megvalósítását elősegíti a feladattal kapcsolatos elméleti háttér ismerete, így az információ-visszakeresés elméletének, a webes keresési eszközök tulajdonságainak ismerete is javíthatja keresési készségünket.

A hallgatók jelentős része nem ismeri a lehetőségeket egy-egy információ keresésekor, ezért növelni kell a diákok tudatosságát a keresési stratégiák, s a stratégiának megfelelő keresőeszközök kiválasztásánál és használatánál. A tapasztalat azt



mutatja, hogy a többség szabadszavas keresőt használ akkor is, amikor tematikus kereső használata lenne indokolt.

Felmérésünk szerint a hallgatók többsége nem ismeri az összetett és részletes keresés nyújtotta lehetőségeket. Az információ-visszakeresés hatékonysága érdekében meg kell ismertetni velük a visszakeresésekor alkalmazható különböző technikákat, amellyel az információs igény pontosítható, s ezáltal az eredményhalmaz szűkíthető, a keresésre szánt idő csökkenthető.

A szabadszavas keresők tulajdonságainak ismerete nemcsak az információs igény optimális megfogalmazásánál segít, de jól hasznosítható a webre szánt dokumentumok elkészítésénél is. A weblapok elkészítésének tanításakor hasznos azt is ismertetni, hogy az egyes szabadszavas keresők a dokumentumok mely tulajdonságait veszik figyelembe az eredményoldalak rangsorolásánál. Weboldalaink optimalizálásával javíthatjuk weboldalunk rangsorbeli helyezését.

A keresési készség kialakításánál növelni kell a tanárok szerepét. Kísérletünk igazolta, hogy már egy-két, a didaktikai elvárásoknak megfelelő tanítási óra után is jelentősen növelhető a hallgatók keresési készsége.

## **1. Introduction, Verification of the Choice of the Topic**

Today's knowledge centered society supposes needs and encourages its members' continuous learning. One condition of the long term and dynamic development of the global economy is that a greater and greater majority of people should have access to intellectual capital and use it. The already acquired intellectual capital devolves fast, expertise and knowledge become outdated. Thus individual success does not depend on actual specific knowledge or the amount of information but on the ability to acquire more, that is, on learning above all.

A Memorandum on Lifelong Learning [EC, 2000] issued by the Commission of European Communities in October 2000 puts the emphasis on a new learning culture and also includes forms of learning outside school. The emphasized basic skills are the following: competence in information technology, foreign languages, technical culture, enterprises and social skills (like self confidence, independence and risk taking).

Another basic and very important skill is the ability to learn and acquiring this ability also means adapting to changes and finding one's way in the ever more extended information flow. New methods and techniques help people to learn and one of the most important of these is the use of the computer and the internet and the possibilities provided by web interfaces. The Internet is becoming more and more widespread in everyday life and thus in education as well. E-learning systems supply carefully chosen and designed knowledge in formal and less formal ways. Of course there are question related to the topic under discussion to which the answers are not contained in the material provided, but there are questions coming up in everyday life the answers to which are searched for according to individual interest and previous knowledge. When doing so, apart from consulting a competent person (teacher, librarian, expert), information available on the web is more and more frequently used. The ability to deal with this huge amount of information is a vital need for the members of our information society and a necessary condition of successful informal web based learning.

## **2. Research Goals**

The web can be considered a huge mass of information, navigating on which is what our time expects from us. When seeking information on the web, the most widely used tool, free word search engines, are the best known of all information seeking systems.

The research contains three parts regarding information seeking on the web:

### **2.1. Review of the Theory of Information Retrieval**

Information retrieval is a very wide area which includes information storage, display and retrieval as well. It incorporates many research fields (e.g. algorithms, use of metadata, multimedia retrieval, artificial intelligence, and natural language processing, etc.).

The literature of information retrieval related research fields is rich [Blackwell, Amazon], but due to the large extension of the field, there are only a few books written in English [Baeza-Yates&Ribeiro-Neto,1999] to give a systematic overall view of the topic, and there are no books available in Hungarian. [OSZK]. In the first part of our dissertation we undertake to give an overall review of this wide and rich field, from our point of view focusing on information retrieval from the web, and make it available in Hungarian as well.

The development and maintenance of the main tool used for searching the web, that of free word search engines, is more a business and also involves patented technology on a strong market. This implies that we know about how they work only what the developers consider safe to publish and this usually means few technical details. The main tools for learning are articles available on the web, conference material, information available on special web ports as well as trials and inferences.

- In the first part of our thesis discuss the basic concepts of information retrieval, how search engines work, models of information retrieval, and the characteristics of documents. We consider the languages used to describe the form and content of documents. We review query types and we also discuss some techniques used in queries. Through examples we explain methods developed for indexing documents and searching them. We also show the possibilities of searching in multimedia objects.
- At the beginning of the chapter *Information Retrieval from the Web* we overview the problems occurring while searching the web for information and the distribution of the web. It is widely known that free word search engines map only a smaller part of the web, so we need to mention the existence of then invisible web and the possibilities to search it. We discuss three basic types of web search: search based on key-word, topic hyper-reference and the tools used. We overview the structure of free word search engines, being the most widely used tools ,and the general and specific characteristics of the gathering, indexing, querying and ranking parts. We discuss multiple and detailed searches that make information needs more specific. We also show solutions to improve ranking. We give an example of searches based on hyper-references, and discuss the taxonomy of web-mining.
- In the chapter *Optimizing Web* pages, we study what influences the order of the search results. If we know what features determine where our web page is put in the ranking, then optimizing pages designed for the web can improve the place taken by our page and with that better ranking the information can reach more people. Thus knowing about the characteristics of free word search engines can be useful in designing documents for the web.

## 2.2 A Survey on How Students Seek Information on the Internet

Navigating among the information available on the Internet has become an expectation for the members of the information society we are living in. This especially applies to students of higher education, the intellectuals of the future.

There are a large number of papers discussing users' searching strategies and behaviour [iProspect,2004, Greenspan,2002, Nielsen,2001, Environics Research Group, 2001]. A Canadian study [Gunn&Hepburn,2003] undertaken among 12<sup>th</sup> grade high school students examining their Internet search skills was of special interest to us.

We made a survey among the 1<sup>st</sup> year students of the Faculty of Technical Engineering at Debrecen University on how they can navigate on the web. During the seminars we could see that:

- our students search using only a few key words;
- the majority is not familiar with the possibilities provided by free word search engines to specify the required information, and they cannot narrow down or filter the result set;
- they are not aware that certain information needs require different search tools.

Seeing the flaws we were also curious to find out how widespread this phenomenon is, that is how students at other faculties of the university search the web.

1. With this object we developed questionnaire that was answered by 200 full time students of the University of Debrecen (UD) representative of the ratio of the number of students at the faculties. We were seeking to answer the following questions in connection with the students' knowledge of information location:
2. What Internet services do the students use and how often?
3. What information sources do they prefer in their study related assignments next to their textbooks?
4. How do they perceive their own ability to locate information on the Internet?
5. What Internet access possibilities do students have?
6. How do students learn to seek information on the Internet?
7. What knowledge do students have of the various Internet search engines?
8. What strategies and techniques do students use to enhance information seeking on the Internet?
9. What influences students' search skills?
10. What searching habits do students have?
11. How do they rate the effectiveness of search engines?

We compiled the answers given to our questions and analysed them and also compared them with other similar survey results.

### **Main Results:**

- Among the students of Debrecen University the most frequently used Internet services are the web (99%) and the e-mail (86%). These percentages are the

same as those in international surveys [Greenspan,2002, Environics Research Group, 2001]. Our students search for information on the web for personal purposes more often than for their studies.

- Most students at the University of Debrecen equally are in favour of the World Wide Web and books, but all in all there is a majority of those who prefer print sources in addition to textbooks. We find this fact a bit surprising as textbooks contain the course material, while the corresponding research results are mostly available on the World Wide Web first: it takes a considerably longer time to publish print material than publishing on the World Wide Web, and due to the limited financial resources many print materials aren't available in the libraries of the institutions.
- Although most students use few effective strategies when searching World Wide Web for information and use even fewer services to limit and refine the search results they consider their search skills to be average or good. Positive self assessment is even more obvious with Canadian students.
- Most students at UD pleased possibilities to connect to the Internet in their educational institutions. Half of students reported that they had Internet connection at home.
- The results of the surveys show that students rely mostly on themselves (60%) and their mates (44%) when learning how to locate information on the World Wide Web, and teachers have a less important role (21%).
- The 3 most popular search engines in the USA (Google, Yahoo, MSN) [SearchEngineWatch] are among the best-known among the students. Ask Jeeves, which is popular among American and Canadian high school students, is fairly unknown among our students. The most popular search engine in Hungarian is Vizsla24 (with 85%), then come Heuréka and Góliát.
- According to our survey students use 2.35 search engines on average at a time and 1.13 search engines regularly, which gives an average of 3.48 search engine use with some regularity. This number is higher, than the corresponding one in the Canadian survey, where this number is 1-2 among the users. [Greenspan,2002, Gunn&Hepburn,2003]
- When specifying their information requirements 76% of the students at Debrecen University do not use logical operators, 54% do not use expressions, 66% do not use substitute characters, and 64% do not use refined search. The results among Canadian students were very similar.
- 90 % of our students are not familiar with techniques for narrowing down search results (e.g. filter by file format, date, occurrence, domain and language).
- When analysing our survey we would have liked to answer the question what students' search skills depend on, more precisely if it is influenced by the student's sex, their interest in computer science, their access to the Internet or at which faculty they study. Based on the answers of the students we could not statistically find a significant correspondence between access to the Internet

(either at home or in the educational institution) and search skills. Statistical tests have shown, on the other hand, that search skills are influenced by the interest in computer science. Various tests have shown a correlation between the students' sex and their search skills, but further investigations (two-conditional variation analysis) have proved that sex influences search skills through the interest in computer science, but on its own, it is independent from it. We have come to the same conclusion examining the search skills of students studying at the various faculties: there was a significant difference in results among the students studying at the Faculty of Informatics and the students studying at the other faculties.

- When the first search is not successful is typical for our students to refine the search query (85%) and only 12% will try another search engine. Our student overview a larger portion of the search results than average users do.
- Various estimates are available on the web on the size of the invisible web, which say that it is more than 10 times larger than the visible web [Bergman, 2001, BrightPlanet]. There have been technical efforts taken to make the invisible web visible [Princz, Rutkovszky, 2004]. The invisible web is growing very fast, so making its content available is an urgent need. Using free word search engines users can find a sufficient amount of information on the web in almost any topic, so they are likely to suppose that they can reach all information available on the web with this software. This is what our survey also shows [Princz, 2006]. The estimates, given by students at the various faculties regarding the size of indexed web sites, ranged from 48% to 72%, which gives an average of 64% at the university. These are very high percentages and they show an overestimation of search engines.

### **2.3 A Pedagogical Experiment to Improvement Searching Skill**

One of our survey results is that in most cases students rely on themselves and their peers when learning search skills, and only 5% of them claimed that their teachers also helped them.

Our hypothesis is that completing suitable search tasks with teacher guidance, discussing the applied strategies and techniques can improve students' search skills.

The independent variables of the experiment were:

- a task sheet we developed to familiarize search strategies and techniques,
- a teaching strategy suitable for the goals and didactic needs of the experiment.

The dependant variable in the experiment is the varying performance depending on the various information requirements.

Our experiment is a single group experiment [Falus et al.,2000], where the initial state is measured on the basis of performance given while searching for information requiring various search strategies and techniques.

For the teaching of search skills, just like for basic skills and knowledge, direct methods are the best [Falus et al.,2003]. The main idea of direct instruction is that students acquire the basic skills and knowledge if we break the goals down into sub-goals and with firm teacher guidance students are led along the process. In this method students' individual practicing is vital. At the end of the experiment the result is provided by the performance given in a test asking for the acquired knowledge.

The result of the experiment supports that after teaching information retrieval from the web the use and familiarity with search tools and techniques needed to gain the required information have considerably improved.

### **3. Summary**

Our globalized world expects the students to learn how to widen their knowledge rather than acquiring specific knowledge. Finding information is of vital importance. The Internet is an endless warehouse of human knowledge, where information search skills are basic for everybody.

It is a rule of the thumb that knowing the theoretical background to a practical problem helps solving it, so knowing about the theory of information retrieval and the characteristics of web search tools can improve our search skills.

Most students are not familiar with the possibilities, so their awareness needs to be increased in search strategies and in choosing and using the appropriate search tools. Experience shows that the majority uses free word search engines even when a thematic engine would be appropriate.

According to our survey most of our students do not know about multiple and refined searches. So to increase their efficiency they need to learn about the various techniques applicable in information retrieval, with which the information need can be made more specific and the result set narrowed down, and thus the time consumption decreased.

Knowing about the characteristics of free word search engines helps not only with the better specification of the information requirement but also with designing web documents. When teaching web page design it is also worth discussing what features of the document influence their ranking in certain free word search engines. By optimizing our web pages, we can improve their ranking.

The role of the teacher in developing search skills should be increased. Our results have proved that one or two didactically appropriate classes can considerably improve students' search skills.

## Irodalomjegyzék, References

- [1] Amazon <http://www.amazon.com/>
- [2] Arocena,G., Mendelzon,A., Mihaila,G. (1997) Applications of a Web Query Language, *Computer Networks and ISDN Systems archive Volume 29*
- [3] Attar,R., Fraenkel,A.S. (1977) Local feedback in full-text retrieval systems. *Journal of the ACM*
- [4] Baeza-Yates,R., Ribeiro-Neto,B., (1999) Modern Information Retrieval, Addison Wesley
- [5] Barabasi, A-L., Albert, R., Jong H.(1999) Diameter of the World Wide Web. In: *Nature*
- [6] Barlow, L. (2001). The spider's apprentice: A helpful guide to Web search engines. Monash Information Services from <http://www.monash.com/spidap.html>
- [7] Bergman, M.K. (2001). The Deep Web: Surfacing Hidden Value. *Journal of Electronic Publishing*, 7(1).
- [8] Beyond General World Wide Web Searching, UC Berkeley Library <http://www.lib.berkeley.edu/>
- [9] Bharat,K., Broder. A.Z. (1998) *A technique for measuring the relative size and overlap of public web search engines*. In WWW7
- [10] Bilal,D. (2000) Children's use of Yahoo! Web Search Engine: 1. Cognitive, physical, and affective behaviors on fact-based search tasks. *Journal of the American Society for Information Science*, 51 (7)
- [11] Blackwell Online <http://bookshop.blackwell.co.uk/>
- [12] Billelloch,G. (1997) Algorithms in the Real World <http://www.cs.cmu.edu/~guyb/real-world/indexing/>
- [13] Breadth first search and depth first search (1996) ICS 161: Design and Analysis of Algorithms, Lecture notes for February 15, 1996 <http://www1.ics.uci.edu>
- [14] Brewington, B.E. & Gybenko, G. (2000) How dynamic is the Web? WWW9.org. <http://www9.org/w9cdrom/264/264.html>
- [15] BrightPlanet <http://www.brightplanet.com>
- [16] Brin,S. és Page,L. (1998) The Anatomy of a Large-Scale Hypertextual Web Search Engine WWW7 / *Computer Networks* 30(1-7), pp. 107-117, <http://www.stanford.edu>
- [17] Brooks, T. (2003) Web search: how the Web has changed information retrieval, *Information Research*, Vol. 8 No. 3, April 2003
- [18] Bujdosó Gy., Students should learn to use word processors, *Teaching Mathematics and Computer Science*, Debrecen
- [19] Children and their Use of the Internet <http://www.pages.drexel.edu/~ab352/>
- [20] Cho,J. és Garcia-Molina,H. (2000) The Evolution of the Web and Implications for an Incremental Crawler. To appear in the 26th Intl. Conf. on Very Large Databases



- [21] Cockburn, A. & McKenzie, B. (2001) 'What do Web users do? An empirical analysis of Web use' *International Journal of Human-Computer Studies*, **54**(6), <http://www.cosc.canterbury.ac.nz/andrew.cockburn/papers/ijhcsAnalysis.pdf>
- [22] Croft, W. B. és Harper, D. (1979). *Using probabilistic models of document retrieval without relevance information*. *Journal of Documentation* 35
- [23] Development of a European Service for Information on Research and Education <http://www.lub.lu.se/desire>
- [24] Dictionary of HTML META Tags <http://vancouver-webpages.com/META>
- [25] Dublin Core Metadata <http://dublincore.org>
- [26] Environics Research Group (2001) *Young Canadians in a wired world: the students' view* Report prepared for the Media Awareness Network and the Government of Canada. <http://www.media-awareness.ca/>
- [27] European Commission (2000) Memorandum az egész életen át tartó tanulásról, <http://www.om.hu/main.php?folderID=1027>
- [28] Falus, I. et al. (2000) Bevezetés a pedagógiai kutatás módszereibe, Műszaki Könyvkiadó
- [29] Falus, I. et al. (2003) Didaktika : elméleti alapok a tanítás tanulásához, Nemzeti Tankönyvkiadó
- [30] Friedman, E., Uher, M., Windhager, E. (2003) Keresés a Világhálón, *Híradástechnika* 58(3)
- [31] Greenspan, R. (2002) Search Engine Usage Ranks High, *CyberAtlas*, [http://cyberatlas.internet.com/markets/advertising/article/0,5941\\_1500821,00.html](http://cyberatlas.internet.com/markets/advertising/article/0,5941_1500821,00.html)
- [32] Griffiths, J., Brophy, P. (2002) Student searching behaviour in the JISC Information Environment, <http://www.ariadne.ac.uk/issue33/edner/>
- [33] Gulli, A., Signorini, A. (2005) The Indexable Web is more than 11.5 billion pages, WWW14, <http://citeseer.ist.psu.edu/context/2655616/0>
- [34] Gunn, H. és Hepburn, G. (2003) Seeking Information for School Purposes on the Internet *Canadian Journal of Learning and Technology*
- [35] How Much Information: Internet (2000), <http://www.sims.berkeley.edu/research/projects/how-much-info/internet.html>
- [36] Hu, Chen, Schmalý, Ritter (2001) An Overview of World Wide Web Search Technologies, <http://www.eng.auburn.edu/users/wenchen/publication/overview.ps>
- [37] Information Seeking Behavior <http://www.librarysupportstaff.com/infoseek.html>
- [38] Inktomi WebMap. <http://www.inktomi.com/webmap/>
- [39] Invisible Web <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet>
- [40] iProspect (2004) Search Engine User Attitudes Survey <http://www.iprospect.com>
- [41] Kleinberg, Kumar, Raghavan, Rajagopalan, Tomkins (1999) The Web as a graph: Measurements, models and methods. Invited survey at the International Conference on Combinatorics and Computing
- [42] Kleinberg, J. (1998) Authoritative sources in a hyperlinked environment, 9th ACM-SIAM Symposium on Discrete Algorithms

- [43] Koehler, W. (1999) Digital libraries and the World Wide Web sites and page persistence. *Information Research*, 4, (4) <http://informationr.net/ir/4-4/paper60.html>
- [44] Lawrence, S., Giles, L. (1998) Searching the World Wide Web, *Science*, April 3
- [45] Lawrence, S., Giles, L. (1999) Accessibility and Distribution of Information on the Web, *Nature*, Vol. 400, pp. 107-109
- [46] Markwell, J. & Brooks, D.W. (2002) Broken links: just how rapidly do science education hyperlinks go extinct? Lincoln, NE: University of Nebraska-Lincoln. Department of Biochemistry. [http://www.class.unl.edu/biochem/url/broken\\_links.html](http://www.class.unl.edu/biochem/url/broken_links.html)
- [47] Melnik, S., Raghavan, S., Yang, B. és Garcia-Molina, H. (2001) Building a Distributed Full-Text Index for the Web WWW10, Hong-Kong
- [48] Morris, C. So What's the Problem? [http://www.wdvl.com/Internet/Dead\\_SearchEngines](http://www.wdvl.com/Internet/Dead_SearchEngines)
- [49] Mozaik Kiadó <http://www.mozaik.info.hu/HOMEPAGE/HUNMWRI/HIREK.HTM>
- [50] Navarro, G. (1998) Approximate Text Searching, PhD tézis, Dept. of Computer Science, Univ. Of Chile
- [51] Nemzeti Tankönyvkiadó <http://www.nettankonyv.hu/web/guest/home>
- [52] Nielsen, J. (2001) Search: Visible and Simple *Alertbox*, <http://useit.com>
- [53] On the size of the World Wide Web (2001) <http://www.pandia.com/sw-2001/57>
- [54] O'Neil, E., Lavoie, B., Bennett (2003) Trends in the Evolution of the Public Web, *D-Lib Magazine*
- [55] Országos Széchényi Könyvtár [http://www.oszk.hu/index\\_hu.htm](http://www.oszk.hu/index_hu.htm)
- [56] Perinotti, T. (1997) How Search Engines Work, *Microsoft Interactive developer*
- [57] Princz, M. (2001) Knowledge Discovery in Databases, ICAI, Eger
- [58] Princz, M., Rutkovszky, E. (2004) Content Discovery of Invisible Web, ICAI, Eger
- [59] Princz, M. (2006) A Survey on How Students Seek Information on the Internet, Teaching Mathematics and Computer Science, Debrecen
- [60] Recommended Sites and Search Techniques <http://library.albany.edu/internet/search.html>
- [61] Rutkovszky, E., Princz, M. (2001) XML - the ideal language for weaving a better web, MicroCAD International Computer Science Conference
- [62] Salton, G. and Buckley, C. (1990) *Improving retrieval performance by relevance feedback*, *Journal of the American Society for Information Science*, vol. 41.
- [63] Schneiderman, B. (1997) Designing the User Interface: Strategies for Effective Humancomputer Interaction, Addison-Wesley
- [64] SearchEngines <http://www.searchengines.com/>
- [65] SearchEngineShowdown <http://www.searchengineshowdown.com>
- [66] SearchEngineWatch <http://searchenginewatch.com>
- [67] Shapiro, Y., Lehoczky, E. Factors that influence search engines rankings <http://www.searchengines.com>

- [68] Shivakumar, N.; Garcia-Molina, H.(1998) Finding near-replicas of documents on the web, Workshop on Web Databases
- [69] Sullivan,D. (2001) Search engine math. *Search Engine Watch*,  
<http://searchenginewatch.com/facts/math.html>
- [70] The Web Robots FAQ...<http://www.robotstxt.org/wc/faq.html#log>
- [71] Vargha, A. (2000) Matematikai statisztika, Pólya Kiadó
- [72] W3 <http://www.w3.org/>
- [73] W3C Metadata Area <http://www.w3.org/Metadata/>
- [74] WebReference <http://www.webreference.com>
- [75] What Is Natural-Language Searching? <http://www.nwc.com/1120/1120f1side2.html>
- [76] Wikipedia <http://www.wikipedia.org/>
- [77] Witten,I., Moffat,A., Bell,T. (1999) *Managing Gigabytes: Compressing and Indexing Documents and Images*, Morgan Kaufmann Publishing, San Francisco, ISBN 1-55860-570-3.
- [78] Zobel,J.,Moffat,A.,Ramamohanarao,K.. (1998) Inverted Files Versus Signature Files for Text Indexing

## **Publikációs jegyzék, List of Publications**

### **Referált publikációk**

#### *Reviewed publications*

- [1] Princz M. Knowledge Discovery in Databases  
5<sup>th</sup> International Conference on Applied Informatics, Eger, 2001, p. 201-212,  
[Zentralblatt pre02065284]
- [2] Princz M. A Survey on How Students Seek Information on the Internet  
Teaching Mathematics and Computer Science,  
Debrecen, 4/1 (2006), p. 153-165
- [3] Princz M. Computer Aided Problem Solving  
Teaching Mathematics and Computer Science,  
Debrecen, megjelenés alatt

### **Lektorált publikációk**

#### *Refereed publications*

- [4] Princz M. A weben lévő információk hozzáférhetősége  
Debreceni Műszaki Közlemények, 2004/1, p. 51-66
- [5] Princz M. E-learning  
Debreceni Műszaki Közlemények 2005/1, p. 93-113

### **Egyéb publikációk**

#### *Other publications*

- [6] Princz M. Keresési stratégiák a Web-en  
DE MFK Tudományos Közlöny 2000, p. 209-212., Debrecen

### **Konferencia kiadványban megjelent cikkek és előadások**

#### *Conference proceedings and lectures*

- [7] Princz M. Systems to access information in the Web  
MicroCAD'2000 International Computer Science Conference, p 169-173,  
Miskolc, 2000
- [8] Princz M., Rutkovszky E. Datamining  
MicroCAD'2001 International Computer Science Conference, p 117-125,  
Miskolc, 2001
- [9] Rutkovszky E., Princz M. XML - the ideal language for weaving a better web  
MicroCAD'2001 International Computer Science Conference, p 209-214,  
Miskolc, 2001
- [10] Princz M., Rutkovszky E. Content Discovery of Invisible Web  
6<sup>th</sup> International Conference on Applied Informatics, Eger, 2004
- [11] Princz M. Információkeresési stratégiák a weben és tanításuk tapasztalatai  
14<sup>th</sup> International Conference in Computer Science and Education  
ISBN 973-86097 p 28-34, Kolozsvár, 2004.

- [12] Princz M. Hallgatók webes Információkeresési készségének felmérése  
16<sup>th</sup> International Conference in Computer Science and Education  
ISBN-10 973-7840-12-7 p.56-68, Szováta, 2006.
- [13] Princz M. Az informatika oktatásának helyzete a KLTE MFK-n  
Informatika a felsőoktatásban II. kötet p. 641-645., Debrecen, 1999
- [14] Princz M., Rutkovszky E. Trendek az adatbányászatban  
NetworkShop 2001, Sopron, CD melléklet  
<http://www.niif.hu/rendezvenyek/networkshop/>
- [15] NetworkShop 2001, Sopron, CD melléklet  
<http://www.niif.hu/rendezvenyek/networkshop/>
- [16] Princz M., Rutkovszky, E. Ismeret reprezentáció a weben  
NetworkShop 2002, Eger, CD melléklet  
<http://www.niif.hu/rendezvenyek/networkshop/>
- [17] Princz M., Husi G. A webes keresők használatának tanítása  
Informatika a felsőoktatásban, Debrecen, 2002, CD melléklet
- [18] Husi G., Princz M. Informatikával támogatott oktatási projekt a Hajdú-Bihar  
Megyei Rendőr-főkapitányságon  
Informatika a felsőoktatásban, Debrecen, 2002, CD melléklet
- [19] Princz M. A weben lévő információk hozzáférhetősége  
NetworkShop 2003, Pécs, CD melléklet  
<http://www.niif.hu/rendezvenyek/networkshop/>
- [20] Princz M. A webes keresők tanításának tapasztalatai  
E-learning alkalmazások a hazai felsőoktatásban,  
Budapest, 2003., p. 185-190  
<http://www.szamalk.hu/okk/E-learning/Program.htm>
- [21] Princz M. Tájékozódás a weben  
NetworkShop 2004, Győr, CD melléklet  
<http://www.niif.hu/rendezvenyek/networkshop/>
- [22] Princz M., Nagy J. A hallgatók keresési szokásai  
Informatika a felsőoktatásban, Debrecen, 2005, CD melléklet

### **Elektronikus formában elérhető tananyagok**

#### *Electronic subject-matters*

- [23] Műszaki informatika II (MF MUI 02M03) tárgy előadásainak és gyakorlatainak  
tananyaga (DE MFK) <http://delfin.unideb.hu/~pmaria/> (folyamatos  
fejlesztés alatt)

### **Előadások**

#### *Lectures*

- [24] Nagyszámítógépes és mikrogépes rendszerek működése a SZÜV-ben  
Csillag Pál Szimpózium, Balatonlelle, 1988
- [25] Az Internet Kukac  
SZÜV, Debrecen, 1990

- [26] Adatbányászat  
INFO ÉRA 2000, <http://www.szikszi.hu/info/adatbanyaszat.html>  
Békéscsaba, 2000
- [27] Információkeresési stratégiák  
DE MFK ÁMI Intézeti tudományos nap, Debrecen, 2000
- [28] Adatbázisrendszerek jövője  
DE MFK ÁMI Intézeti tudományos nap, Debrecen, 2000
- [29] Keresőgépek  
INFO ÉRA 2001, <http://www.infoera.hu/infoera2001/info4.shtml>
- [30] E-business  
DE MFK Gépészeti Szakmai Napok, Debrecen, 2003
- [31] A hallgatók keresési szokásai  
INFO ÉRA, Békéscsaba, 2005
- [32] Felmérés a hallgatók keresési készségéről  
Menedzsment és Alkalmazott Informatika Szeminárium, Cserépfalu, 2006

## Szoftver

### *Software*

#### *A Szerzői Jogvédő Hivatalnál bejegyzett szoftvertermékek*

1983-1991 között a SZÜV, Debrecen munkatársaként az alábbi szoftverek fejlesztésében vettem részt programozóként. A szoftverek PL1-ben, illetve Clipper programnyelven íródtak.

- [33] Elő- és utókalkulációs rendszer – Medicor (kb. 3500 sor)
- [34] Bérrendszer – Hajdú Megyei Tejipari Vállalat (kb. 1900 sor)
- [35] Termelésirányítási rendszer – Hajdú Megyei Tejipari Vállalat (kb. 1500 sor)
- [36] Értékesítési rendszer – Barnevál (kb. 2700 sor)
- [37] Üzletelszámoltatási rendszer – Hajdú Megyei Vendéglátóipari Vállalat (kb. 1500 sor)
- [38] Termelés elszámolás – Hajdú Megyei Húsipari Vállalat (kb. 1700 sor)
- [39] Állóeszköz nyilvántartási rendszer – TŰZÉP, Nyiregyháza (kb. 7000 sor)
- [40] Osztott-keret nyilvántartás – DOTE (kb. 2500 sor)

#### *Egyéb szoftvertermékek*

Az alábbi rendszereket 100%-ban magam készítettem:

- [41] Menetlevél nyilvántartási, számlázási rendszer – HOMEX Kft, 1992 (kb. 5000 sor)
- [42] Fuvar-elszámoltatási rendszer – AGROS Kft, 1993 (kb. 4000 sor)
- [43] Menetlevél nyilvántartási, számlázási rendszer – Herczeg Kft, 1993 (kb. 5000 sor)
- [44] Hardver nyilvántartási rendszer – Ybl Miklós Műszaki Főiskola, 1994 (kb. 2000 sor)
- [45] Főiskolai Gopher adatbázis létrehozása – Ybl Miklós Műszaki Főiskola, 1994
- [46] Általános és Menedzserképző Intézet honlapja – DE MFK, 2000