



MASZEKER HÍRLEVÉL - 1. szám - 2009. november

A Hírlevél az ALL (Applied Logic Laboratory, Budapest) és az SZTE (Szegedi Tudományegyetem) Nemzeti Technológia Program keretében futó közös kutatás-fejlesztési projektje, a MASZEKER (Modell-Alapú Szemantikus Kereső Rendszer) részeként jelenik meg. (Ld. <http://www.maszeker.hu/>)

Szerkesztő:

Pajor Enikő pajor@jgypk.u-szeged.hu

Középpontban a sziget-ontológia: elindult a MASZEKER-program.

A 2009 elején indult kutatás célja a jól bevált „search by matching strings” paradigmán alapuló hagyományos keresők és a sokat ígérő, de az internet lényegi szabályozatlansága miatt a szigorú követelmények megvalósítására kevés esélyt nyújtó „szemantikus web” két nagy paradigmáján túllépve egy harmadik, a keresés fogalmának mélyebb és általánosabb értelmezésére épülő, új elveken alapuló, ún. „sziget-ontológiás”, integrált multimodális keresőrendszer prototípus-szintű kifejlesztése.

A projekt során számos szövegnyelvészeti rész-fejlesztésre kerül sor, és két konkrét alkalmazási területen, szabadalmi adatbázisokban és néprajzi szövegtörzsekben folyik majd az eszköz tesztelése.

A munkálatokkal párhuzamosan a kutatók igyekeznek folyamatosan figyelemmel kísérni a keresőrendszerek fejlesztésével kapcsolatos legfrissebb fejleményeket, és azokról ebben a hírlevélben folyamatosan tudósítani.

A tartalomról

A hírlevél e száma először a nemzetközi szakirodalom közismert szerzőpárosának több évtizedes, az információkeresés elméleti és gyakorlati szempontjait összegző monográfiáját mutatja be. Tekintettel arra, hogy az információkeresés egyik fontos technikai elemeként a szerzők kiemelik a keresőfelület, az interfész szerepét, a hírlevél e száma a továbbiakban összegyűjti a legfontosabb ezzel kapcsolatos kutatásokat, a felhasználó és a gép kapcsolatára vonatkozó jellemzőket, majd olyan régebbi-új keresők neve és URL címe zárja a gyűjtést, amelyek keresőablaka, vagy találatainak a megjelenítése valamilyen szempontból érdekes lehet a MASZEKER kutatói számára.

Egy kiemelkedően fontos mű



Ingwersen, Peter-Jarvelin, Kalervo: The turn : integration of information seeking and retrieval in context. Dordrecht : Springer, 2005. XIV, 448 p. (The Kluwer international series on information retrieval; 18.) >>> [2009-11-11]

Ism.: Dudás Anikó: A fordulat: információkeresés-kutatás és perspektívája: a kognitív szemlélet In: Könyvtári Figyelő, 18. (54.) évf., 2008. 3. sz. 505-511. p. >>>

A kötetből és Dudás Anikó ismertetéséből csak azokat a részeket emelem ki, amelyek a kereső gyakorlati megvalósítása szempontjából érdekeseek lehetnek a kollégáknak.

A kötet megismerését a MASZEKER kutatói számára elengedhetetlenül fontosnak tartom, mert pontosan azt elemzi, hogy a "hatékony keresőeszközök megteremtéséhez éppen annyira nélkülözhetetlen az emberi információs magatartás, a feladatorientált kontextus figyelembevétele, mint amennyire megkerülhetetlen a problémamegoldó helyzet, a tudásállapotok, kognitív struktúrák vizsgálatánál azoknak a technikai eszközöknek a figyelembevétele, amelyeket az emberek információszerezésre használnak."

Kiindulópontjuk a visszakeresésekre használt különféle algoritmusok elemzése az ún. Cranfield-paradigma segítségével, amelyet azért fejlesztettek ki, hogy a visszakeresések hatékonyságát fokozzák és eredményeit mérhetővé is tegyék. Ezt sok példával illusztrálják a szerzők, hogy megvilágítsák azt a tényt, hogy bármilyen módon is szerkesztjük meg keresési algoritmusunkat - kontrollált, prekoordinált fogalmi szótár, tezaurusz vagy posztkoordinált, szövegszavas automatizált index, relevancia-kalkulációval előállított lista segítségével dolgozik a keresőgép/felhasználó -, hiba, félreértés mindig marad a rendszerben. Ennek két kiküszöbölése lehetséges:

1) A Peter Ingwersen által kitalált "relevancia teljesítményt fokozó polireprezentáció elvén alapuló ún. kognitív átfedés (cognitive overlap) kiaknázása, melyek során a más-más funkciót szolgáló adatfélések megfeleltetésével érhető el magas relevanciaszint. Az eljárás pl. a dokumentum címstruktúrája, külső hivatkozásai (out-link), a dokumentumra mutató hivatkozások, vagy az azt idéző utalások (in-link), tezauruszok, folyóiratcímek, indexelők által hozzáadott deskriptorok, műfaji megjelölések stb. adatelemeinek tartalmi átfedéséből "keveri ki" a legrelevánsabb találatot.

2) A **használoi felület fejlesztése**, hogy ne csak közvetítsen és biztosítsa az információ visszakeresését, hanem magyarázataival, szemléltető példáival segítőként be is kapcsolódjon a keresési folyamatba helyzetérzékenyen ott, ahol a felhasználó megakad.

A keresésről és a keresési technikákról

AZ USA-ban állítólag egyre okosabban keresnek az internetezők: eljutnak akár a négy kulcsszó alkalmazásáig is!

Melius WEIDEMAN és Corrie STRÜMPFER már 2003 óta kutatásaiban arra kerestek választ, hogy a kulcsszó megválasztása és a demográfiai jellemzők hogyan befolyásolják a keresés sikerét. Három kontinens 46 intézményében 1109 kereső magatartását vizsgálták és arra a megállapításra jutottak, hogy a használt kulcsszavak száma erőteljesen befolyásolja az eredményt éppúgy, mint a felhasználók kora (kiemelkedően jó a 20-25 évesek keresése) és a fajhoz való tartozás is. A fehérek a leggyesebbek, majd a feketék, végül az ázsiaiak keresnek a legkisebb sikerrel. >>>

Hogy mennyire meghatározó az ember-gép interakciója során a közérthetőség, mi sem bizonyítja jobban, mint az, hogy a marylandi egyetem kutatói már egy évtizede kifejlesztették a **QUIS = Questionnaire for user interaction satisfaction** nevű kérdőíves rendszerüket, melynél arra törekedtek, hogy olyan általánosságokra kérdezzenek rá, amelyek általában mindenféle számítógépes alkalmazásnál előfordulnak. A kérdőív negyedik pontja foglalkozik részletesen az interfész megjelenítésének kérdéseivel. A hosszabb változatú kérdőívet **on-line változatban le lehet tölteni és ki lehet próbálni** saját alkalmazásunkon is:

A weboldalak ergonómiájának és design-jának egyik legismertebb kutatója Jakob Nielsen és kutatócsoportja minden évben a világ 4 pontján Usability Week címmel workshop jellegű konferenciát rendez, melyen saját és más jelentős kutatók eredményeit mutatják be az érdeklődőknek. **Az ez évi Usability week 2009 konferencia** szeptember 14-én kezdődik és négy hónapon keresztül négy különböző városban (Edinburgh, Las Vegas, Berlin, San Francisco) egy-egy hétig elméleti és gyakorlati előadások, workshopok keretében vizsgálja a webre tett tartalmak, alkalmazások és a felhasználók kapcsolatát. A csoport által szervezett konferenciák és weboldalak elsősorban az e-kereskedelem, az üzleti információ, a B2B (business to business) stb. területeihez kapcsolódó felhasználói magatartásmintákat és jellemzőket elemzik, de lehetnek olyan architektúrák a dokumentációkban, amelyek egyes elemei az ALL jelenlegi munkájához ötletet adhatnak. **A publikációk elérése**

Az egyik konferencia keresésekkel és a felhasználók navigálásával kapcsolatos ismeretei (**Search and navigation**) különösen érdekelhetik az ALL fejlesztőket.

Tanulmányok, cikkek, hírek a felhasználói felülettel kapcsolatban

E kérdéskörhöz kapcsolódó, ám jelenleg csak fizetés ellenében elérhető két részes cikk végső következtetésére hívom fel az ALL kutatóinak a figyelmét. >>>

Az amerikai Karen MARKEY (MARKEY, Karen: Twenty-five years of end-user searching Part 1-2. In: Journal American Society for Information Science and Technology, 58. vol. 2007. no. 8. 1071-1081. p.; és 1123-1130. p.)

megvizsgálta az on-line használói információkeresési szokások utóbbi 25 évét, az erről írott cikkeket és tapasztalatokat.

Fő kérdéskörei: Mi a jellemző a használók által feltett kérdésekre? Milyen eszközöket használnak a kereséskor és mire volna szükségük a kérdés pontosításához? Hogyan reagálnak a kapott találati halmazra?

Végkövetkeztetés:

- a használók a lehető legegyszerűbb megoldásokat szeretik (kevés kifejezés, egyetlen lépés, egy kulcsszó)
- a fejlesztőknek a továbbiakban is az információkereső felületek egyszerűségére kell törekedniük

A felhasználók begyakorlott, általában mindig azonos keresési technikákat alkalmaznak, amely tapasztalataik szerint már jól bevált. A szakirodalom ezt mentális modellnek nevezi. Bizonyított tény, hogy keresés közben három fontos elem megjelenését várják a keresőmotortól:

a) egy keresőablakot, amelybe beírhatják a keresett kifejezést,

b) egy olyan gombot, amelyre rá van írva a KERES kifejezés (Ha a Go vagy nyíl vagy egyéb módon jelölik, már veszít hatékonyságából! Vö: User interface. a szerk.)

c) és egy lineárisan olvasható eredménylistát - pontosan úgy - mondják -, ahogy kedvenc keresőmotorjaik (Google, Yahoo, MSN) dolgoznak.

A Jakob Nielsen által írott **Mental Models For Search Are Getting Firmer**

cikk azt vizsgálja, hogy az ettől eltérő interfész milyen veszteséget okoz a kereső használatánál. Javaslatokat ad arra, hogy hogyan lehet ezt elkerülni. Pl. Nincs értelme kiemelni néhány találatot egy, a "legrelevánsabb találatok" feliratú dobozba, jobb, ha a lineáris lista elején szerepelnek ezek.

Igen érdekes Niensennek az a kutatása is, amely 232 web felhasználónak a szemmozgás-követése alapján arra a megállapításra jutott, hogy **a felhasználók a megjelenített lapokat F alakban pásztázzák át**. Erre érdemes lesz figyelni a keresőoldal webdesign-ja megtervezésekor! Magyarul lásd: **A mágikus F betű**

Az eredeti cikk angolul

Amint látható, az internetes keresők felhasználóit vizuális és gyakorlati szempontból két terület

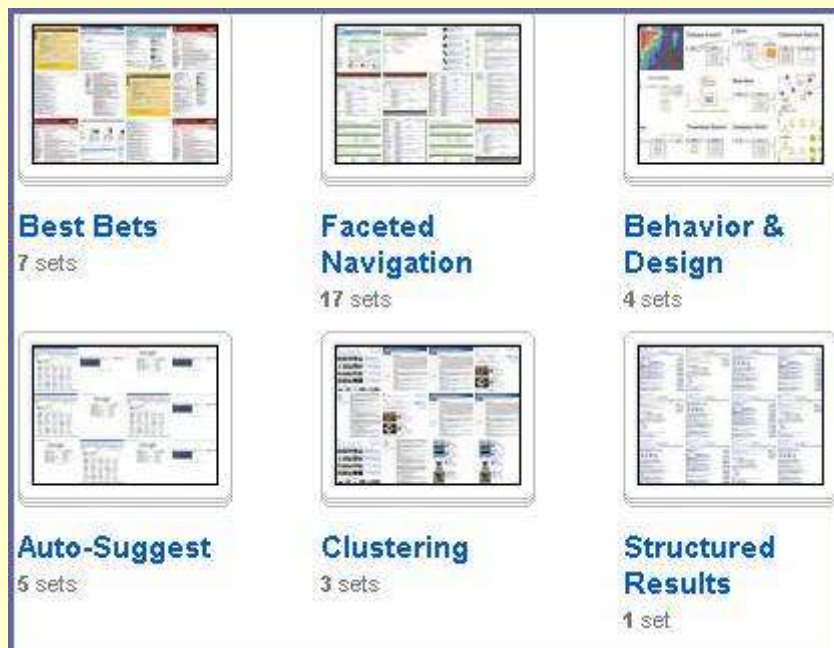
érdekl:

a keresőablak, a találatok és az "user interface", a felhasználó felület külső megjelenése és mindezek funkcionalitása, könnyen kezelhetősége

A **Search User Interface and User Experience** weboldal elsősorban tanácsokat és számos linket ad a fenti kérdésre vonatkozóan. Az összegyűjtött linkek a technikai megoldásokat éppúgy körüljárják, mint a pszichológiai, gyakorlati területeket. Szemléltetésül álljon itt néhány jótanácsuk a kereső felhasználói felületét illetően:

- A site minden oldalán legyen egy egyszerű, de megfelelően hosszú keresőmező
- Közérthető szavakkal, minden technikai szakszó mellőzésével kell elmagyarázni a használatot
- A kitett ikonok mindegyikéhez hozzá kell rendelni egy-egy label-t/etikette-t
- Feltétlenül kerülni kell a teljesen új szerkezetű használói felület alkalmazását, mert az meg fogja zavarni a felhasználókat
- Célszerű a legnagyobb, legbefutottabb keresőgépek legjobbnak tartott interfész megoldásait követni stb.

Az egyszerű, az összetett és egyéb keresési formák pattern-jeiből gyűjtött össze számos illusztrációt a témáról "**Information Architecture**" (2006) címmel könyvet is író Peter Morville, aki ezt mondja: "I hope to add patterns that illustrate user behavior and the information architecture of search." >>>



Morville weblogja

XIE, Hong Iris: Online IR system evaluation : Online databases versus Web search engines

In: **Online Information Review**, 28. 2004. no 3. 211-219.

című tanulmányában kétféle típusú on-line információkeresési rendszert értékelt a használói kritériumok alapján: on-line adatbázisokat és internetes keresőgépeket. *Négyféle keresőgép összehasonlító vizsgálatát is bemutatja: tematikus linkgyűjtemények, általános keresőgépek, meta-keresőgépek, és speciális keresőgépek.*

Az eredmények azt mutatják, hogy a használók három elemet tartanak lényegesnek az információkeresési rendszerek értékelésénél: 1) használói felület, 2) a rendszer teljesítménye 3) és a gyűjtemény tartalma.

A résztvevők a könnyű használat és az intuitív felület fontosságát emelték ki, ugyanakkor a kapott

információk megbízhatóságát és hasznosságát is lényegesnek tartották. Az on-line adatbázisok és az internetes keresőgépek előnyeinek és problémáinak elemzése alapján a szerző szempontokat javasol az információkeresési rendszerek tervezéséhez.

Az eredeti mű a fizetős Emerald szolgáltatáson keresztül érhető el, szükség esetén beszerezem az ALL kollégák számára.

Az aktuális divatáramlatok általában érdeklik az embereket.
Erre figyelnek a **Web Design Trends For 2009**

>>>

szakemberei is, akik már az év elején csokorba gyűjtötték a legdivatosabb és legsikeresebb felületeket.
Persze itt vigyázni kell, hogy ami az átlag felhasználónak tetszik, bevált, az sikeres lehet-e az ALL keresőjét használó szakembereknek is?

Az **UI Design Newsletter**

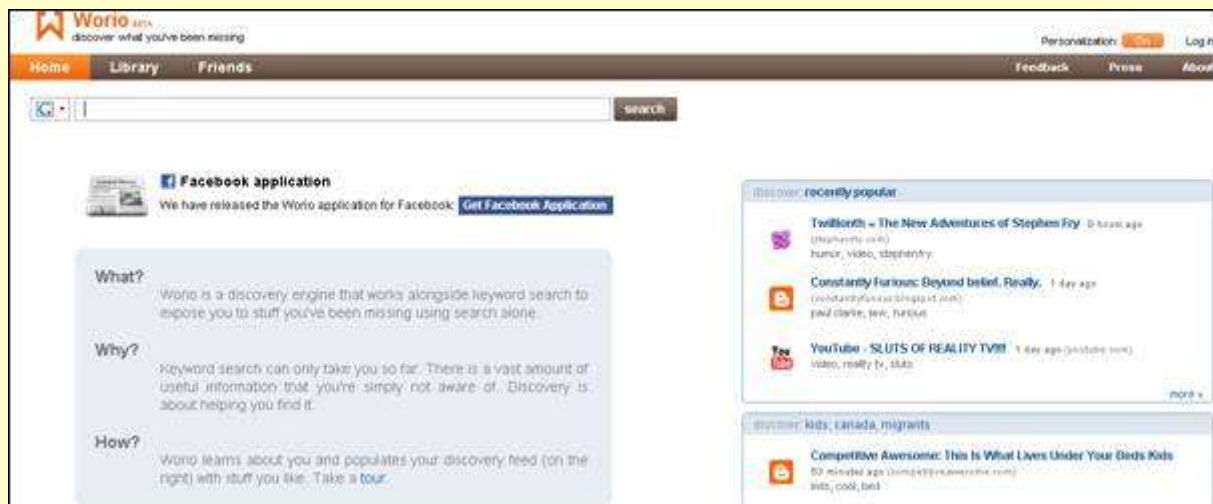
1998-2009. évi számai egy-egy problémakört járnak körül.

A **10 Useful Techniques For Good User Interface Design** cikk után megjelent a **12 Useful Techniques For Good User Interface Design** is, amelyben a szerző ötleteket, megoldásokat, illusztrációval ellátott kész példákat ad arra, hogy milyen módon tehetjük felhasználóbarátabbá internetes kereső és egyéb webfelületeinket.

Keresőmotorok

A "felfedezési" kereső: a **WORIO**

"Worio is a discovery engine that works alongside keyword search to expose you to stuff you've been missing using search alone."



A "slideshow"-t játszó kereső: az **OAMOS**

*"At Oamos, we have a single mission:
To play and discover content you'll love - and nothing else."*

Oamos

Meta-Inspiration Engine

Az Oamos úgy kereső, hogy nem kereső, noha annak nevezi magát. A beírt keresőszóra nem találati listát ad, hanem a találatok megnyitott weboldalaiból egy interaktív slideshow-t rögtönöz zenei aláfestéssel. A három perces bemutató a téma weboldalaival zsúfolja tele az állandó mozgásban lévő képernyőt.



Egy újítani szándékozó, zömében azonban megszokott metakereső: a **Myriad**

"Combines the results of Google, Yahoo, MSN Search and Ask Jeeves in one advertisement-free stream, highlighted to show the source of each result."





A sokból kettő, akik mindenáron meg akarják törni a Google egyeduralmát: a szemantikus keresést megcélzó Cuil és a Bing

"Cuil's goal is to solve the two great problems of search: how to index the whole Internet-not just part of it-and how to analyze and sort out its pages so you get relevant results."



2008. július 28. hír

"A Bing nem rosszabb a Google-nél, ..a gond az, hogy nem is jobb."



És aki "nem akar jobb lenni", mégis összefog a Binggel: a Wolfram Alpha

"A Wolfram Alpha egy hatalmas tudományos adatbázist és a hozzá kapcsolódó matematikai modellt használja a válasz kiszámításához."



2009. nov. 12. hír
Összefogott a Bing és a Wolfram Alpha

Wolfram Alpha Builds Ring Around Google

2009. május 18. hír
Elindult az első tudományos kereső

A találatokat interaktívan vizualizáló motor: a Redz (2009) és elődje a RedZee (2005)



A keresőmotorok értékelésének/tesztelésének szempontjai

Mind az információkereső szakemberek, mint a webes szolgáltatók részéről igény van arra, hogy az interaktív információkereső rendszerek, többek között a keresőmotorok teljesítményének értékeléséhez szempontokat dolgozzanak ki. A két szakmai közösségnek vannak közös problémái és egymás számára hasznosítható eredményei.

A keresőmotorokat eddig legtöbbször úgy értékelték, hogy összevetették őket, melyik eszköz "jobb" a másikkal. A klasszikus szempontokat (relevancia, teljesség, pontosság, visszahívás) megújító kritériumrendszerben figyelembe kell venni a web dokumentumainak összetettségét és a használóközpontú új szemléletet. Az értékelés öt lépése a következő:

1. Forgatókönyv készítése - annak eldöntése, mit tesztelünk, ki számára és milyen célból. Ki kell tűzni, mi a célunk: a különbségek feltárása, a jobbítás, tervezési ötletek tesztelése, bizonyos kritériumokkal való összehasonlítás, általános elvek meghatározása vagy kipróbálása. Olyan kérdésekre kell itt választ adni, hogy azt akarjuk-e meghatározni, melyik keresőmotor a legjobb, vagy azt, hogy mennyire hasznos egy bizonyos keresőmotor stb.

2. A vizsgálati kritériumok meghatározása - A relevancia értékelésénél figyelembe kell venni a webdokumentumok sajátosságait (a találatok között lehetségesek duplumok, például a tükrözés miatt; a csatolások lehetnek inaktívak; a találatok között szerepelhetnek csatolásgyűjtemények is; a legrelevánsabb találatok a lista elején találhatóak; a dokumentumok lehetnek idegen nyelvűek). A pontosságra nézve a releváns oldalak számát lehet megadni, jelezni a találati halmaz összesített relevanciáját; igénybe lehet venni egy rangsoroló algoritmust. A visszahívás mértékét Clarke és Willett úgy határozza meg, hogy az egyes szolgáltatásokból kinyert releváns dokumentumokat vizsgálja. Az egyes adatbázisokban újabb keresést kell végezni a más szolgáltatások által felderített dokumentumokra. Az újabb keresésnél hozzáférhető, de az eredeti találati listán nem

szereplő dokumentumok számát hozzá kell adni azokat a releváns dokumentumok induló számához - ez lesz egy mutatószám nevezője. A számlálóba pedig a releváns dokumentumok kiinduló mennyisége kerül. A relevancia-rangsor összeállításához két módszer használatos. A találati lista első felében lévő releváns dokumentumok számát elosztják a halmazban lévő összes releváns dokumentum számával, vagy a halmazban a dokumentumoknak a legrelevánsabb dokumentumhoz képest elfoglalt helyük szerint adnak egy értéket. A lefedés kiszámításához egy konkrét keresés esetében a hozzáférhető releváns dokumentumok számát elosztják az összes keresőmotor adatbázisaiban talált összes releváns dokumentum számával. A használhatósággal kapcsolatos szempontok a másodlagos szolgáltatásoknál: a források kiválasztása, a leírások szintje, tárgyi feltárás, értékelési kritériumok. A duplumok, az inaktív és tükrözött csatolások mennyisége azt jelzi, mennyi zajra számíthat a használó. Az állandó elérhetőség és a zavartalan keresés érdekében ellenőrizni célszerű, hányszor küld egy-egy keresőszolgáltatás "nem érhető el" vagy hibaüzenetet.

3. A keresőgépek elemzése - Az értékelendő keresőgépekről a lehető legtöbb adatot össze kell gyűjteni (az adatbázis nagyságáról, azaz pl. az indexelt weblapok számáról; a szerkesztés manuális vagy automatikus módjáról; arról, hogy az internetnek mely részét indexeli, milyen mélységig, milyen gyakori a frissítés és a teljes frissítés; a dokumentumok mely része van indexelve, regisztrálják-e a csatolásokat). További szempontok: a keresőképernyők, a keresési módszerek, a keresési eredmények megjelenítése, kiegészítő szolgáltatások.

4. A kísérlet céljai és környezete - A kutatók számára nagy segítséget jelent, ha valódi használók kéréseit használják fel, vagy akár velük végeztetik el a kereséseket (és irányelvek birtokában az értékelést). Minél több keresést célszerű elvégeztetni, lehetőleg időben egymáshoz minél közelebb. Előre meg kell határozni, mi tekinthető relevánsnak, nem relevánsnak (a duplumok, az inaktív és a tükrözött csatolások), mikor nem értékelhető egy találat (ha egy csatolás többször nem érhető el). A nyelvük miatt nem értékelhető dokumentumokat a találati halmazból egy másikkal kell helyettesíteni. Csak eredeti webdokumentumokat kell értékelni. A használói vizsgálatok tapasztalatai alapján a találati lista elején lévő húsz dokumentumot szokták a további számításoknál figyelembe venni.

5. Az eredmények értelmezése - Az eredményeket a vizsgálat célja szerint kell elemezni. Az egyes kereséseknél kapott értékeket átlagolni kell, és a keresőgépekre nézve a középértéket ki kell számítani. Egyéb statisztikai tesztekre és kiterjedtebb tesztelésre is szükség lehet az eredmények hitelesítéséhez.