

Az alábbi kivonat Marti A. Hearst: Search User Interfaces c. munkájából készült a MASZeKer projekt számára. M. M.

Tartalomjegyzék

1. Keresőinterfészek tervezése

1.1. Legyen egyszerű az interfész

1.2. Korszakváltás a keresőinterfészek tervezésében

1.3. A keresőinterfész tervezésének folyamata

1.4. Tervezési irányelvek keresőinterfészekhez

1.5. Hatékony és informatív visszajelzés

1.5.1. Keresési eredmények azonnali megmutatása

1.5.2. Informatív szurrogátumok megmutatása; Kulcssavak kiemelése

1.5.3. Eredmények különböző szempontok szerinti rendezése

1.5.4. Keresési kifejezések ajánlása

1.5.5. A relevancia indikátorok óvatós használata

1.5.6. Azonnali válasz támogatása

1.6. Egyensúly a felhasználói irányítás és az automatizált folyamatok között

1.6.1. Sorba rendezés a webes keresésben

1.6.2. Kérések átalakítása

1.7. Rövidtávú memória terhelésének csökkentése

1.7.1. Keresési javaslat a keresőmezőben

1.7.2. Előzmények támogatása

1.7.3. Navigáció és keresés integrálása

1.8. Kombinációk lehetőségének biztosítása

1.9. Hibák csökkentése

1.9.1. Tartózkodjunk az üres találati listáktól

1.9.2. A szótárprobléma megoldása

1.10. Figyelmet kell szentelni a részleteknek

1.11. Az esztétika fontossága a tervezésben

1.12. Összefoglalás

2. A keresőinterfész kiértékelése

2.1. Információvisszakeresés standard kiértékelése

2.2. Informális használhatósági tesztelés

2.3. Formális tanulmányok és irányított kísérletek

2.3.1. A formális vizsgálatokban alkalmazott technikák

2.3.2. A kísérleti feltételek prezentálását ki kell egyensúlyozni

2.3.3. Résztevők gyűjtése

2.3.4. A felhasználók preferenciáinak mérése

2.4. Longitudinális kutatások

2.5. A keresőmotor logfájljainak elemzése

2.5.1. Session határok azonosítása

2.5.2. A felhasználó identitásával kapcsolatos kérdések

2.6. Átfogó, logfájl alapú használhatósági tesztelés (vödör(?) teszt)

2.7. Speciális megfontolások keresőinterfész kiértékelésekor

2.7.1. Tartózkodni kell a kutatói elfogultságtól

2.7.2. A résztvevők motiválása

2.7.3. Számolni kell az individuális különbségekkel

2.7.3.1. A résztvevők tudása a szakterületről és feladatokról

2.7.3.2. A résztvevők korábbi keresési tapasztalatai

2.7.3.3. A résztvevők kognitív képességei

2.7.4. Számolni kell a kérések és a feladatok különbözőségével

2.7.4.1. Számolni kell a feladatok változatosságának hatásával

2.7.4.2. A keresőkifejezésben megjelenő variabilitás kontrollálása

2.7.4.3. Tartózkodni kell az előítéletektől a kérés és feladat megválasztásakor

2.7.5. A teszt kollekció karakterisztikájának figyelembevétele

2.7.6. Számolni kell a feladatok végrehajtásának időtartamában megjelenő különbségekkel

2.7.7. Összehasonlítás egy (túl) erős alappal

2.8. Konklúziók

3. Az információkeresés folyamatának modelljei

3.1. Az információkeresés standard modellje

3.2. Az információkeresés kognitív modelljei

3.3. A dinamikus (bogyószedő) modell

3.4. Fokonkénti információkeresés

3.5. Az információkeresés mint stratégiai folyamat

3.5.1. Stratégiák mint különböző taktikák szekvenciái

3.5.2. Költségszerkezet elemzés és információ szerzés elmélet

3.5.3. Bőngészés vs. keresés mint információkeresési stratégia

3.5.4. Információs nyomok információs struktúrákban történő navigációhoz

3.5.5. Tájékozódás és egyéb inkrementális stratégiák

3.5.6. Értelemalkotás: a keresés egy nagyobb folyamatba ágyazva

3.6. Értelemalkotás: a keresés egy nagyobb folyamatba ágyazva

3.7. Információs szükségletek és a keresések intenciója

3.7.1. Weblog alapú keresőkifejezés taxonómiák

3.7.2. A keresőkifejezések weblog alapú témába sorolása

3.7.3. A keresőkifejezések kétértelműségének weblog alapú elemzése

3.7.4. Újralátogatási minták weblog alapú elemzése

3.7.5. A megfigyelt keresési viselkedés osztályozása

4. A kérés specifikálása

4.1. Szöveges kérés megfogalmazás

4.1.1 Keresés szurrogátumokban vs. teljes szövegekben

4.1.2. Kulcsszavas keresések

4.1.2.1. Kulcsszavas keresések statisztikája

4.1.2.2. Kulcsszavas keresések: statisztikai sorrend vs. kapcsolat

4.1.2.3. Kulcsszavas keresések vs. természetes nyelvi keresések

4.1.3. Automatizált válaszadás

4.1.3.1. Automatizált válaszadási technikák

4.1.3.2. Választípusok kérdésszerű kérésekre

4.1.4. Bekezdésszintű szöveges keresések

4.1.5. Szöveges kérések automatikus átalakítása

4.2. Keresőkifejezések megfogalmazása beviteli mezőkön keresztül

4.3. Dinamikus kifejezés-javaslat a keresőkérés megfogalmazása során

4.4. A keresőkifejezés megfogalmazása Boolean és más operátorok segítségével

4.4.1. Kifejezések távolsága Boolean kérésekben

4.4.2. Utólag koordinált és facettázott Boolean kérések

4.4.3. Web alapú hatékonyságnövelés Boolean kéréseknél

4.4.4. Operátorhasználati statisztikák

4.5. Parancssori nyelvek használatával megfogalmazott keresőkifejezés

4.6. Konklúziók

5. A keresési eredmények megjelenítése

5.1. Dokumentum szurrogátumok

5.2. KWIC avagy kérésközpontú összefoglalók

5.2.1. Mondatválasztás KWIC-ek kapcsán

5.2.2. Összefoglalók hossza KWIC-ek esetén

5.3. Keresőkifejezések kiemelése

5.4. Az eredmények listázásának további jellegzetességei

5.5. A találati lista rendezésének hatásai

5.6. Találati eredmények vizualizációja

5.7. Következtetések

6. A keresőkifejezés átformálása

6.1. Az átformálás igénye

6.2. Helyesírás ellenőrzés és javítás

6.3. Automatizált kifejezésajánlás

6.3.1. Prisma

6.3.2. Kifejezések ajánlásának más módjai

6.3.3. A kérés pontosítására szolgáló ajánlások webes keresőinterfészekben

6.4. Kedvelt célpont ajánlása

6.5. Relevancia visszajelzése

6.6. Kapcsolódó cikkek megmutatása

6.7. Következtetések

7. A keresési folyamat támogatása

7.1. Kiindulópontok a keresés számára

7.1.1. Kiindulópontok a webes keresésben

7.1.2. Kiindulópontok online könyvtárkatalógusok esetében

7.1.3. Interaktív dialógusok mint kiindulópontok

7.2. A keresés történetének támogatása

7.3. A keresési folyamat egészének támogatása

7.4. A keresés integrálása az értelemalkotással

7.5. Konklúziók

8. A navigáció és a keresés integrálása

8.1. Kategóriák navigációhoz és szűkítéshez

8.2. Kategóriák a keresési eredmények csoportosításához

8.3. Kategóriák az eredmények rendezéséhez és szűréséhez

8.4. Találati eredmények rendezése tartalomjegyzékhez hasonló nézet szerint

8.5. A webes tartalom hierarchikus navigációjának hanyatlása

8.6. Facettás navigáció

8.6.1. A Flamenco facettás navigációs interfész

8.6.2. Facettás navigáció az eBay Express-szen

8.7. Navigáció közösségi címkéken és közösségi könyvjelzőkön keresztül

8.8. Klaszterizálás a keresőinterfészekben

8.8.1. Klaszterezés dokumentumok közötti hasonlóság alapján

8.8.2. Klaszterezés megosztott kifejezés alapján

8.8.3. Klaszterezés a weben

8.9. Klaszterek vs. kategóriák a keresőinterfészekben

=====
=====

1. Keresőinterfészek tervezése

1.1. Legyen egyszerű az interfész

A keresőinterfész feladata, hogy segítse a felhasználót:

- információk szükségletének kifejezésében,
- a keresőkifejezés megformázásában,
- a keresési eredmények megértésében,
- az információkeresési folyamat nyomonkövetésében

A tipikus keresőinterfész mindazonáltal egy egyszéles interfész, írd-be-a-kulcsszavakat, nézd-át-az-eredményt-vertikális-listában sémára épül. Miért ilyen egyszerű a standard keresőinterfész?

- A keresés egy eszköz valamilyen cél elérése érdekében, mintsem cél önmagában. Teleologikus, ahogyan a filozófus mondaná.[M.M.] A keresés valamilyen átfogóbb feladat elvégzésébe épül be. A felhasználó nem szereti, ha megszakítják mindenféle kérdésekkel a feladat elvégzésének folyamatát.
- A keresés szellemileg intenzív feladat. Tehát minél kevesebb tényezővel kell a felhasználónak foglalkoznia, annál használhatóbb az interfész.
- Mivel szinte mindenki, aki a webet használja keres is, ezért az interfésznek érthetőnek és tetszetősnek kell lennie széles felhasználói spektrum számára, óriási variabilitású információk szükségletek kielégítésére alkalmazva.

[M.M.] Hearst továbbmegy, azt állítja, hogy még a fent említett végletekig egyszerűsített interfész sem minden felhasználó számára teljesen magától értetődő. Kutatások (Aula és Käki (2005); Hargittai (2004)) rámutatnak, hogy a találati lista egyszerűsítése bizonyos esetekben (időseknél) jelentősen csökkenti a hibák számát, illetve hogy néhányan még az egyszerű kulcsszó meghatározáson alapuló keresés alapjait sem értik meg. A Maszeker esetében a keresést képzett szakértők fogják végezni, tehát ezzel a kérdéssel nem kell foglalkoznunk.

1.2. Korszakváltás a keresőinterfészek tervezésében

[M.M.] A web megjelenése előtt az információkeresés egy szűk demográfiai csoport érdeklődésére tartott számot, azonban a web megjelenése fordulatot hozott ezen a területen is. Hearst ebben az alfejezetben áttekinti a web előtti keresést. A Maszeker szempontjából ez az áttekintés nem lényeges.

1.3. A keresőinterfész tervezésének folyamata

A felhasználói interfész egy fontos minősége a **használhatósága**, egy kifejezés, amely az interfész azon tulajdonságaira utal, amelyek meghatározzák mennyire egyszerű annak használata. A használhatóság öt összetevője Schneiderman és Plaisant (2004) illetve Nielsen (2003b) alapján:

- **Tanulhatóság:** Mennyire könnyen tudnak a felhasználók az interfésszel történő első találkozáskor alapvető feladatokat végrehajtani?
- **Hatékonyág:** Milyen gyorsan tudják a felhasználók elvégezni feladataikat, ha már megtanulták az interfész használatát?
- **Memorabilitás:** Milyen gyorsan tud a felhasználó jártassága visszatérni egy kihagyott periódus után?
- **Hibák:** Hány hibát követnek el a felhasználók, milyen súlyosak ezek a hibák és milyen könnyen tudják javítani ezeket a hibákat?
- **Megelégedés:** Mennyire kellemes és kielégítő az interfész használata?

A felhasználóközpontú (user-centered design) tervezés megvalósításához elengedhetetlen az interfész tervezés és a használhatóság területén szerzett jártasság, alapos ismeret, ezek holisztikus szemlélete, a HCI (Human-Computer Interaction).

Ebben a tervezési folyamatban először egy szükséglet kiértékelés történik meg, amelyben a tervező megvizsgálja a felhasználókat, igényeiket, céljaikat és az azok elvégzéséhez szükséges feladataikat. A második fázis a feladat elemzése, amikor a tervező meghatározza azokat a **lépéseket**, amelyeket a felhasználónak a **feladat** elvégzéséhez meg kell tennie, eldönti, hogy azok mely **cél** elérését támogatják és **scenáriókat** készít, amelyek szemléltetik a feladatok felhasználók által történő végrehajtását. A felhasználó céljainak és feladatainak meghatározása után a tervezés egy tervezés-tesztelés-újratervezés ciklusban folyik, amelynek során prototípust kell készíteni, felhasználói reakciókat nyerni és újraértékelni a tervezést a reakciókra építve. Ezt lehetséges, hogy néhányszor meg kell ismételní a kielégítő eredmény elérése előtt. Itt érdemes alkalmazni a „discount” tesztelést. Amikor ezen a terv már jól szerepel, akkor kell alaposabb finomításokat végrehajtani különböző variánsokat és kiválasztani közülük a leginkább megfelelőt. Ezzel a folyamattal nagyon alaposan foglalkozni kell [M.M.].

1.4. Tervezési irányelvek keresőinterfészekhez

Sok kutató és gyakorló tervező a HCI mezőben nagyon sok különböző irányelvre tett már javaslatot. Néhányan kiemelt figyelmet szenteltek a keresőinterfészeknek. Egy meghatározó dolgozat (Schneiderman et al. (1997)) nyolc tervezési kívánalmat határoz meg a keresőinterfészek tervezéséhez általában:

- Informatív visszajelzés
- Felhasználói irányítás támogatása
- Rövidtávú memória terhelésének csökkentése
- Shortcutok biztosítása képzett felhasználóknak
- Hibák csökkentése; egyszerű hibakezelés lehetővé tétele
- Törekvés a konzisztenciára
- Cselekvések egyszerű megfordításának lehetővé tétele

- Kompakt tervezés

Ezen irányelvek követése mindazonáltal nem olyan egyszerű, mert nem explicitek (ahogy az irányelvek általában) és nem mindig zárják ki az ellentmondást egymásnak. Hearst ehhez a listához hozzáteszi még Nielsen (1993) „beszélj a felhasználó nyelvét” irányelvét. Mindazonáltal az irányelvekkel is alaposabban fog még foglalkozni a könyv[M.M.].

1.5. Hatékony és informatív visszajelzés

Alapvető fontosságú, hogy a felhasználó mindig kapjon visszajelzést a rendszer állapotáról és a vonatkozó interakciókról. Ismerős példa a homokóra.

1.5.1. Keresési eredmények azonnali megmutatása

Számos tanulmány megmutatta, hogy nagyon fontos a felhasználó számára a keresési eredmény azonnali megmutatása. Ez segíti a keresőt, eldöntheti, hogy jó nyomon jár-e vagy sem. Sok kísérleti rendszer elköveti azt a hibát, hogy túl sok segítő információt tár a felhasználó elé, mielőtt a keresési eredményt megmutatná. Információmegjelenítő rendszerek, amelyek dokumentumokat pontokként mutatnak meg súlyos használhatósági gondokkal küzdenek, mert a kereső nem látja a szöveg címét és a szurrogátumot.

1.5.2. Informatív szurrogátumok megmutatása; Kulcssavak kiemelése

A legtöbb keresési eredmény vertikálisan jelenik meg és különböző információkat tartalmaz a dokumentumról (cím, url, szöveges összefoglaló – ez a szurrogátum). A dokumentum összefoglalója tipikusan néhány soros szöveg a dokumentumból kiemelve.

Fontos visszajelzés a keresőkifejezésnek és kontextusának megmutatása a szurrogátumban, ezáltal összekapcsolódik a dokumentum és a keresés. A keresőkifejezések távolságának megmutatása is nagyon hasznos a keresés hatékonyságának növelésében. Tehát ahol lehetséges a keresőkifejezésnek meg kell jelennie a szurrogátumban. De például ha minden kulcsszó megjelenik a dokumentum címében akkor nem szabad megjeleníteni azokat az összefoglalóban is, mert az így egyéb használható információt jelenít meg.

A kulcsszavak **vizuális kiemelése** szintén nagyon fontos tulajdonsága a keresőinterfészeknek. Ez azt jelenti, hogy a kulcsszavak szövegének háttérszínét megváltoztatjuk (eye-catching), ezáltal kiemeljük a szövegből. Ez segíti a keresőt az őt érdeklő szövegrészek megtalálásában, megmutatja, hogy az egyes kifejezések milyen közel vannak egymáshoz. **Mindazonáltal fontos, hogy ne túl sok kifejezést emeljünk ki, mert ekkor a kiemelés pozitív hatásai elveszhetnek.**

Meg kell találni az egyensúlyt a hosszú informatív összefoglalók és az egyes találatok megjelenítésére szánt képernyőméret között.. A kutatások eredményei nem egyértelműek abból a szempontból, hogy az összes keresőkifejezést tartalmazó mondat releváns részét meg kell-e jeleníteni, vagy koherens szövegrészeket és kevesebb kulcsszót. **Mindenesetre az biztos, hogy különböző típusú keresések különböző hosszúságú találati összefoglalókat kívánnak meg.**

1.5.3. Eredmények különböző szempontok szerinti rendezése

A visszajelzés egy másik hatékony formája, ha lehetővé teszi a felület, hogy a felhasználó különböző szempontok/dimenziók szerint rangsorolja a találatokat. Ezáltal a felhasználó érvényesíthet rangsorolási szempontokat és a változásokat azonnal láthatja. A találatok csoportosítása szintén hasznos lehet.

1.5.4. Keresési kifejezések ajánlása

Hasznos lehet a keresés elindítása után automatikusan generált keresési javaslatok és finomítások ajánlása a felhasználónak. Ezek lehetnek nyelvtani hibák javításai, vagy kapcsolódó, alternatív kifejezések is. Kifejezés kiterjesztése (term expansion): a használhatósági tanulmányok összességében pozitívan nyilatkoznak róla, erősen javasolt alkalmazni. Ennek lényege, hogy ahogy a felhasználó üti be a kulcsszó első betűit, a rendszer automatikusan generál egy listát, amit különböző szempontok szerint állít össze. Hearst egy repülőjegyek árusításával foglalkozó oldalon mutatja be: a reptér első néhány betűjének vagy azonosítójának beírása után a rendszer a legközelebbi találatot mint lehetőséget adja meg. Pl: elkezdi beírni, hogy s , akkor a rendszer az ország vagy reptér első betűje alapján leválogatja a következő listát: Palma Mallorca Spanyolországban, Suvarnabhumi Bangkokban, és az SFO San Fransiscoban, Kaliforniában. Egy betű, pl „f” hozzáadása első helyre teszi az SFO-t, de olyan, első betűleütés után kevésbé releváns találatokat is megjelenít, mint Sfax El Maou Tunéziában. Itt beszél még Hearst a „sliding trail” eszközről, amely lenyílik és különböző szemantikai szempont szerint rendezett találatokat mutat. Ez a mindennapi felhasználó szempontjából zavaró lehet, hiszen nem feltétlenül érti azokat a szemantikai szempontokat, de egy szakértő felhasználó biztosan haszná tudja venni egy ilyen típusú segítségnek[M.M.]. Hearst azt mondja még, hogy a kifejezések eltávolításának lehetősége nem szerencsés, mert a specifikumok megszüntetése után a kérés eredményhalmaza nagyobb lesz, nem pedig kisebb.

1.5.5. A relevancia indikátorok óvatós használata

Kutatások azt mutatják, hogy a relevancia szerinti rendezés nem javasolt weboldalak találati listáinak megjelenítésekor. Hearst azt mondja, hogy „a relevancia szerinti értékelés jelentése áttetsző a felhasználók számára”, illetve a találatok vertikális listázása és az egyes dokumentumoknak a listában elfoglalt pozíciója elegendő az adott dokumentum relevanciájának értékelésére. Innovatív megjelenítési rendszerek azonban vannak, amelyek grafikusán mutatják a kulcsszavak eloszlását a dokumentumokban. Ezeket elsősorban szövegelemzésre használják, szóval ezeknek mindenképp utána kell nézni[M.M.].

1.5.6. Azonnali válasz támogatása

Fontos, hogy a rendszer, amennyire csak lehet rezponzív legyen. A találati eredmények rezponzív és interaktív megjelenítése és a gyors válaszidő lehetővé teszi a felhasználó számára, hogy kérdéseit gyorsan újrafogalmazza, vagy több kérdést is fel tudjon tenni egyszerre. Ezekre tehát mindenképpen törekedni kell. Amennyiben az adott feladat elvégzése több időt vesz igénybe, ezt a felhasználó tudomására kell hozni (grafikus animáció).

1.6. Egyensúly a felhasználói irányítás és az automatizált folyamatok között

Hearst két fontos szempontot említ, amelyeket egy interfész tervezése során figyelembe kell venni: az egyensúly fenntartását a felhasználói irányítás és a rendszerfolyamatok között. Mit is jelent ez pontosabban? Hearst a digitális fényképezőgép példáját hozza: az előre definiált set-ekkel nagyon jó képeket lehet gyorsan készíteni, azonban mikor a felhasználó meg akarja változtatni a beállításokat (nagyon kevés fényhez pl.), akkor nehézségbe ütközhet a feladat végrehajtása során. **Ezért szükséges a „tradeoff” az áttetsző rendszerirányítás és a transzparens felhasználói irányítás viszonyában. A keresőinterfészek esetében ez az eredmények rendezését és a kérések átalakításának lehetőségét érinti.**

1.6.1. Sorba rendezés a webes keresésben

Hearst ebben az alfejezetben jórészt a webes keresők rendezési sajátágaival foglalkozik, pl a Hotbot keresőmotorja által bevezetett konjunktív (AND alapú) kulcsszóelemzést, ami azt jelenti, hogy minden kulcsszónak benne kell lennie a találati lista összes dokumentumában. Ez sokkal átláthatóbb a felhasználó számára mint a statisztikai alapú keresés. A 90-es évek végén a Google javította ezt tovább úgy, hogy nagyobb súlyt adott azoknak a dokumentumoknak, amelyekben a kulcsszavak közel helyezkedtek el egymáshoz. A Google vezette be továbbá a népszerűség mércéjét PageRank nevű szolgáltatásával. Napjainkban azonban úgy tűnik az AND alapú konjunktív keresés háttérbe szorul, mivel a felhasználók egyre hosszabb kéréseket alkotnak és megjelentek a természetes alapú keresők, amelyek ezeket támogatják is. Ahogyan nő a kulcsszavak száma, úgy kell lazítani az AND szabályán is. Más esetekben fontos lehet a kronológiai szempontú rendezés is, elsősorban hírek, publikációk stb. kapcsán.

1.6.2. Kérések átalakítása

A másik fontos szempont az, hogy a felhasználó hogyan alakíthatja át kéréseit, milyen lehetőségei vannak és milyen átalakításokat végez a rendszer (pl vs. -> versus). Egy másik példa az áttetsző rendszerre a megállítószavak negligálása. Ezek olyan szavak, amelyek a leggyakrabban használtak közé tartoznak (closed-class): a, an, the, or stb. Például a 'to be or not to be' keresés nem hozott volna eredményt a web korai szakaszában. 1996-ban pl egyedül az AltaVista kereső tudott találatot adni a fenti kérésre, a többiek ignorálták a megállítószavakat.

A szótövesítéssel óvatosan kell bánni, hasznos lehet, de csak mértékkel alkalmazva.

A helyesírás ellenőrzésével is hasonlóképpen óvatosan kell bánni, nem kell feltétlenül ráerőltetni a rendszer által helyesnek vélt formát, de azt sem szabad megengedni, hogy nyilvánvalóan helytelen kifejezéseket keressen a felhasználó.

1.7. Rövidtávú memória terhelésének csökkentése

A felhasználónak mindig releváns információt kell mutatni, mintsem megkívánni tőle, hogy emlékezzen rá vagy észben tartsa.

1.7.1. Keresési javaslat a keresőmezőben

Manapság elterjedt minta lett, hogy ne az üres keresőmezőt kapja a felhasználó, hanem valamilyen informatív szöveget. Például megkülönböztethető a teljes kollekción vagy a találati halmazon végzett

keresés. Mindazonáltal szerencsésebb egy rádiógombbal lehetővé tenni a választást, hiszen a felhasználó nem mindig veszi észre a lehetőséget.

1.7.2. Előzmények támogatása

A kutatás azt mutatja, hogy a felhasználók nagyon gyakran fogalmazzák újra korábban feltett kérdéseiket, ezért hasznos egy előzménylista elérhetővé tétele, ami egyrészt felgyorsítja, másrészt áttekinthetővé teszi ezt az interakciót. Ezt a webböngészők már alkalmazzák. Hasznos lehet a Maszeker esetében is, bár kérdés, hogy mekkora eséllyel fognak a felhasználók ugyanolyan kereséseket végrehajtani...

A weboldalak ezt a „breadcumb” nevű alkalmatossággal integrálják navigációjukba, ami nyomon követi a felhasználó útvonalát az egyre mélyebb információs szinteken.

1.7.3. Navigáció és keresés integrálása

Az emberi emlékezet egy jól megalapozott tulajdonsága, hogy egyszerűbb felismerni egy szót vagy nevet, mint emlékezni rá. A böngészhető információs struktúrák áttekintést adnak a felhasználónak a lehetőségekről, lehetővé teszik, hogy érdeklődésének megfelelően lépjen előre, szűkítse a halmazt. Hogy teljesen hatékony legyen, a navigációs interfésznek és a keresésnek összefésülhetőnek kell lennie egy egységes információs struktúrába.

A keresőinterfészek esetében a **kategóriarendszerek** azok az eszközök, amelyek segítségével a felhasználó a tartalomban navigál. A kategóriarendszer értelmes címkék halmaza, amelyek szervezése visszatükrözi a releváns tárgyterület fogalmi struktúráját.

Használata a következő esetekben ajánlott:

- dokumentumhalmaz részhalmazának kijelölése, ezáltal a találatok szűkítése
- elemek csoportosítása, ezáltal részhalmazok létrehozása
- használhatóak továbbá rendezésre (ordering and sorting) is

Struktúrájuk lapos, hierarchikus vagy facettás lehet.

A legjobb kombinációnak Hearst a hierarchikus facettás metaadatolást javasolja. A felhasználók így több szempont szerint böngészhetik az információs teret. Ha a facetták szerinti rendezés túl nagy, akkor az elemeket / facettákat hierarchiába lehet szervezni, ami lehetővé teszi, hogy egy elem más kontextusban is megjelenjen.

Tanulmányok azt mutatják, hogy nagyon jó interfészként használható (valamennyire) homogén információtartalmak navigációjához. A weben mindenesetre nagyon sokan használják már, mert nagyon jól működik. Ez a Maszekerbe is kelleni fog.

Nagyon jó példa erre a zvents.com weboldal, ahol különböző szempontok szerint lehet eseményeket keresni. Kifinomult használhatóságú oldal.

1.8. Kombinációk lehetőségének biztosítása

Arra vonatkozik, hogy gyakorlott felhasználók számára rövidítéseket kell létrehozni, amikkel bizonyos feladatokat gyorsabban tudnak elvégezni. Klasszikus példája a billentyűkombináció(CTRL+C, CTRL+V). A rövidítéseknek egyetlen hátrányuk van, amire figyelni kell: meg kell őket tanulni.

Egy alternatíva lehet erre a „mélylinkelés”, ami a webes keresésben már megjelent (a találati domain leglátogatottabb oldalainak a listája megjelenik a találat alatt, ezáltal a felhasználó egyből egy szinttel tovább lát).

A rövidítés másik formája, hogy a keresőmotor megpróbálja kitalálni, hogy milyen információs szükséglet vezérelte a keresést. Például a Google a „rentals seattle” kifejezésre már eleve olyan találati listát ad vissza, ahol a felhasználó tovább szűkítheti az ingatlanokra vonatkozó igényeit.

1.9 Hibák csökkentése

1.9.1. Tartózkodjunk az üres találati listáktól

Nem szerencsés a felhasználó számára üres találati listát adni egy kérésre. A helyesírás ellenőrzése és a kifejezés kiterjesztése ezen segíthet. A fent említett hierarchikusan facettázott rendszerek esetében előfordulhat ilyen (recept oldal esetében a desszert, alacsony zsírtartalmú és sajt facetták együttes kijelölése nagy valószínűséggel üres lesz....), de azt mindig a felhasználó tudomására kell hozni, hogy adott kombinációra nincs találat(keresőkifejezés előnézeti képe).

1.9.2. A szótárprobléma megoldása

A hiba egy másik forrása lehet ha a felhasználó nem ismeri az interfész szavait. Ez azt jelenti, hogy nem egyértelműek a navigációs jelek, a menük nevei vagy maga a keresés (keresőkifejezések). A kulcsszó illesztéssel történő keresés általános problémája a „nyelvi produktivásban” vagy másként a „szótárproblémában” rejlik. Ez azt jelenti, hogy különböző kifejezések nagyon sokféleképpen fejezhető ki. Az egyetértés mértékével kapcsolatos vizsgálatok azt mutatták, hogy egy speciális tárgyterület esetében egy kifejezés 15 különböző formájának használatakor 60-80%-os az egyetértési ráta. Egy kifejezés használatakor ez 11%... A kifejezések kiterjesztésének lehetősége jó megoldás lehet.

A tény, hogy különböző emberek különbözőképpen fejeznek ki hasonló fogalmakat mélyreható következményekkel bír információs rendszerek tervezésekor. **Azt is megmutatja, hogy az interfész elemek címkézését nagyon alaposan meg kell fontolni.** Segédeszköz a kártyarendezés (card sorting), aminek célja a legmegbízhatóbb és kiszámíthatóbb kategóriák és címkék kiválasztása egy adott információs struktúrához.

1.10. Figyelmet kell szentelni a részleteknek

Hearst nagyon érdekes példákat mutat be olyan esetekről, ahol az interfész nagyon kis változásai a felhasználók viselkedését nagy mértékben befolyásolták. Például egy kísérletben hosszabb keresőmezőt kaptak a felhasználók és megfigyelték, hogy megnövekedett a keresőkifejezésük hossza is. A másik érdekes példa a Google helyesírás javítójának integrálása a találati oldalak szerkezetébe.

1.11. Az esztétika fontossága a tervezésben

A szín, a tipográfia, a layout megtervezése nagyon fontos. Ez egyértelmű. Nem pusztán a szoftver elfogadását könnyíti meg, de egy kísérlet szerint a feladatok végrehajtásának idejét is a felére tudja csökkenteni egy grafikai szempontból igényesen megtervezett interfész. Fontos tervezési szempontok: csoportosítás, sűrűség, rendezés és méret. Ehhez a részhez meg kell nézni Jennifer Tidwell: User Interface Design című könyvéből a vonatkozó részeket.

1.12. Összefoglalás

Irányelvek:

- hatékony és informatív visszajelzés
- egyensúlyban a felhasználó irányítása és a rendszer automatizmusa
- rövidtávú memória terhelésének csökkentése
- rövidítések biztosítása
- hibák csökkentése, kizárása
- apró részletek fontossága
- esztétikum fontossága

2. A keresőinterfész kiértékelése

Az első fejezet a következő kritériumok szerint osztályozta az interfészeket: tanulhatóság, hatékonyság, memorabilitás, hibák csökkentése, felhasználói elégedettség. Azonban az interfészek kiértékelése más szempontok szerint történik: hatásosság, hatékonyság, megelégedettség.

- hatásosság: pontosság és teljesség, amivel a felhasználók elérik meghatározott céljaikat
- hatékonyság: erőforrások szükségessége a pontosság és teljesség eléréséhez
- megelégedettség: mentesség a kellemetlenségektől és pozitív attitűd a szoftver további használatát illetően

2.1. Információvisszakeresés standard kiértékelése

A hagyományos visszakereső rendszerek legfontosabb kiértékelési szempontja a találatok relevanciája volt. Egyáltalán nem vették figyelembe a felhasználót, pusztán az algoritmus hatékonysága számított a kiértékelésnél. Ennek legprominensebb megnyilvánulása a TREC (Text Retrieval Conference) versenye, amelyben különböző intézmények algoritmusai versenyeznek egymással, bármiféle felhasználói interakció nélkül.

A kiértékelő algoritmus kiértékelésének szempontjai: Pontosság, előhívás, F-mérték és Mean Average Precision (MAP). A pontosság úgy határozható meg, mint az a releváns találatok és a találatok hányadosa. Az előhívás a megtalált releváns dokumentumok és az ismert releváns dokumentumok hányadosa. Az F-mérték a kettő közötti egyensúlyt jelenti. Úgy határozható meg mint a két érték súlyozott középértéke = $(2 * P * R) / (P + R)$. A legnagyobb probléma – több kritika mellett – az, hogy nem számol a felhasználóval, a felhasználói felülettel, egyáltalán nem vonja be ezeket keresési folyamatba. Ezek bevonása egy sokkal árnyaltabb értékelést tesz lehetővé. Pi Pickens et al. (2008) megkülönböztet azokat a dokumentumokat, amiket a kereső találatként visszaadott, azokat amelyeket a felhasználó megnézett és azokat amelyeket a felhasználó relevánsnak minősített –

lépcsőzetes relevanciaértékelés. Fontos fogalom még az azonnali pontosság, ami azt jelenti, hogy a felhasználók általában csak egy-két dokumentumot néznek meg a találati listából. Azon kérések száma, amik legalább egy releváns dokumentumot eredményeztek k számú dokumentum áttekintése után. Például 80%-os azonnali pontosság azt jelenti, hogy a felhasználó a kérések 80%-ában legalább egy releváns dokumentumot talált a találati lista első két dokumentuma közül.

2.2. Informális használhatósági tesztelés

Amikor grafikai terveket mutatunk meg a leendő felhasználóknak és rögzítjük a reakcióikat, akkor informális tesztelést végzünk. Ez hasznos lehet a nagyobb hibák kiszűrése céljából és a terv pozitív aspektusainak megerősítésében. Egy ilyen teszt megerősíthet bizonyos tervezési mintákat és elvetésre ítélni másokat. Ez az a tesztelés, ami az első fejezetben „discount usability testing” néven jelent meg.

A tervezés korai fázisában javasolt ún. **low-fidelity prototípusok bemutatása** a leendő felhasználóknak, gyakran papír vázlat formájában. Ezeket gyorsan meg lehet változtatni és lehetővé teszik, hogy a lényeges tervezési elemekre fókuszáljon a felhasználó, ne pedig a grafikai csilivilire, illetve azt, hogy több konkurens verziót is teszteljen a tervező.

Ezek kiértékelése nagyon hatékony útmutatóul szolgálhat a továbbiakban. Általában 2-3 kiértékelő van jelen egy résztvevő megvizsgálja a rendszert. Egyikük elmagyarázza a terv célját és a feladatokat, amiket a felhasználónak végre kell hajtania, míg a másikuk kicseréli a papír prototípusokat az aktuális feladat végrehajtásának megfelelően. A harmadik egy kicsit távolabbról figyeli az eseményeket, rögzíti a résztvevő megjegyzéseit, benyomásait, zavarait stb. A résztvevőket bátorítani kell arra, hogy hangosan gondolkodjanak.

Miután a low-fidelity terv kiértékelése megtörtént egy high-fidelity terv készül a kiértékelésnek megfelelően, amibe már egy kevés interaktivitás is kerülhet. Ezeket újra ki kell értékelni, először csak néhány résztvevővel. Ez szintén iteratív folyamat. Sokan sokféleképpen nyilatkoznak arról, hogy hány résztvevővel tekinthető eredményesnek és megbízhatónak a kiértékelés, Nielsen szerint 5 felhasználó már kiszűri a hibák 85%-át. Mások ezt megcáfolták, illetve azt írták le (Lindgaard és Chattratichart, 2007), hogy nem az a fontos hány személlyel történik a kiértékelés, hanem az, hogy a különböző feladatok közül mennyit végeznek el(!).

A „discount” tesztelés egy másik formája a **heurisztikus kiértékelés**, amely azon a feltételezésen alapul, hogy egy használhatósági szakértő képes kiszűrni és felismerni a potenciális hibákat. Ebben a változatban néhány szakértő tanulmányozza a tervet és irányelvnek megfelelően értékeli azt.

A kettő kombinációja kiválóan működhet a tervezés korai fázisában.

A harmadik informális használhatósági teszt az ún. **„field study”**, amelyben a szakértők a munkavégzés helyszínén rögzítik és megfigyelik a felhasználók feladatvégzését, akár munkahelyen, akár otthon. Feltételezhető, hogy egy jól ismert környezetben a felhasználó sokkal realisztikusabban viselkedik.

Fontos, hogy a tervező el tudjon vonatkoztatni az általa favorizált terv erőltetésétől és észrevegye és megfigyelje azokat az apró jeleket, amelyek a legjobb megoldás irányába mutatnak. Gyakori hiba alulértékelni a negatív és felülértékelni a pozitív visszajelzéseket.

2.3. Formális tanulmányok és irányított kísérletek

Segítik a tárgyterület megértését, illetve azt, hogy a felhasználók hogyan használnak interfészeket ezen a területen belül, milyen tervezési koncepciók működőképeseek milyen körülmények között és miért.

Ebben a megközelítésben bizonyos határok között végezzük a felület kiértékelését, ahogyan egy gyógyszer klinikai tesztelése is jól meghatározott keretek között zajlik.

2.3.1. A formális vizsgálatokban alkalmazott technikák

A klasszikus formális használhatósági vizsgálatot használhatósági laborban végezzük, amelyben a megfigyelők detektívtükör mögött vannak. Mindazonáltal egy csendes szoba asztallal, székekkel és egy számítógéppel is megteszi. Az eseményt video- vagy hangrögzítőre kell felvenni a későbbi elemzés céljából.

A résztvevők adatainak védelme kiemelt fontosságú! A résztvevők a vizsgálatot bármikor különösebb ok nélkül félbeszakíthatják.

Bevett gyakorlat bemelegítő feladatokat adni a résztvevőknek, amelyek ráhangolhatják őket a rendszerre. Kezdetben ezeket közvetlen munkatársakkal is el lehet tesztelni, mielőtt a résztvevőknek odaadnánk.

A tesztelés végrehajtásához nem kell feltétlenül a résztvevővel azonos helyen tartózkodni, léteznek ún. távoli használhatósági tesztek, amelyek a terepen végzett tesztelés hatékonyságát elérik. Segítségével nagyobb diverzitású, könnyebben szervezhető felhasználói tesztcsoportot lehet tesztelni. Hátránya a közvetlenség (nonverbális kommunikáció) bizonyos fokú hiánya, illetve az, hogy a kísérletet végzők és a résztvevők közötti információáramlás nehézkes lehet. Nem minden esetben oldható meg továbbá az, hogy a szoftvert a résztvevő a saját gépén tesztelje (a szoftver futtatási környezetének speciális szükségletei miatt).

2.3.2. A kísérleti feltételek prezentálását ki kell egyensúlyozni

A felhasználó tesztek egyik problémája annak a kiküszöbölése, hogy a különböző kísérleti interfészverziókat milyen sorrendben prezentálják az alanyoknak. A felhasználó több interfésszel többféle módon ismerkedik a kísérlet során. Ezért nagyon fontos, hogy az egyes feladatok elvégzésének időtartamát az egyes interfészverzióknak megfelelően rögzíteni kell.

Ezeket a kísérleteket meg kell tervezni. Szerencsére rendelkezésre állnak módszerek, amelyek segítségével ez kivitelezhető. Egy ilyen, nagyon gyakran és sikeresen alkalmazott módszer a Latin Négyzet. Ebben egy négyzetet kell készíteni: pl 3 sor x 3 oszlop. Az első négyzet első sora 1,2,3 volna, a második 2,3,1, a harmadik pedig 3,2,1. Ez, jóllehet nem tartalmazza az összes lehetséges variációt, már elegendő alap lehet a prezentációs hatások felméréséhez.

Hearst ezután egy kísérletet ír le, amelyben alkalmazzák ezt a módszert. Igen tanulságos és meggyőző, mindenképp kellene alkalmazni nekünk is.

2.3.3. Résztvevők gyűjtése

A legjobb természetesen a jól meghatározott leendő felhasználói csoporttal végrehajtani a tesztelést (nekünk szerencsére ez adott), de kollégákkal vagy hallgatókkal is lehet tesztelni a rendszereket.

Vannak még olyan érdekes kezdeményezések (crowsourcing) is, mint az Amazon Mturk, ahol rövid (1 perc – 10 perc) feladatokat kell végrehajtani egy csekély (\$0.01-\$0.5) összegért cserébe.

2.3.4. A felhasználók preferenciáinak mérése

A különböző preferenciákat a Likert skálával mérhetjük. A Likert (alighanem) a Like és Alert kifejezések összevonása. A felhasználó kiválaszt egy pontot egy minőségi tartomány két végpontja között (tetszett, nem tetszett, egyetért, nem ért egyet). A legtöbb Likertnek 5,7 vagy 9 értéke van. Ezt a kérdéshez igazodva célszerű meghatározni. Pl. fontosabb kérdések esetében 9, kevésbé fontosak esetében 5.

Egy tanulmány szerint jobb standard kérdőíveket használni mint ad hoc jellegűeket. Az egyik ilyen kérdőív a [Questionnaire for User Interaction Satisfaction \(QUIS\)](#), aminek a student liszenszdíja \$50.

Az esztétikai benyomások nagy fontossággal bírnak a használhatóság és elfogadás szempontjából, azonban alapvetően nem befolyásolják egy feladat elvégzésének gyorsaságát. Ben-Bassat et al.(2006) azt találták, hogy a szubjektív benyomások számítanak egy kérdőív alapú kiértékelésnél, azonban egyáltalán nem egy cselekvés alapúnál(ahol a felhasználót valami ösztönzi a feladat elvégzésére).

Néhány esetben célszerű teljesen elválasztani a használhatóságot és a vizuális tervezést. Például egy keresőmotor rendezőalgoritmus vizsgálataánál A és B találati eredménynek ugyanolyan semlegesnek kell lennie.

2.4. Longitudinális kutatások

Egy keresőinterfész ökológiai érvényességét (ecological validity – tervezési minták használata és értéke) hosszútávú vizsgálatok igazolják. Ezek a vizsgálatok nagyobb időtartamot ölelnek át, ellentétben a fentebb tárgyaltakkal. Itt már lehetőség van olyan finomhangolásokra, amelyek az interfész használata során kialakult mintákat is figyelembe veszik. Hearst két példán mutatja be ezek hasznosságát.

2.5. A keresőmotor logfájljainak elemzése

A legtöbb keresőmotor rögzít valamilyen információt a felhasználóról. Ezek nagyon hasznosak lehetnek különböző elemzések elvégzésekor. Ezek általában a következők: kulcsszavak, keresés időpontja, felhasználó IP címe. Ezekből – ha elegendő áll rendelkezésre – fontos következtetéseket vonhatunk le mind a keresőalgoritmus mind pedig az interfész hatékonyságával kapcsolatban.

2.5.1. Session határok azonosítása

Az információkeresés gyakran több lépésből áll. A lépések egymásutánisága sokat elárulhat a felhasználó eredeti céljáról vagy az adott feladat elvégzésének hatékonyságáról. Ehhez kiváló segítséget nyújt a logfájlok elemzése. A felhasználó feladatának elvégzését úgy rögzíthetjük, hogy megnézzük az első cselekvés idejét, meghatározunk egy időtartamot (5-60perc), ami ha eltelik, akkor a felhasználó új feladatba kezd. A feladatokat ezek alapján csoportosítjuk.

2.5.2. A felhasználó identitásával kapcsolatos kérdések

Egy logfájlban a felhasználót IP címe azonosítja, tehát elvileg anoním marad. Azonban ez nem mindig van így, illetve bizonyos ismertetőjegyek alapján így is be lehet azonosítani az egyes felhasználókat.

Mindent meg kell tenni annak érdekében, hogy a logfájlok és ezáltal a felhasználók semmilyen adatai ne kerüljenek illetéktelen kezekbe. Vannak jó anonimizáló algoritmusok.

2.6. Átfogó, logfájl alapú használhatósági tesztelés (vödör(?) teszt)

Nagy forgalmú weboldalak esetében lehetőség van arra, hogy egy tervet a felhasználók nagyobb csoportján teszteljünk azáltal, hogy bizonyos százalékuknak az új verziót mutatjuk meg, a többieknek pedig a régit (a Youtube alkalmazza ezt mostanában pl). Ezáltal lehetőség van nagy számú felhasználó által generált logfájlok elemzésére és nagyon gyorsan lefuttatható egy verzióteszt. Ez a fajta tesztelés persze nélküli a formális tesztelés néhány előnyét (nincsenek meghatározott feladatok, nincs explicit visszajelzés), viszont kisebb funkcionális vagy grafikai változtatások tesztelésére kiválóan alkalmas. Ez a fajta tervezés kiválóan alkalmasnak bizonyult tervezési döntésekkel kapcsolatos problémák feloldására. Az Amazon például gyakran használja ezt és így lehetősége nyílik a gyors és használható innovációra. Ezeket a változtatásokat nagyobb forgalmú oldalak esetében fokozatosan kell végrehajtani (eBay: 30 nap alatt változtatta meg az oldal háttérszínét szürkéről fehérre...).

2.7. Speciális megfontolások keresőinterfész kiértékelésekor

2.7.1. Tartózkodni kell a kutatói elfogultságtól

Gyakori probléma a használhatósági tesztek kiértékelésénél a „minden szentnek maga felé hajlik a keze” effektus, ezért célszerű független elemzővel kiértékelteni a felületet. Azonban gyakran nem praktikus külső szakértőt bevonni, a kísérletet végzőnek ki kell tudnia küszöbölnie ezt az elfogultságot. Fontos például, hogy ne „szivárogtasson” információt az interfészről, illetve hogy egy **semleges nyelven** fogalmazza meg kéréseit / kérdéseit a felhasználó felé.

2.7.2. A résztvevők motiválása

A résztvevők motivációja vagy ösztönzése jótékony hatással van a kiértékelésre. Fontos azonban, hogy a felhasználónak ne véletlenszerű kereséseket kelljen végrehajtani, hanem a szakterületének megfelelőt. Ezért már az is egy motiváció, ha a meglévő információs szükséglet kielégítését el tudja végezni a felhasználó.

2.7.3. Számolni kell az individuális különbségekkel

Nielsen (1993) szerint a két legfontosabb tényező a használhatósági tesztben a felhasználó feladatai és a felhasználók individuális különbségei.

2.7.3.1. A résztvevők tudása a szakterületről és feladatokról

Hearst különböző tanulmányokra és kutatási eredményekre hivatkozva megállapítja, hogy minél jobban ismeri valaki azt az adott tárgyterületet, amelyen a keresést végre kell hajtani, annál jobb lesz a keresés hatékonysága (relevánsként mentett dokumentumok száma osztva a megtekintett dokumentumok számával). Nekünk ebből a szempontból szintén szerencsénk van.

Fontos az is, hogy a kiértékelés előtt meg kell tudni, hogy a felhasználók ismerik-e az adott keresőkérdésre a választ. Sokat torzíthat a teszten, ha valaki egyből a jó irányba indul el.

2.7.3.2. A résztvevők korábbi keresési tapasztalatai

Egy kutatásban a kezdő felhasználók (akiknek még nem volt tapasztalata az interfésszel) hamarabb elvesztek félrevezető kérdések labirintusában, próba-szerencse módszer szerint haladtak a cél felé, míg a tapasztaltabb felhasználók nem estek kétségbe, mert tudták, hogy a rendszerben előbb-utóbb megtalálják a keresett információt.

Más kutatások nem találtak összefüggést a keresés sikeressége és a tárgyterület ismeretének szintje között, jóllehet utóbbi elmélyülése magával hozta a keresési stratégiák megváltozását.

2.7.3.3. A résztvevők kognitív képességei

Számos tanulmány megállapítja, hogy a résztvevők kognitív képességeinek szintje összefüggésben van a keresési hatékonysággal.

2.7.4. Számolni kell a kérések és a feladatok különbözőségével

Nielsen (1993, p. 185) megjegyzi, hogy az alapszabály a feladatok tesztelésein az, hogy amennyire csak lehetséges reprezentatívak legyenek az élesben végzett használati esetek vonatkozásában. Javasolja továbbá a feladatok kis egységekre bontását.

2.7.4.1. Számolni kell a feladatok változatosságának hatásával

Az emberek különbözőképpen reagálnak bizonyos keresésekre. Ezek lehetnek emocionálisak is, vagy, ha a résztvevő az adott terület szakértője, magabiztosak is. Különböző kutatások szerint érdemes az elvégzendő kereséseket egy előzetes tesztnek alávetni és a kialakuló minták szerint csoportokba rendezve prezentálni a tesztalanyoknak. Így biztosítható a feladatok homogenitása.

2.7.4.2. A keresőkifejezésben megjelenő variabilitás kontrollálása

Egyes felhasználói teszteléseinél előre leírják, hogy a tesztalanyok milyen keresőkifejezéseket használhatnak.

Ez nem minden esetben szerencsés, hiszen meglehetősen valószerűtlen szituációkat eredményezhet, azonban olyan speciális interfészeknél mint a Maszeker célszerű egy jól meghatározott készletből gazdálkodva kereséseket végeztetni a felhasználókkal, hiszen ezáltal összemérhetővé válnak az egyes felhasználók által végzett keresési statisztikák és módszertanok. Nyilván, amikor a keresési eredmények használhatóságának kiértékelése a cél, kevésbé fontosak a konkrét kérdések. Amikor viszont például egy ajánlórendszer tesztelése történik akkor meglehetősen valószerűtlen.

2.7.4.3. Tartózkodni kell az előítéletektől a kérés és feladat megválasztásakor

A kísérletet végző könnyen becsempészheti előítéleteit a kísérletbe a nem megfelelően kiválasztott feladatok vagy keresések kijelölésével.

Összefoglalva: az előítéletek a következő okokból merülhetnek fel: az eredmények megjelenítésének nem megfelelő variálása, arra kérni a felhasználót, hogy egy feladatot megismételjen, a feladatok elvégzésének idejét befolyásolhatja, hogy bizonyos adatokat előtölt a rendszer, míg másokat nem.

2.7.5. A teszt kollekció karakterisztikájának figyelembevétele

Fontos, hogy a tesztelések során keresőalgorithmus és a keresőinterfész a célkorporusz minél nagyobb részét keresve tudjon működni. Nincs értelme pár száz elemből álló dokumentumhalmazon használhatósági tesztet végezni.

2.7.6. Számolni kell a feladatok végrehajtásának időtartamában megjelenő különbségekkel

Nagyon fontos metrika az interfészek kiértékelésénél az egyes feladatok elvégzésének ideje, erőfeszítés mértéke, illetve az elkövetett hibák száma és súlyossága.

A feladatok elvégzésének ideje azonban sokmindentől függhet, ráadásul bizonyos interfészek esetében pont az a cél, hogy egy feladat elvégzése sokáig tartson (a felhasználó többet tudjon meg egy adott témáról).

Käki és Aula (2008) rámutat, hogy nincs egy standard időegység egy feladat elvégzésének sebességére, ezért bevezetik a **minőségi keresési idő** fogalmát. Ez a válasz/perc arányt méri, a minőségi válaszok szempontjából. Ennek segítségével összehasonlíthatóvá válik A és B rendszer a releváns válasz/perc és az irreleváns válasz/perc értékek alapján.

Az időtartamok mérését még a hálózati kapcsolat minősége is befolyásolhatja, illetve néhány más technikai tényező.

2.7.7. Összehasonlítás egy (túl) erős alappal

Egy másik lehetséges csapdahelyzet az, amikor a keresőrendszer képességeit egy valószerűtlen rendszerével hasonlítják össze.

2.8. Konklúziók

A fejezet keresőinterfészek és interfészek kiértékelésének széles spektrumát írta le. Ezek a következőképpen foglalhatóak össze:

a standard információkeresés kiértékelése, ami a keresőmotorok találati hatékonysága alapján osztályozza a keresés, de nem számol a felhasználói interakcióval

a következő az informális és a formális keresőinterfész kiértékelések voltak, majd a hosszútávú kutatások, amelyek lehetővé teszik a kutató számára, hogy az interfészt hosszab időtartam használata után újraértékelje, és a nagy léptékű kutatások, amelyek létező, nagy méretű felhasználói csoporton hasonlítanak össze interfész verziókat

Felmerül a kérdés, hogy mindezek tudatában mégis hogyan kell nekikezdeni az interfész tervezésének? Egy lehetséges stratégia, hogy amikor új interfészt tervezünk, akkor a kiértékelés felől közelítsük az meg, mintegy visszafelé haladva.

Szintén fontos, hogy realiztikus feladatokat definiáljunk, amik összhangban vannak a felhasználói bázis igényeivel. Ezeket a feladatokat széleskörű tesztelésnek kell alávetnünk, amelyek során a résztvevőknek motiváltaknak kell maradniuk. Az informális kiértékelések segítségével gyorsan meghatározhatjuk, hogy mi az ami működik és mi az ami nem. Formálisabb kutatásokat akkor kezdünk el, amikor az informálisak már jó alapot biztosítanak ehhez.

3. Az információkeresés folyamatának modelljei

Ha jól működő keresőinterfészt akarunk tervezni, elkerülhetetlen, hogy megértsük az emberi információkeresés folyamatát, illetve azokat a stratégiákat, amelyeket az emberek ezen folyamat során alkalmaznak. A fejezet bemutatja a legelterjedtebb modelleket.

3.1. Az információkeresés standard modellje

Sokan azt feltételezik, hogy az információkeresés egy körkörös interakció, amely az információs szükséglet azonosításából, a keresőkérés megfogalmazásából a találatok vizsgálatából és ha szükséges a keresés finomításából áll. Ez a kör addig ismétlődik, amíg a felhasználó meg nem találja az általa keresett választ.

Ezt a modellt Sutcliffe és Eniss (1998) dolgozta ki alaposan, amit a következőképpen formuláltak:

- Probléma azonosítása
- Információs szükséglet megfogalmazása
- Kérés megformálása
- Eredmények kiértékelése

Shneiderman et al. (1997) a következő főbb lépéseket állítják meg:

- Kérés megformálása
- Cselekvés (kérés futtatása)
- Eredmények felülvizsgálata
- Pontosítás

Marchinini és White (2008) leírása az információkeresésről a következőkből áll:

- Az információ iránti igény felismerése
- Az igény kielégítésére vonatkozó cselekvés kihívásának elfogadása
- Probléma megfogalmazása
- Információs szükséglet kifejezése a keresőrendszerben
- Eredmények vizsgálata
- A problémának és kifejezésének újrafogalmazása
- Eredmények használata

Ezek jelentik a központi cselekvéseket egy információkereső feladatban.

A webes keresők ezek közül jórészt a kérés megfogalmazását, az eredmények vizsgálatát és bizonyos fokig a kérés újrafogalmazását teszik lehetővé.

Vannak olyan rendszerek, amelyek az értelemalkotást(sensemaking) is támogatják. Ezek a probléma megfogalmazását, az információ újraszervezését és az összegyűjtött információból történő új reprezentációk létrehozását is támogatják.

3.2. Az információkeresés kognitív modelljei

Norman (1988) meghatározó tanulmányából vezethető le, amely általában foglalkozik a feladat végrehajtásával. E szerint egy személy rendelkezik egy céllal, amit el akar érni. Ezután belső reprezentációs modelljét a világról (mentális modell) arra használja, hogy eldöntse, milyen feladatok végrehajtásával lehet azt a célt elérni, hogyan hathat ez a világra stb.

HCI-ben gyakran használt fogalomvá vált a mentális modell különböző rendszerek és interfészek tervezése kapcsán. Ez egy dinamikusan változó belső megjelenítése egy problémának, szituációnak vagy rendszernek, amely inputokat fogad a világból és különböző, az inputokra adott hatások predikciójával tér vissza.

Norman ezt kettéosztja a cselekvésre (doing) és az eredmény kiértékelésre (checking). A különbséget az elvárt és az elért között a cselekvés öbleként jelöli, míg annak meghatározását, hogy teljesültek-e a célok a kiértékelés öblének. Interfész tervezés esetén törekedni kell arra, hogy ezek az öblök minél kisebbek legyenek, mert annál használhatóbb lesz a rendszer. Ez a modell úgy is tekinthető, mint ami kognitív alátámasztást ad a standard modell számára.

3.3. A dinamikus (bogyószedő) modell

A standard modell szerint a felhasználó információs igénye statikus, nem változik a keresés folyamán és a siker a kérés hatékony finomításán múlik és csak az eredeti igénynek megfelelő dokumentumokat nyertük ki. Mindazonáltal a keresés folyamán a kereső egyre többet tud meg a keresés tárgyáról, és új alkérdéseket fogalmaz meg, miután bizonyos alkérdéseket megválaszoltak az eredmények. A standard információkeresés modelljét fenti okokból kifolyólag ezért többen is kihívták.

Bates (1989) a „bogyószedő” modellt javasolja, amelyben az információkeresés a bogyók szedéséhez hasonlít. Két fő pontban jellemezhető: 1. a kereső a keresés során feldolgozott információból tanul és ez folyamatosan megváltoztatja a kereső információs igényét és kéréseit. Az információ egy pontból több irányba is vezethet. 2. A kereső információs igényét általában nem elégíti ki egy végső dokumentumkészlet, hanem inkább választások sorozat és információs falatok, amelyeket útközben talált. Ez nagy kontrasztban áll az előző két verzióval. Nagyon sok megfigyelés támasztja azonban alá. Azt találták, hogy a keresések egy céllal összefüggésben újabb célokat jelöltek ki és az érték nem a végső dokumentumhalmaz megtalálása volt, hanem maga a keresési folyamat, amelynek során új ismeretekre tett szert a kereső.

3.4. Fokenkénti információkeresés

Kuhlthau (1991) foglalkozott először érdemben ezzel a témával. Szerinte megállapítható, hogy a keresők mind kéréseik mind tudásuk szempontjából fokozatokon haladnak keresztül, amelynek alátámasztására egy kiterjedt kísérletet is lebonyolított. A kísérletben jórészt hallgatóknak kellett könyvtári információs adatbázisokban keresve dolgozatot írniuk egy előre meghatározott és számukra kijelölt témában. Azt is kérte a kísérletvezető, hogy a résztvevők ne csak információkeresési szokásaikat rögzítsék a kísérlet során, hanem érzelmi állapotaikat is.

A következő, alapvetően közösnek mondható információkeresési folyamatokról és érzelmi mintákról számol be:

Kezdet: a feladat az információs szükséglet felismerése. A keresők általános háttértudással rendelkeznek. Ahogy a résztvevő rájön, hogy hiányos a tudása a bizonytalanság és a nyugtalanság

érzései felerősödnek. A gondolatok a feladat megértése körül forognak és próbálják a már meglévő tapasztalatra vonatkoztatni a problémát.

Kiválasztás: a feladat egy általános topik vagy megközelítés keresése. A gondolatok általánosak és indifferenciáltak és lehetőségek körül forognak. Gyakori az optimizmus érzése a kiválasztás után.

Felfedezés: A feladat információszerzés az általános topikról, ezáltal a megértés kiterjesztése. Gyakori az információs konfliktus a már meglévő tudással. Érzelmek: zavar, bizonytalanság, kétely.

Kifejezésre juttatás: Fordulópontot jelent a folyamatban, fókuszált perspektíva áll rendelkezésre a topikról, konfliktusok feloldásra kerülhetnek. Hipotézist igazoló keresések. Érzelmek változnak, bizonytalanság csökken, magabiztosság nő. Sajnálatos módon azonban a résztvevők felének nem sikerült elérnie egy fókuszált perspektívát a keresési folyamat során.

Gyűjtés: Ezen a szinten a legproduktívabb a keresés a résztvevő számára. A keresések azt a célt szolgálják, hogy meghatározzák, kiterjesszék vagy támogassák a fókuszot. Relevanciaítéletek egyre pontosabbakká válnak.

Prezentáció: Ebben a fázisban a keresés készen van. A kereső már vagy redundáns vagy indifferens információt talál csak. Felszabadultság és megelégedettség érzései töltik el a felhasználót.

3.5. Az információkeresés mint stratégiai folyamat

Néhányan stratégiai folyamatnak tekintik az információkeresést, ahogyan Marchionini (2000) megjegyzi: „a keresés analitikus és interaktív problémamegoldó stratégiák kölcsönhatásában jön létre”. Néhány esetben ezek szakértők által tudatosan megtervezett viselkedést jelentenek.

3.5.1. Stratégiák mint különböző taktikák szekvenciái

Bates (1979) szerint a kereső viselkedése különböző keresőtaktikák szekvenciájaként írható le. Ezek belső választások vagy cselekvések a figyelem aktuális fókuszában álló keresésnek megfelelően. Segítenek elérni valamilyen részcélt. Bates felsorolja a keresőtaktikákat, amelyeket négy fő kategóriába sorol:

1. Terminus taktika: szavak és kifejezések kérésen belüli igazítása. Kifejezés ajánlók(term suggestion) és teauruszok lehetnek segítségére a keresőnek.
2. Információs struktúra taktika: információ vagy linkstruktúrán keresztül történő mozgás a forrás felé. Idézettségi lánc. Kategórián belüli keresés vagy hyperlinkek követése.
3. Kulcsszó újrafogalmazás taktikája: kulcsszó hatókörének szűkítése, Boolean operátorok használata
4. Monitorozó taktika: a helyzet kifejlésének figyelemmel követése. Bates szerint az egyik legnehezebb feladat a keresési stratégiákkal kapcsolatban annak megállapítása, hogy mikor kell megállni. A monitorozás ezen segíthet.

Azonban ezt nagyon nehéz lehet pontosan meghatározni. O'Day és Jeffries (1993) meghatároztak néhány olyan kioldót (trigger), amelyek a keresőt egy másik stratégiára történő váltásra vehetik rá:

- egy lépés teljesítése és egy másik logikus lépés elkezdése egy tervben

- valamilyen szemléletformáló újdonsággal történő találkozás az információkeresés során, amely új szempontból mutatja be a problémát, rávilágít egy másik aspektusára
- előzetes elvárások változásával vagy sérülésével való találkozás ami további vizsgálódást követel
- előzetes megértés inkonzisztenciájával vagy hiányosságaival való szembesülés, amely további vizsgálódást követel

Fenti szerzőpáros megkísérelte a megállító feltételeket is meghatározni – azokat a körülményeket, amelyek alatt az emberek úgy döntöttek abbahagyják a keresést:

- nincs több kényszerítő trigger
- elegendő mennyiségű információ gyűlt össze
- specifikus gátló tényező jelent meg (pl kiderült, hogy a megcélzott piac túl kicsi ahhoz, hogy megérje kutatni)

Bates (1979) mindenestre megjegyzi, hogy néhány információs igény eléggé egyszerű ahhoz, hogy ne kelljen stratégiát alkalmazni(sőt, alighanem ez a gyakoribb).

3.5.2. Költségszerkezet elemzés és információ szerzés elmélet

A fent említett monitorozási stratégia kapcsán kerül szóba a költségek és a keresés folytatásából adódó nyereség súlyozása. Russell et al. (1993) a költségszerkezet elemzéseként határozzák ezt meg. Ez a megközelítés feltételezi, hogy a felhasználó a leginkább kifizetődő stratégia alkalmazását választja. Ha a különböző lokális taktikai választásoknak megfelelően egy stratégia jobbnak mutatkozik a jelenleginél, akkor a felhasználó átvált arra.

A költségszerkezet elemzés módszertanát Pirolli és Card (1999) fejlesztették tovább az információ szerzésének elméletévé. Ez az elméleti keretrendszer az evolúcióval áll kapcsolatban, egészen konkrétan az ősember kognitív és észlelési struktúráinak fejlődésével az élelemszerzéssel összefüggésben. A tanulmány szerint a modern ember az élelemszerzés stratégiáit átülteti az információszerzésre (és fogyasztásra). Az elmélet megkísérli megjósolni az emberek navigációs stratégiáit információs rendszerekben. Ezt Nielsen (2003a) két kérdésben foglalja össze:

1. Milyen nyereséget remélhetek egy specifikus információs diótól?
2. Mi a lehetséges költsége idő és erőfeszítés szempontjából a fenti dió felfedezésének és elfogyasztásának?

Így az információ fogyasztója összeveti az információ azonnali kiértékelésének és elfogyasztásának költségét a további keresés költségével.

Az információ szerzése közben az információ foltokban jelenhet meg, ami azt jelenti, hogy egy kis darab információ hasznosnak bizonyul az egyik weboldalról, míg egy másik darab egy másiktól.

3.5.3. Bőngészés vs. keresés mint információkeresési stratégia

Alapvető pszichológiai eredmény a kognitív tudományban a felismerés előbbre sorolása a visszaemlékezésnél. Sokkal egyszerűbb felismerni valamit mintsem előhívni az emlékezet mocsarából. A találati lista átfutása helyett a rendszer áttekinthető struktúráját biztosít a felhasználó számára ami jellemzi az adott információt.

A keresőkérések eredményei új, ad hoc kollekciót tárnak a felhasználó elé, amit azelőtt nem

rendeztek össze. A navigáció vagy böngészés linkek és kategóriák segítségével előre definiált információs tereket hoz létre, amely össze vannak kapcsolva egymással. Aula (2005) a következőket írja:

„Kognitív szempontból a keresés sokkal analitikusabb és erőforrás igényesebb az információkeresésben mint a böngészés, mivel magában foglal néhány fázist, mint például a kérés tervezése és futtatása, az eredmények kiértékelése, a kérés finomítása, míg a böngészés csupán annyit kíván a felhasználótól, hogy az ígéretesnek tűnő linkeket felismerje.”

Ennélfogva bizonyos esetekben sokkal egyszerűbb linkek listáját átnézni mint megfogalmazni a megfelelő keresőkifejezést ami lefed egy információs igényt. Ezt persze egyensúlyozni kell, hiszen túl sokáig nem jó nézelődni a linkek között, ha azok nem vezetnek használható információforráshoz.

A böngészés addig működik jól, amíg megfelelő irányba vezetnek a listában szereplő linkek.

Az információépítészet megkülönbözteti az információs és navigációs struktúrákat. Előbbin keresztül mozgunk utóbbival. Szerencsés esetben előbbit jól lefedi utóbbi, ezért a mozgás könnyű és gyors.

3.5.4. Információs nyomok információs struktúrákban történő navigációhoz

Információs nyomok (illatok) olyan jelek, amelyek megbízható információval látják el a keresőt arról az információról, ami nem áll közvetlenül rendelkezésre. Furnas (1997) navigálhatósági tételként jellemzi egy cél jó megtalálhatóságát az információs struktúrán keresztül azáltal, hogy jó információs nyomokkal rendelkezik minden csomópontnál/kapcsolatnál.

A találati eredményeknek információval kell szolgálniuk arról, hogy melyikre kattintson a felhasználó; erre a problémára alkalmazható az információs nyom elmélete.

3.5.5. Tájékozódás és egyéb inkrementális stratégiák

Egy gyakran megfigyelt információkereső stratégia szerint a felhasználó egy nem teljesen pontos kereséssel az információs tér számára fontos részére ugrik és innen apróbb keresési lépésekkel jut el az információig. Ez az információkeresés tájékozódási formája, amelyben a kereső apró lépésekkel nem pedig nagy, komplex keresőkérdésekkel jut el az információig.

Jónéhány tanulmány rámutatott, hogy a keresők rövid vagy általánosabb terminusokkal kezdenek és ezután az eredményeknek megfelelően változtatják a keresőkérdést úgy, hogy mindig újabb járulékos részekkel egészítik ki. Iteratív. Bates (1979) megjegyzi, hogy a komplex kérések apróbb részkerésekké történő felbontása és a probléma ezek mentén történő megválaszolása egy nagyon elterjedt és jól működő információkereső és problémamegoldó módszer.

Teevan et al. (2004) elvégeztek egy kísérletet, amelyben a fenti tájékozódási stratégiát vizsgálták. Megfigyelték, hogy a tárgyterület bizonyos szintű ismerete után a résztvevők „teleportálnak” az információs struktúrában, azaz célirányosan írnak be hosszabb, összetettebb kéréseket.

Ez a modell kevésbé megerőltető, használata nagyon elterjedt a webes keresésben. Amikor a felhasználó orientálódik, akkor gyakran fogalmazza újra a kérést és próbálkozik, amíg egy látszólag megfelelő találatot nem kap. Ezt azután alaposabban is megszemléli. Általános kérések felől indulva jut egyre konkrétabb megfogalmazásokhoz. A MASzeKer esetében ez nem így lesz, hiszen itt nagyon

pontosan megfogalmazott kérések állnak majd rendelkezésre (kontrollált nyelv). Itt tehát a teleportálás nem lesz annyira lényeges momentum.

3.6. Értelemalkotás: a keresés egy nagyobb folyamatba ágyazva

Az információhoz való hozzáférés folyamata két fő komponensre bontható: információkeresés keresésen és böngészésen keresztül valamint az eredmények elemzése és szintézise. Ez utóbbi az értelemalkotásnak is nevezett folyamat. Az értelemalkotás olyan iteratív folyamat, amely nagy tömegű információt fogalmi egység alá von.

Patterson (2001) megjegyezte, hogy a következő igényeket kell egy ilyen eszköznek kielégítenie:

szöveg kisebb egységein végzett

- egyszerű manipuláció
- megtekintés
- címkézés

valamint segítséget nyújt adatok közötti viszonyok azonosítására, követésére, róluk alkotott ítéletek felülbírálatára. Szükség volt tovább arra, hogy bővítsék az adatokat és versengő hipotéziseket zárjanak ki, illetve fel kellett ismerniük az információ hiányát, vagy hiányosságokat az információban.

Erről a 7. fejezetben alaposabban is szó lesz.

3.7. Információs szükségletek és a keresések intenciója

Az információs szükséglet terminust keresztül-kasul használja az interfész irodalom. Wilson (1981) rámutat a definíciókísérlet problematikus természetére:

Információs szükséglet alatt ne valamilyen kognitív, belső, érzelmi szükségletet értsünk, hanem az információ egy fogalmát mint eszközt, amellyel azoknak a sokkal alapvetőbb szükségleteknek a kielégítése felé haladunk.

3.7.1. Weblog alapú keresőkifejezés taxonómiák

A webes keresők előtti időkben a keresési kérések teljesen más jellegűek voltak, csak egy meghatározott szakértői réteg foglalkozott velük és gyakran hosszú és összetett kérdések voltak. Ezzel szemben a webes keresések módszere teljesen más, rövid kifejezések, széles tárgyterületi spektrumok átfogására alkalmas módszerek. Broder (2002) létrehozott egy taxonómiát a webes keresésekre alapozva:

- navigációs: a közvetlen cél egy bizonyos oldal elérése (24% felmérés / 20% log alapján)
- információs: a cél valamilyen információ szerzése, ami feltételezhetően jelen van egy vagy több webolalon (38% felmérés / 48% log alapján)
- tranzakcionális: a cél valamilyen web közvetítette cselekvés (36% felmérés / 30% log alapján)

Ez egy általánosan használt taxonómia.

3.7.2. A keresőkifejezések weblog alapú témába sorolása

A keresőkifejezések osztályozhatóak a kérés témája szerint, függetlenül az információs szükséglet típusától. Például az időjárásról lehet egyszerű előrejelzéseket is keresni, de gazdag és összetett információkat is megtudni.

Itt tulajdonképpen a webes keresési trendek témákra vonatkoztatott változásairól, hangsúlyeltolódásairól van szó.

3.7.3. A keresőkifejezések kétértelműségének weblog alapú elemzése

Kétértelmű kifejezések azok, amelyek két vagy többféleképpen is érthetőek (pl. apple). Egyértelműsítéssel kapcsolatos kérdések.

3.7.4. Újralátogatási minták weblog alapú elemzése

Gyakran előfordul, hogy az emberek már bevált információkeresési minták mentén indulnak el új információk keresésekor.

3.7.5. A megfigyelt keresési viselkedés osztályozása

Kellar et al. (2006) tanulmányukban statisztikai adatokat gyűjtöttek arról, hogy egyes felhasználók milyen típusú feladatokat végeznek el milyen gyakorisággal. Ezeket a következő főbb kategóriákba sorolták:

- Tények keresése (Fact finding): konkrét tényekkel kapcsolatos feladatok, viszonylag rövid ideig tartó, gyorsan végetérő tevékenységek. Például időjárási adatok, receptek, stb.
- Információ gyűjtése: Olyan feladat, amely magában foglalja információ gyűjtését különböző forrásokból. Több napig is eltarthat. Fuzzy határokkal rendelkezik, nem egyértelmű hogy a feladat meddig tart-hol ér véget.
- Böngészés: nézelődős információs szükséglet, semmi konkrét, mi újság típusú információkeresés.
- Tranzakciók: Online cselekvések. Emaillel kapcsolatos cselekvések, ebank, blogolás stb.
- Egyéb: egyéb tevékenységek, pl weboldal karbantartása.

Az osztályozás alapvetően kifejezi azokat a legnagyobb csoportokat, amelyekbe a legtöbb webbel kapcsolatos feladat befoglalható, de vannak hiányosságai is. Például az alfeladatokkal már nem tud érdemben mit kezdeni.

3.8. Konklúzió

Ez a fejezet összefoglalta az információkeresés legismertebb modelljeit:

- standard modell
- kognitív modell
- dinamikus (bogyószedő) modell

- fokenkénti információkeresés
- információkeresés mint stratégiai folyamat
- értelemalkotás

4. A kérés specifikálása

A kérés megfogalmazása során a kereső úgy fejezi ki információs szükségletét, hogy a belső absztrakt fogalmait átfordítja a nyelvbe, majd ezt a kifejezést egy olyan keresőkérdéssé alakítja, amellyel a keresőrendszer tud mit kezdeni. Ez a fejezet az információs szükségletek kifejezésének módjaival foglalkozni. A kérés megfogalmazásának két fő folyamata van:

1. A kereső által szolgáltatott információ. Nagyon széles spektrumot ölel fel. Természetes nyelvi kérések, kulcsszavak vagy kulcs mondatok, parancssori kérések.
2. Az interfész mechanizmus, amellyel a felhasználó dolgozik az információn. Parancssori interfészek, grafikus interfészek stb.

4.1. Szöveges kérés megfogalmazás

Kérések megfogalmazása szöveges információra általában szöveges formában történik.

4.1.1 Keresés szurrogátumokban vs. teljes szövegekben

A keresők első verziói még nem tettek lehetővé keresést a teljes szövegben, csak a könyvtáros által megadott kulcsszavakban és a címben. Mára azonban a teljes szövegekben történő keresés a megszokott.

4.1.2. Kulcsszavas keresések

A web megjelenésével jött el a kulcsszavak dominanciája a keresésekben. Ezek a keresések egy vagy több szó listáját tartalmazzák, céljuk olyan dokumentumok megtalálása, amelyek relevánsak a felhasználó számára a lista alapján.

4.1.2.1. Kulcsszavas keresések statisztikája

A kulcsszavak hosszával és összetételével kapcsolatos statisztikákat a web korai szakasza óta rögzítik. Tendencia a kulcsszavak számának és hosszának növekedése, az egy kulcsszavas keresések száma csökken. Átlagos hosszúságuk 2.4 szó volt 1997-ben 31%-ban egyszavas kifejezésekkel. Ez az arány 2005-re 18%-ra csökkent, a keresés hosszúsága pedig 2.8-ra nőtt.

4.1.2.2. Kulcsszavas keresések: statisztikai sorrend vs. kapcsolat

A web korai időszakában a keresőmotorok statisztikai relevancia alapján állították fel a találatok sorrendjét, azonban ez olykor inkább zavaró és káros volt mintsem hasznos.

A keresőkifejezések összekapcsolásának jelentése egyszerűbben megérthető mint a statisztikai alapon nyugvó értékelés. A keresőkifejezések összekapcsolása azt jelenti, hogy magasabb értéket kapnak azok a találatok, amelyekben az egyes kulcsszavak közelebb helyezkednek el egymáshoz. Sokkal elterjedtebbek lettek az évek során mint a statisztikai besoroláson alapuló algoritmusok.

4.1.2.3. Kulcsszavas keresések vs. természetes nyelvi keresések

A kulcsszavas keresés egy alternatívája az, ha lehetővé tesszük a felhasználó számára, hogy kérését természetes nyelvi mondatokban fogalmazza meg. Kutatások azt mutatják, hogy ez a módszer sokkal intuitívabb és azok az emberek, akik először találkoznak keresőinterfészsel feltételezik, hogy a kérdések efféle feltevése a jó kezdet.

Megfigyelhető, hogy a kevés keresési tapasztalattal rendelkezőkre jellemzőbb a természetes nyelven megfogalmazott keresőkérdések feladása a keresőrendszer számára. A Google egyik vezető mérnöke szerint a keresési stratégiák a „mutasd meg amit beírtam” típusról áttevődnek a „mutasd meg amit akarok” típusúra.

4.1.3. Automatizált válaszadás

Fenti súlypontát helyezés hatására egyre nagyobb hangsúlyt fektettek a különböző természetes nyelven megfogalmazott keresőkifejezések kezelésére. Ezek kormányzati támogatásból és például a TRAC versenyzőiből építkező modellek lettek, amelyek mára egy olyan rendszert eredményeztek, amely képes tényszerű kérdések megválaszolására.

4.1.3.1. Automatizált válaszadási technikák

A válaszadó rendszerek nem a semmiből próbálnak találatot keresni, hanem a természetes nyelvi kérdés alapján megtalálni a leginkább releváns bekezdést vagy dokumentumot. Ezek a rendszerek a nyelvi struktúrák alapján határozzák meg a lehetséges releváns dokumentumok halmazát, illetve a kapcsolódó részjelentések alapján szintetizálják a választ. Sok startup cég próbálkozott a keresés szent gráljának is tekinthető természetes nyelven megfogalmazott kérésekre történő válaszadással, de a nagyobb webes keresők egyelőre nem támogatják ezt a fajta információkeresési modellt.

4.1.3.2. Választípusok kérdésszerű kérésekre

A kérdések megválaszolásának egy érdekes aspektusa, hogy a keresőkifejezés kulcsszavait a válaszul adott dokumentumban nem szabad kiemelni, hanem a választ magát kell valamilyen módon megkülönböztetni, hiszen a kereső arra kíváncsi.

Lin et al. (2003) elvégeztek egy használhatósági kísérletet, ahol azt találták, hogy a résztvevők legnagyobb százaléka (53%) a bekezdés méretű részeket preferálta a leginkább, hozzátéve, hogy a legtöbb esetben egy mondat nem szolgáltat elég információt a kérdéssel kapcsolatban, míg a teljes dokumentum túlságosan nagy lehet a gyors áttekintéshez. 23% részesítette előnyben a teljes dokumentum megjelenítését és 20% a mondat megjelenítését.

Tényekkel kapcsolatos keresések esetén a frázis vagy mondatszintű találati eredményeket preferálták a résztvevők.

4.1.4. Bekezdésszintű szöveges keresések

A szöveges keresés egy másik formájában a felhasználó az általa keresett nagyon hosszú keresőkifejezéshez vagy dokumentumhoz lényegi hasonlósággal kapcsolódó dokumentumot keres. Ezt a fajta kérést sokat tanulmányozták a 80-as 90-es években. A hangsúly a magas előhíváson volt hiszen a kutatók számára fontos, hogy a lehető legtöbb releváns dokumentum kerüljön elő. A magas előhívás kívánalma a TREC versenybe is beépült, ahol a cél 1000 releváns dokumentum megtalálása volt az adott keresőkifejezésre vonatkoztatva.

Hosszú kérések esetében nem valószínű, hogy az összes keresőkifejezés megtalálható a dokumentumokban, ezért a keresőalgorithmus különböző súlyozási technikákkal állítja fel a találati sorrendet.

A webes keresések esetében nem teljesen világos, hogy a kereső mennyire szeretne egymáshoz hasonló dokumentumokat találni, azonban vannak olyan területek, ahol ennek kiemelt fontossága van. Ilyen a szabadalomkutatás is.

4.1.5. Szöveges kérések automatikus átalakítása

Az egyik leggyakrabban alkalmazott módszer a szótövesítés eljárása, amelynek során a szavakat szótövekre egyszerűsítik és azzal végzik a keresést a keresőmotorok (házaknak -> ház).

Más típusú átalakítások is előfordulhatnak, például a gyakori megállítószavak eltávolítása vagy kifejezések kiterjesztése szinonimákkal vagy kapcsolódó fogalmakkal.

Marumatsu és Pratt (2001) végeztek egy kutatást annak meghatározására, hogy az egyes szövegtranszformációs eljárásokat hogyan használják a felhasználók. A következő megállapításra jutottak:

Boolean operátorok összehasonlítása (két keresőmotor, az egyik automatikusan OR-olta a másik AND-elte a keresőkifejezéseket. A résztvevők nagy része nem tudta megmagyarázni a keresési eredmények közötti különbséget.

Megállítószavak eltávolítása. Két keresőmotor közül az egyik eltávolította ezeket, míg a másik nem. A résztvevők a „to be or not to be” kérést adták a rendszernek. Abban az esetben, amikor a megállítószavakat eldobta a kereső, a 14 résztvevő közül 9 nem tudott magyarázatot adni az üres találati listára.

Toldalékok kiterjesztése. A 14 résztvevő közül mindössze 3 tudott magyarázatot adni a runner, running (run) keresésre adott üres találati listára.

Kifejezések sorrendjére való érzékenység. Két keresőmotor összehasonlítása során az egyik súlyozta a találatokat a keresőkifejezések távolsága alapján (fire boat és boat fire). 14 résztvevőből 5 gondolta úgy, hogy a két keresés más találati listát fog adni.

A tanulmány szerint a szótövesítés hasznos lehet, de a megállítószavak eltávolítása veszélyes lehet használhatósági szempontból.

4.2. Keresőkifejezések megfogalmazása beviteli mezőkön keresztül

Manapság a keresések megadására használatos standard interfész a keresőmező egy aktiválógommbal. Gyakran tartalmaznak legördülő ajánlásokat, amelyek olyan előzetesen végrehajtott keresések listáját jelenítik meg, amelyek első néhány karaktere megegyezik a már begépelte karakterekkel (autocomplete). Weboldalak, ahol a keresést különböző kategóriákon belül is végre lehet hajtani lehetőséget adnak a célkategóriák kijelölésére.

Frantzen és Karlgren (2000) két különböző méretű keresőmezőt készítettek, az egyik egysoros volt, míg a másik hat. Megfigyelték, hogy a különböző hosszúságú keresőmezők különböző hosszúságú keresőkérdés felé orientálnak. Az első esetben az átlagos keresőkifejezések száma 2.81 szó volt, míg a második esetben 3.43.

Befolyásolhatják a keresést a keresőmező környezetében elhelyezett kifejezések. Egy bőbeszédűbb interfész arra biztatja a felhasználót, hogy legyen ő is bőbeszédű (Belkin et al. (2003)).

A keresőinterfész bemutatásakor el kell kerülni, hogy a felhasználót választások sorozata elé állítsuk, mielőtt a keresőmezőhöz hozzáférhet.

4.3. Dinamikus kifejezés-javaslat a keresőkérés megfogalmazása során

Mostanában megjelentek olyan interfészek, amelyek a keresőkifejezés bevitele közben tesznek javaslatot a keresőkifejezés pontosítására. Ezek a dinamikus keresőkifejezés ajánlók általában még a találati eredmények megjelenítése előtt használhatóak. Nagyon ígéretesnek tűnnek.

Néhány ilyen ajánlórendszer azokat a javaslatokat mutatja csak meg, amelyeknek a felhasználó már begépelte néhány karakterét. Logelemzések azt mutatják, hogy a felhasználók 30-37%-a használta a dinamikus javaslatokat keresése során.

Három fő típusát különböztethetjük meg:

1. találati lista megjelenítésekor bemutatott javaslatok (standard kifejezés ajánlás)
2. keresőkifejezés bevitele (karakterenként) közben felajánlott kifejezések
3. egyes keresőkifejezés bevitele után szűkített találati listából felajánlott kifejezések (pl első nő űrben = Tyereskova)

Mindazonáltal egyensúlyozni kell mennyi releváns kifejezést mutatunk meg a felhasználónak a keresőkérés bevitele közben, hiszen fentebb az információkeresési modellek jónéhánya megfogalmazza azt a tételt, hogy a felhasználó a keresés folyamata közben változtatja a keresés irányát, ami viszont olykor tévutakra is vezethet.

4.4. A keresőkifejezés megfogalmazása Boolean és más operátorok segítségével

A keresés korai időszakában a parancssorból végzet Boolean és egyéb operátorokkal megtűzdelt keresések voltak a megszokottak a drága gép- és keresésidő miatt (a Boolean sokkal olcsóbb). Ebben az időben még nem állt rendelkezésre a teljes dokumentum, csak a különböző minőségű bibliográfiai

adatok és valamilyen összefoglaló. Egy klasszikus példa a Dialog Pocket Guide-ből (szabadalomkutatással foglalkozó cég):

(PCR OR POLYMERASE(W) CHAIN (W) REACTION? OR DNA(W)SEQUENCE?)

AND (CANCER? OR PRECANCER? OR NEOPLASM? OR CARCINO?)

A kérdőjel csonkolást jelöl, a (W) jelölés azt mutatja, hogy a kifejezések alárendeltek egymásnak az adott meghatározott sorrendben. A két VAGY-ot egy ÉS köti össze egymással. Elég világos. Hearst több problémát is lát azonban a Boolean kérések futtatásával.

1. A találati lista túlságosan nagy lehet bizonyos esetekben, például ha a felhasználó túl sok keresőkifejezést kapcsol össze VAGY-nyal, vagy üresek, ha túl sok ÉS-t használ a találati lista szűkítése érdekében. Ez abból fakadhat, hogy a felhasználó nem ismeri a kollekción tartalmát, vagy nincs tisztában a kifejezések szerepével a kollekción belül. Ez a MASzeKer esetében nem áll szerencsére.
2. A másik probléma abban jelentkezik, hogy sokan logikátlanak tartják a szemantikáját. A legtöbb angolul beszélő azt feltételezi, hogy a hétköznapi jelentések kapcsolódnak a Boolean kifejezésekhez. Pedig nem.
3. A harmadik probléma pedig az, hogy nem a hasonlóságuknak megfelelően sorolják be az előhívott dokumentumokat. Egy tisztán Boolean típusú rendszerben egy dokumentum vagy kilégíti a kérést vagy nem.

Az általánosságban nem túl jó használhatóságuk ellenére azonban a Boolean operátorokat a legtöbb keresőmotor támogatja.

4.4.1. Kifejezések távolsága Boolean kérésekben

A keresőkifejezések távolságával kapcsolatos információk igen hasznosak lehetnek a keresés pontosságának növelésekor. A weben leggyakrabban használt az idézőjel operátor, ami a benne foglalt kifejezésekről megköveteli, hogy közvetlenül egymás mellett legyenek a találati dokumentumban. Az AOL például ezt különböző operátorok használatával segíti (dogs NEAR/3 cats – ez azt jelenti, hogy a dogs kifejezés környezetében maximum 3 szónyira lévő cats kifejezést is tartalmazó dokumentumok).

A legjobb stratégia ezzel kapcsolatban, ha lehetővé tesszük a felhasználó számára, hogy pontos szókapcsolatokat használjon, de úgy tekintünk rájuk, mint amik távolságmértéket indikálnak, feltehetően egy exponenciálisan csökkenő súlyozását.

4.4.2. Utólag koordinált és facettázott Boolean kérések

Az utólag koordinált Boolean kérések találati listája azt jelenti, hogy ha valaki például a „hal kutya macska egér” kifejezésekre keres, akkor a találati listán azok a dokumentumok amelyekben mindegyik keresőkifejezés megtalálható előbb szerepelnek, mint azok amelyekben például 30 találat van a kutyára. Az eBay vezetett be egy ilyen segítséget kereséseinél, amikor üres találati lista esetén a keresőkérésekből kivett egyet-egyet és azt mutatta meg a felhasználónak, hogy az így redukált keresőkérés hány találatot eredményezne.

Az utólag koordinált keresést nagyon hatékonyan összeköthetjük a facettázott kereséssel, így mindkét módszer egyszerre használható(111p.!!!!).

4.4.3. Web alapú hatékonyságnövelés Boolean kéréseknél

Manapság az „Összetett keresés” link mögött található Boolean kérések megadására alkalmas felületet, amely nem parancssorból beírt formulák futtatását teszi lehetővé, hanem legördülő menü segítségével facettákra fókuszált keresést tesz lehetővé.

Egy másik hasonló kezdeményezés az AltaVista keresőben jelent meg, ez a + jel, amivel a keresőkifejezések fűzhetőek, de nem mint az ÉS a Boolean kérések esetében, hanem úgy, hogy azok a keresőkifejezések, amik a + előtt vannak feltétlenül benne kell legyenek a találati listában, amik utána állnak opcionálisan.

4.4.4. Operátorhasználati statisztikák

Jóllehet szinte mindegyik webes kereső lehetővé tesz valamilyen operátorhasználatot, a statisztikák azt mutatják, hogy a felhasználók nem igazán kedvelik a keresés pontosításának ezen eszközeit. Úgy tűnik, hogy a gyakorlott keresők használják inkább őket.

4.5. Parancssori nyelvek használatával megfogalmazott keresőkifejezés

A legtöbb rendszer, amelyik a Boolean kifejezések használatát támogatja, valamilyen parancssori nyelvbe ágyazottan teszi azt. Ezek általában parancsokat (commands) (vagy „igék”) használnak argumentumokkal ellátva. Azonban a felhasználónak a rendszer ritka használata esetén problémát okoz a szintaxis megjegyzése és a szintaxis általában a legkevésbé sem rugalmas, nem igazodik a felhasználói igényekhez. Néhány esetben azonban ezek a nyelvek méltó riválisai az adatbázislekérdező nyelveknek. Mindazonáltal használatuk összetett és nehézkes lehet, sok esetben külön képzést igénylő. A legrégebbi keresőrendszereket természetesen leginkább képzett könyvtárosok használták.

Példakeresés a Dialogban:

?s (biometric? and security)/TI, LP /** szótövesítéssel Boolean ÉS keresés a címekben és a lead bekezdésekben a biometric és security kulcsszavakra**/

1621 BIOMETRIC?/TI,LP /** egyik részkeresés találati halmaza**/

40268 SECURITY/TI, LP /** másik részkeresés találati halmaza **/

S1 678 S (biometric? and security)/TI, LP /** kettejüké együtt **/

?s S1 and CS=(HARVARD AND MEDIC?) /** az S1-ként tárolt keresés további pontosítása **/

S2 52 S1 and CS=(HARVARD AND MEDIC?) /** és találati halmaza **/

?T S2/3/1-10 /** amely halmaz első tíz elemét a 3-as formátum/interfész séma szerint megjeleníti)**/

Ez a MASzeKer számára is nagyon fontos lesz valószínűleg.

A web megjelenésével azonban elkerülhetetlenné vált, hogy az úrlap alapú keresőfelületek vegyék át a vezetést (gyakoriság szerint legalábbis). Ezek általában különböző interfész elemekkel segítik a keresés megfogalmazását (legördülő menüvel, rádiógombokkal, csúszkákkal stb). Azonban például a Google-ban vannak ún. keresési rövidítések, amelyek segítségével a megszállott kereső közvetlenül egy információ konkrétumára kérdezhet rá. Pl a SWA 49 kérés a Southwest Airlines légitársaság 49-es járatáról ad információt, míg a stocks:NOK a Nokia részvényeinek aktuális értékéről. Kérdés, hogy nekünk kell-e alkalmaznunk?

4.6. Konklúziók

A keresőkérés a híd a felhasználó saját információs szükségletének megértése és a rendelkezésre álló információ között. A keresőkérés meghatározása manapság a leggyakrabban kulcsszavak begépelése egy beviteli mezőbe vagy linkek követése útján történik (a linkek keresőkifejezésként történő használatáról a 8. fejezetben lesz szó részletesebben).

Ez a fejezet a keresőkérés megfogalmazásának két fő aspektusát tárgyalta: a kereső által biztosított információ fajtáját és az interfész stílusát, ami ezen információ kifejezésére használható. A keresőkérés meghatározása úgy fogalmazható meg, mint egy cselekvés, amelyben a felhasználó meghatároz:

kulcsszavas kérést

természetes nyelvi kérést, beleértve a kérést mint kérdést

szöveg bekezdéseit

kéréseket, amelyek Boolean operátorokat tartalmaznak és

kéréseket parancssor alapú nyelvtannal.

A fejezet tárgyalta továbbá a különböző keresőrendszerek módosításait a bevitt szövegen, például a szótövesítést és a megállítószavak eltávolítást. Bemutatásra került a grafikus beviteli mező keresőkérés meghatározásához, továbbá a növekvő fontossága és népszerűsége azoknak a dinamikus keresési javaslatoknak amelyek aközben jelennek meg ahogyan a felhasználó beviszi a kérést. Más, a keresést módosító eljárások, mint például a helyesírás javítása a 6. fejezetben kerülnek tárgyalásra.

A szintaxis alapú keresőnyelvek hanyatlása (kivéve a specialisták által használtakat) egy nagyon tanulságos példája annak, hogy azok az interfészek, amelyeket arra terveztek, hogy minél egyszerűbben implementálhatóak legyenek a számítógépen átadják a helyüket azoknak, amelyek intuitívabbak a felhasználók számára. Lenyűgöző, hogy napjainkban az emberek különböző webes felméréseken azt a választ adják, hogy a webes keresőmotorokat egyszerűen tudják használni; létezésük első néhány évében az emberek még azon elégedetlenkedtek, hogy milyen zavaros a viseledésük. Nemcsak az algoritmusok, a feltételezések mögöttük is változtak, nagyobb fókuszot helyezve a hétköznapi ember intuitívására.

A jövőben valószínűleg a beszélt nyelv általi kéresemeghatározások lesznek egyre népszerűbbek; erről a mobil keresésről szóló 12. fejezetben lesz szó. Valószínű továbbá, hogy hosszabb kérések, kérések kérdéseként szervezve egyre növekvő számban lesznek használatban, ahogyan az algoritmusok kérdésekre adott válaszai javulnak.

5. A keresési eredmények megjelenítése

Ez a fejezet azon interfészeket írja le amelyek az információkeresési folyamatban a keresési eredmények megjelenítéséhez szükségesek, a manapság leggyakrabban használt ötletekre fókuszálva.

5.1. Dokumentum szurrogátumok

A találati eredmények megjelenítésének egyik legáltalánosabb formája az előhívott dokumentumok összefoglalójának vertikális listája (ezek a search engine result page-ek, vagy SERP-ek). Tipikus esetben egy elem a dokumentum címéből és fontosabb metaadatok (létrehozva, szerző, forrás, url) készletéből áll, valamint a dokumentum releváns részének rövid összefoglalójából vagy kiemeléséből áll. Ezt az egységet gyakran találatnak is nevezik. A fenti információs készlet pedig a dokumentum szurrogátuma.

A szurrogátum minősége nagy fontossággal bír a felhasználó számára a relevancia elbírálásakor. Ezen minőségek meghatározását tűzte ki célul Clarke et al. (2007) kutatása, amelyben webes keresőmotorok eredményeit tesztelték (logfájlok alapján) abból a szempontból, hogy a szurrogátum melyik részére kattintottak a leggyakrabban a felhasználók. A keresztülkattintás megfordításának trükkjét használták a kiértékelésnél, amelyben annak a szurrogátumnak a tulajdonságait amelyekre kattintottak összehasonlították azéval amelyik közvetlenül felette volt. Ha ugyanis a felhasználó megtörte azt a trendet, hogy a listában fentebb lévő (látszólag relevánsabb) dokumentumra kattintson, akkor a végül kattintott dokumentum szurrogátumának van valamilyen figyelemreméltó minősége. 10000 ilyen kiértékelés alapján Clarke et al. (2007) a következőket találta:

- Összefoglaló megvan B találatban, de nincs meg A találatban
- Összefoglaló hosszabb B találatban (>100 karakter), de rövid A találatban (>25 karakter)
- B címe több keresőkifejezés egyezést tartalmaz mint A címe
- B címe a kérdésben megfogalmazott kifejezéssel kezdődik, de A címe nem
- Cím, összefoglaló és URL együtt tartalmazzák a keresőkifejezést B találatban, de nem A találatban
- B találat összefoglalója tartalmaz egy egyezést minden keresőkifejezésre, A összefoglalója némelyikből többet is, de nem mindegyikből
- B találat URL címe a www.keresőkifejezés.hu, de A URL címe nem
- B URL címe kevesebb per jelet tartalmaz
- B URL címe kevesebb karaktert tartalmaz
- B összefogla, nem pedig A megy át egy egyszerű olvashatósági teszten

Következtetésképpen:

- (1) ahol lehetséges a keresőkifejezéseknek meg kell jelenniük a szurrogátumokban, a hozzájuk tartozó céldokumentumhoz kapcsolódó viszonyuk szerint

- (2) ha a keresőkifejezések megjelennek a találat címében, akkor az összefoglalásában nem kell feltétlenül megjelenniük
- (3) az URL címek hosszát és összetettségét csökkenteni kell amennyire lehet, és az URL-eket a keresőkéréshez való viszonyukat hangsúlyozva kell kiválasztani és megjeleníteni

5.2. KWIC avagy kérešközpontú összefoglalók

A legtöbb találati lista megmutat egy kivonatot az előhívott dokumentumból, ami összefoglalja a tartalmát. Ezt a kivonatot többféleképpen is nevezik: összefoglaló, részlet vagy absztrakt.

A modern webes keresők egy fontos tulajdonsága, hogy a keresőkifejezést is számbavevő összefoglalót jelenítenek meg. Ez a kulcsszó-kontextus (keyword-in-context- KWIC) kivonat. A KWIC nézetekben mondatrövidékek vagy mondatok csoportjai jelennek meg, amelyek tartalmazzák a keresőkifejezést, egyéb szurrogátumok mellett. A KWIC különbözik a hagyományos absztrakttól, amennyiben utóbbi célja a dokumentum tartalmának összefoglalása (azonban ez nem biztos, hogy tartalmaz találatot a keresőkifejezésre), míg előbbi a dokumentumban azon mondatait emeli ki, amelyekben megtalálhatóak a keresőkifejezések, továbbá felfedi azt a kontextust is, amelyben megjelennek ezek a kifejezések.

A kutatások azt mutatják, hogy jobb keresőkifejezés-központú kivonatokat mutatni, mintsem a dokumentum első néhány mondatát (ahogyan régebben volt).

Sok esetben az információs szükségletet már pusztán a keresőkifejezés-központú összefoglaló is kielégíti. Ezek elterjedésének legnagyobb akadályja, hogy a keresőrendszer általában nem tartalmazza a teljes dokumentumot, vagy ahhoz valamilyen szerzői jog kapcsolódik. Mindazonáltal a Google újabban elkezdte tartalmazni a teljes dokumentumot és sikeresen alkalmazza a KWIC találati eredmény prezentálás módszerét.

5.2.1 Mondatválasztás KWIC-ek kapcsán

Jelentős tervezési kérdés a KWIC megfogalmazása és megjelenítése. Néhányan kísérleteket végeztek, amelyekben a mondatokat súlyozták a dokumentumon belül elfoglalt helyüknek, az általuk tartalmazott szavaknak vagy a tartalmazott keresőkifejezés arányának megfelelően. White et al. (2003) például nagyobb súlyt adtak azoknak a keresőkifejezéseknek a találatok között, amelyek valamilyen formázást tartalmaztak (dőlt vagy kövérített).

Varadarajan és Hristidis (2006) bemutatott egy módszert, amely segítségével kérešspecifikus összefoglalókat lehet készíteni, a leginkább kéreš-releváns fragmentumok azonosítása és összekapcsolása által, gráfok segítségével megjelenítve a dokumentum struktúráját.

5.2.2 Összefoglalók hossza KWIC-ek esetén

Kompromisszumra van szükség akkor, amikor el kell dönteni, hogy inkább hosszú informatív összefoglalót vagy pedig a rendelkezésre álló területet jobban kihasználó rövid összefoglalót mutatunk a felhasználóknak.

Peak et al. (2004) elvégeztek egy kísérletet, amelyben különböző hosszúságú összefoglalókat mutattak a felhasználóknak. A következőket interfész megoldásokat hasonlították össze:

1. Standard lista, amelyben a címre kattintás előhívja a weblap teljes szövegét

2. „Azonnali” nézet, ahol egy kattintással kibővült a dokumentum összefoglalója, amelyben keresőkérések voltak láthatóak és a találat is
3. „Dinamikus” nézet, ahol az egérmutató hover állapotára dinamikusan bővült a megjelenített összefoglaló

18-ból 11 résztvevő preferálta a 2. megoldást. Egyikük sem preferálta az elsőt.

Cutrell és Guan (2007) különböző hosszúságú összefoglaló szövegeket adtak meg és azt találták, hogy hosszabb szövegek megadása esetén gyorsabb az információs szükséglet kielégítése.

Ezek és más tanulmányok arra utalnak, hogy a természetüknél fogva inkább kifejtős kérések inkább egy bekezdésnyi hosszú összefoglalót kívánnak meg, illetve azt, hogy a felhasználók nagy része már sejti, hogy milyen hosszúságú összefoglaló adja meg a választ a keresőkérdésre.

5.2.3. Mondattörödékek vs. teljes mondatok a keresőkifejezés központú összefoglalókban

Nem teljesen világos, hogy melyik a jobb, mindkettő mellett és ellen is szólnak kutatási következtetések.

5.3. Keresőkifejezések kiemelése

A kiemelés hasznos interfész elemnek bizonyul már régóta. Azt jelenti, hogy a kulcsszavakra valamilyen vizuális megkülönböztetést alkalmazunk. Ez leggyakrabban a keresőkifejezés háttérének színezését jelenti, amikor a különböző keresőkifejezéseket más-más színnel emelünk ki. Ezzel a kereső figyelmét a dokumentum releváns részeire irányíthatjuk, illetve fontos információt nyerhetünk a keresőkifejezések egymáshoz viszonyított távolságának felmérése által is. A kiemelés megjelenhet a találati listában és magában a dokumentumban is. Azokban a rendszerekben ahol a felhasználó a teljes szöveget is meg tudja nézni, akkor célszerű a teljes dokumentumban kiemelni a kulcsszavakat. Firefox Chrome böngészők például a CTRL+F kombinációval keresnek az aktuális dokumentum szövegében.

Hosszabb szövegek esetében annak áttekintése, hogy a dokumentumokban hol fordul elő a keresőkifejezés nagyon hasznos lehet. Ez a dokumentumok bélyegkép nézetével érhető el például. Ezekben a bélyegképeken is jelölni kell a releváns részeket. Ezt többféleképpen lehet megvalósítani. Az egyik a dokumentum görgetősávját használva. A Thumbar rendszer a gördítősávhoz hasonló eszközt használ a bal oldalon a dokumentum áttekintésére, a jobb oldalon pedig megjeleníti a keresőkifejezéseket és azok gyakoriságát is. Egy használhatósági kutatás szerint a standard, halszem optikai és Thumbar rendszerek közül utóbbi kiemelkedően jól használható hosszabb dokumentumokban kulcsszavak egyezéseinek megjelenítéséhez.

Egy másik kutatásban a Thumbarhoz hasonló rendszerben nem pusztán valamilyen háttérszínnel kiemelték a dokumentumon belül a kulcsszavakat, hanem meg is növelték azok méretét, olvashatóvá tették az áttekintésben a kulcsszavakat. 13 résztvevőből 10 ezt részesítette előnyben másik két interfésszel szemben.

5.4. Az eredmények listázásának további jellegzetességei

A hagyományos metaadatokhoz képest a keresési találatok más információt is tartalmazhatnak.

Oldalankénti találatok száma: a webes keresők tipikusan 10 találatot mutatnak oldalanként, linkkel a többi találati oldalra. Kísérleteztek például 30 találat megjelenítésével is, mert a felmérések szerint a felhasználók szerették volna, de mint kiderült a lista generálásához szükséges idő megnövekedése már sok volt nekik (0.4-ről 0.9 mp-re nőtt).

A relevancia grafikus megjelenítése: A különböző rendszerek évekig valamilyen számformátumban vagy ikonok segítségével mutatták a találat relevanciáját, vagy a számított egyezés mértékét. Az ikonok általában valamilyen oszlopdiaagram, vagy csillagok formájában jelentek meg. Manapság azonban ezeket egyre ritkábban használják. Különböző magyarázatok születtek ennek a jelenségnek a megválaszolására. Először is a relevancia megértéséhez szükség van a háttérrendszer valamelyes ismeretére, amit persze a mezei felhasználotól nem vár el a Google például. Másodsor az első találatok számszakilag nagyon közel vannak egymáshoz, ezért nem sokat jelent a kis különbség illetően megmutatása a relevancia minősítésében. Harmadszor pedig más szurrogátumok sokkal fontosabb információt tartalmazhatnak – például kulcsszavakat –, sokkal informatívabbak lehetnek. Negyedszer a pontszám információt adhat a konkurenciának a keresőalgorithmus működéséről. Végül pedig használhatósági tesztek szerint a felhasználókat egyáltalán nem érdeklik.

A dokumentum tartalmának előnézete: A legtöbb grafikus interfész találati listájában a dokumentum címére vagy ikonikus reprezentációjára történő kattintás a teljes dokumentumhoz vezet tovább. Néhány rendszer kísérletezett a tartalom közvetlenebb megmutatásával a weboldal elhagyása nélkül. Ilyen például a SnapShot, amely egy kis ablakban mutatja meg a teljes dokumentum előnézeti képét.

A keresési eredmények sokféleségének indikátorai: Néhány webes kereső megpróbálja érzékeltetni a keresőkifejezés sokféleségét az első néhány dokumentumban. Ez különösen fontos kétértelmű kulcsszavak esetében. Például a labs (laboratóriumok vagy labradorok) esetében az első néhány találat laboratóriumokra vonatkozik, majd következik egy link, amelyik segítségével a felhasználó egyértelműsíthet „See results for labrador retrievers”, amely alatt erre vonatkozó találatok vannak, a lista harmadik részében pedig vegyes a találati lista.

További/Kapcsolódó találatok indikátorai: Néhány kereső csoportosít linkeket egy weboldaltól és behúzva jeleníti meg őket alattuk linkkel a további találatokra arról az oldalról (5.1. ábra, Adobe Labs).

Oldallinkek: Egyes keresőmotorok ún. mélylinkeléssel teszik lehetővé a felhasználó számára, hogy hozzáférjenek a találati listából a számukra relevánsnak tűnő oldalhoz. Úgy találták, hogy a felhasználók nagyon gyakran használják ezeket a mélylinkeket a találati listából (ez az információ keresési folyamatoknál leírt „teleport” jelensége).

Rövidítések: A keresőmotorok megpróbálnak rövidítéseket biztosítani fókuszált információk szükséglet megfogalmazása esetén, egyfajta válaszológépekké válni, bizonyos kérések esetében. Például a Yahoo a „weather in Berkeley” keresésre egészen részletes időjárás adatokat és előrejelzéseket mutat a hagyományos találati lista elemeinek megjelenítése előtt. Amikre valószínűleg nincs is szükség ebben az esetben.

Találatok és média típusok vegyítése: A webes keresőmotorok egyre gyakrabban vegyítik össze a különböző típusú információs forrásból származó találatokat. Egy nagyon általános kérés futtatása esetén (cicusok) a Hakia rendszer a kérést átkonvertálja (macskák) és annak megfelelően jeleníti meg

különböző típusú találatokat a témában. Tanácsokat a gondozásukkal kapcsolatban stb.

Egy másik kérés (jels) a Microsoft keresőjén elvégezve vegyesen hoz találatokat sport pontszámokról, hírekről (sportklub, repülőgépek, hajtóműveket gyártó cégek). Informális vizsgálatok azt mutatják, hogy a különböző típusú médiát(fotó, videó) nem célszerű az első helyen megjeleníteni, mert – mivel nem minden esetben releváns- a látogatók könnyen elhagyják az oldalt. Ezeket azonban célszerű nagyon általános kérdések esetében használni, mert egyébként könnyen nyomasztóvá válhat a felhasználó számára.

Organikus találatok vs. reklámok: Egyértelmű, hogy meg kell különböztetni egymástól valamilyen módon a valódi és az áltálatokat. Ilyesmivel a MASzeKer esetében szerencsére nem nagyon kell foglalkozni.

5.5. A találati lista rendezésének hatásai

A relevancia szerinti rendezés mellett a találatokat a különböző metaadatok szerint is rendezhetik a keresők, kronologikus sorrendben, vagy idézettség szerint. A kutatások azt mutatják, hogy a keresők szinte egyáltalán nem mennek túl az első találati oldalon. Azt is megállapították, hogy leggyakrabban az első két eredményt nézik meg. Ezeket a megállapításokat alátámasztó kutatások ismertetései következnek.

Granka et al. (2004) egy szemkövetéses vizsgálatban megállapították, hogy az emberek 85%-a kattint az első találati eredményre és csak 10%-a a másodikra.

Joachims et al. (2005) megállapították, hogy a relevanciát a felhasználó számára a sorrend határozza meg még akkor is ha az objektív relevancia nem azonos ezzel.

Mindezek ellenére nem mondhatjuk azt, hogy a felhasználók vakon követik a linkek sorrendjét.

Ugyanők egy kísérletben megfordították a találatok sorrendjét és ahogy várható volt, a felhasználók az alacsonyabbnak vélt relevanciájú összefoglalókat is elolvasták a releváns dokumentum felkutatása érdekében.

5.6. Találati eredmények vizualizációja

Erről a 10. 11. fejezetben lesz szó részletesen.

5.7. Következtetések

A keresési találatok megjelenítése kritikus fontosságú a keresési rutin szempontjából. A fejezet összefoglalta a különböző kutatásokat, megmutatta, hogy mely aspektusokat kell megmutatni a dokumentumból a találati listában, a lista karakterisztikájával együtt. Jóllehet a mai keresők találati listája nagyon hasonlít a tíz évvel ezelőttiekhez, rengeteg apró változást figyelhetünk meg.

6. A keresőkifejezés átformálása

Ahogy a 3. fejezetben Hearst megjegyezte, gyakori stratégia egy általános keresés lefuttatása majd az első néhány találat áttekintése és ha ez nem kielégítő, akkor a keresőkifejezés átformálása. Ezt a ciklust addig ismétli a felhasználó, amíg kielégítő találatot nem kap vagy fel nem adja. Az előző két fejezet a kérés megfogalmazását és a találatok megjelenítését tárgyalta, ebben a fejezetben pedig a harmadik lépésről, a kérés átformálásáról lesz szó.

6.1. Az átformálás igénye

Keresőmotorok logfájljainak elemzése megmutatta, hogy nagyon sokan formálják át kéréseiket. Egy nagy kutatás keretein belül (3 millió rekord) kiderült, hogy a felhasználók 52%-a átformálta valamilyen módon a keresőkérést, 32%-a pedig 3-szor tette meg egy session-on belül.

Szükség van a jó eszközökre a keresés megformálásának folyamatában. A cél a kérések „kukázásának” elkerülése, azaz az eldobott keresőkifejezések számának csökkentése. Megfigyelték, hogy ha nem állnak rendelkezésre a keresés újrafogalmazában segítséget nyújtó eszközök, az emberek frusztráltak lesznek és teljesen irreleváns kéréseket próbálnak meg használni. Ennek elkerülése érdekében a kutatók különböző, a kérés újrafogalmazását segítő eszközt fejlesztettek ki. Ez a fejezet ezeket tárgyalja, illetve a módot ahogyan a felhasználó használja ezeket.

6.2. Helyesírás ellenőrzés és javítás

A keresési logok azt mutatják, hogy a keresések 10-15%-ában találhatóak tipográfiai vagy hejesírási hibák. Ennek megfelelően egy ezeket javító helyesírás ellenőrző hasznos eszköz lehet. A web előtt ezek jórészt szövegszerkesztő programokban voltak használatosak, amelyek egy előre definiált szótárnak megfelelően ellenőrizték a helyesírást. Azonban a webes keresések megjelenésével kiderült, hogy ez jóval komolyabb feladat. Például különböző nyelvjárások használata, cégek nevei, illetve olyanok szavak mint a Shrek, blog stb. Ezzel azonban együtt járt rengeteg felhasználói adat elérhetővé válása is. Így egyes keresőmotorok például óriási előnyre tettek szert, mert az elgépelte schwartzeneger után, ha a felhasználó beírta a helyes schwartzenegger-t is, akkor a rendszer tárolja a hibásat és a legközelebbi hasonló hibánál már tudja, hogy mi lehet a helyes megoldás. Ezzel az elírások 50%-án 88-90%-os hatékonyságot produkálnak (Cucerzan és Brill 2004; Li et al. 2006). Pozíciója a legtöbb esetben a keresőmező alatt, de az első találat fölött van. A javaslatot általában a találati oldal alján is megismétlik. Megfigyelték, hogy minél több tagból áll a keresőkifejezés, annál ritkábban használják, illetve azt is, hogy egy 100 főből álló csoportból 63-an vétettek valamilyen elírást vagy helyesírási hibát. Valamilyen formában tehát mindenképpen figyelni kell erre, azonban a kontrollált nyelv alkalmazása nagyon nagy mértékben csökkenti az ilyen jellegű hibák lehetőségét.

6.3. Automatizált kifejezésajánlás

A kérés újrafogalmazását segítő eszközök másik fontos osztályába a keresőkifejezés szűkítését vagy szélesítését lehetővé tevő javaslatok tartoznak. Ezek lehetnek a fentebb említett keresőlog alapú osztályozáson alapulóak, de lehetnek egy kategóriarendszer szerinti is.

A használhatósági tesztek azt mutatják, hogy ezek jól működnek, ha a felhasználónak nem kell a relevanciával kapcsolatos választást hoznia, illetve nem túl sok kifejezés közül kell választania.

6.3.1. Prisma

Egy kísérletben Anick (2003) integrált egy segítő eszközt, amely nem túl nagy kognitív töltéssel (12 lehetőség) segítette a választást. Ha a felhasználó kattintott valamelyikre a linkek közül, akkor a kereső összekapcsolta azt a kifejezést a keresőkifejezéssel és újra lefuttatta a kérést. A dupla kacsacsőr >> viszont kicserélte a keresési kifejezést arra, amelyikre kattintott (ennél fogva ez a feature nem volt túl feltűnő, ezért nem sokan kattintottak rá. Azt találták, hogy a felhasználók szerették ezt a szolgáltatást és minél hosszabb ideje használták, annál gyakrabban tértek vissza hozzá. A tanulmány a leggyakoribb kereséspontosítási módszernek a következőket találta:

- 25% Módosító hozzáadása vagy megváltoztatása (pl: buckets wholesale -> plastic buckets)
- 24% További információ szerzése (jackson pollock helyett museum of modern art)
- 15% Nyelvészeti „fej” hozzáadása (triassic helyett triassic period)
- 12% A keresőkifejezés más formában történő megfogalmazása (job listings helyett job openings)
- 24% Egyéb módosítások (helyettesítés hyponímákkal, morfológiai variánsokkal, szintaktikai variánsokkal stb)

6.3.2. Kifejezések ajánlásának más módjai

Hearst itt két kísérletet ír le (White et al. 2007, Divoli et al. 2008), amelyek azt bizonyítják, hogy bizonyos szakterületek esetén a kifejezések ajánlása sokkal fontosabb lehet mint a hagyományos webes keresésben. Ezzel a MASzeKer projektben is mindenképpen nagyon alaposan foglalkoznunk kell!

Divoli kísérletében egy biogenetikai információkat tartalmazó adatbázishoz készült keresőinterfész, amelyben a keresőkifejezés ajánlásának lehetőségeit vizsgálták. Az interfészből 4 verzió készült:

1. Az első verzió egyáltalán nem tartalmazott ajánlásokat.
2. A másodikban az ajánlások linkeként jelentek meg, amelyekre kattintva hozzáadódtak a keresőkéréshez.
3. A harmadikban egyenként jelölőnégyzetekkel lehetett kijelölni azokat, amelyeket a felhasználó hozzá akart adni a keresőkéréshez.
4. A negyedikben egy „Add all to query” link jelent meg, amely az összes szinonímát, homológot és egyébeket hozzáadta a keresőkifejezéshez.

A kísérletben a harmadik verziót találták a leginkább használható interfésznek, a második és a negyedik végzett a második helyen és az első lett az utolsó. Fontos, mindenképpen alaposan kell majd foglalkozni vele.

6.3.3. A kérés pontosítására szolgáló ajánlások webes keresőinterfészekben

Sok keresőinterfész használja, nagyon elterjedt lett, mert jól használható eszköz, nagyon jól alkalmazható egyértelműsítésre például.

A Yahoo keresősegédjével (Search Assistant) kapcsolatban az találták, hogy a felhasználók a kereséseik finomításakor az esetek 6%-ában használják.

6.4. Kedvelt célpont ajánlása

White et al. (2007) egy más fajta ajánlási rendszert próbáltak ki, amelyben a felhasználók a találati lista mellett egy kisebb ablakban mások által célpontként elért találatok listáját is visszakapták egy keresés után. Azt találták, hogy használható eszköznek bizonyult abból a szempontól, hogy a felhasználók így „ugrottak” egy nagyot az információs térben.

6.5. Relevancia visszajelzése

A kérés pontosításának egy másik nagyon gyakran használt módszere a relevancia visszajelzése. Ennek eredeti menete úgy történik, hogy a felhasználó valamilyen módon kijelöli a számára releváns dokumentumokat és a rendszer az ezekből levont következtetések alapján hasonlókat próbál találni. Egy másik módszer a pseudo-releváns visszajelzés, amelyben a rendszer feltételezi, hogy a megtalált dokumentumok relevánsak és ezeket használva ajánl fel a felhasználónak újabb dokumentumokat. Igen hasznos eszköznek bizonyult, mindazonáltal nem teljesít jól, ha a legmasabbra rangsorolt dokumentumok irrelevánsak.

Mindazonáltal a mindennapokban használt keresők nem igazán alkalmazzák, ami azt sugallja, hogy nem egy túl hasznos technika. Erre van néhány lehetséges magyarázat:

- A legkorábbi kiértékelések azt feltételezték, hogy a relevancia fontos volt és fontossága az előhívás javulásából ered. Manapság azonban nem az a fontos, hogy minél több releváns dokumentumot találjon a kereső, hanem, hogy minél pontosabbak legyen a találatok (precision).
- A relevancia nem mindig releváns. Néhány esetben valóban jól használható, azonban más esetekben kevésbé.
- A relevanciával kapcsolatos korai kutatások kis méretű szövegtörzseken történtek.
- Végül a valószínűleg legfontosabb szempont az, hogy a relevancia visszajelzése megkívánja a felhasználótól, hogy a relevancia mértékének helyességével kapcsolatos ítéletet hozzon, ami egy fárasztó feladat. Főként akkor, ha nem ismerik az adott tárgyterületet, amelyre a relevancia vonatkozik.

6.6. Kapcsolódó cikkek megmutatása

A relevancia alapján ajánlhatunk hasonló dokumentumokat a keresőnek. Ezek az „egy kattintás” interakciós módszerrel működnek. Például „Hasonló találatok”, „Hasonló oldalak”, „Kapcsolódó cikkek”. Ezeket egy valószínűségi modell alapján ajánlja a rendszer, hogy mennyire egyeznek az eredeti kérés keresőkifejezéseivel. Lin et al. (2008) kísérletében 2M kérésből 360.ezer esetben kattintottak az ajánlóban szereplő linkek valamelyikén (18,5%). Hasznos eszköz ha jól meghatározható szempontok szerint kapcsolódnak az ajánlott linkek a keresőkifejezéshez.

6.7. Következtetések

Ha a kezdeti keresőkifejezés nem hozza meg a várt eredményt, a felhasználó gondban lehet a kérés újrafogalmazásával kapcsolatban. Az ezt segítő eszközök nagyon hasznos részét képezhetik a keresőinterfésznek. A keresőkifejezés újrafogalmazása egy nagyon gyakori keresőstratégia, mint ahogyan azt a fejezetben tárgyalt statisztikák is mutatják – nagyjából a keresések 50%-ában jelen van a keresés újrafogalmazása.

Mind a helyesírás ellenőrzők mind a kifejezést kibővítő eszközök hatékonyan segítik a keresés újrafogalmazását – a keresőkifejezést ajánló eszközöket az esetek 35%-ában használják a felhasználók. Továbbá a kedvelt célpontok megmutatása gyakori kérések esetében és kapcsolódó

cikkek ajánlása kutatások esetében is hasznos lehet. Mindazonáltal a relevancia visszajelzése tradicionális formájában elavult eszköznek számít.

7. A keresési folyamat támogatása

7.1. Kiindulópontok a keresés számára

Az első lépés egy információs igény kielégítése során annak eldöntése, hogy milyen eszközöket használjunk és milyen kollektíókön használjuk azokat, egy folyamat, amelyet gyakran a forrás megválasztásának is neveznek.

7.1.1. *Kiindulópontok a webes keresésben*

Manapság a legegyszerűbb ellátogatni egy webes kereső oldalára vagy a böngészőbe integrált keresőmezőbe gépelni egyből a kérést. A könyvjelzők már eddig is könnyen elérhetővé tették bizonyos gyakran látogatott oldalakat, de a böngészősávból lenyíló nettörténet még egyszerűbbé teszi a gyakran látogatott weboldalak elérését (autocomplete ui egy nagyon hatékony alkalmazása). Amikor a webes keresők még nem voltak elég gyorsak és egyáltalán nem voltak megbízhatóak, akkor többen is katalógusok építésébe kezdtek. A legtovább talán a Yahoo! ment el, amely több mint 200.000 weboldalt rendezett különböző jól strukturált hierarchikus kategóriába. Ezek azonban a web növekedésével fenntarthatatlanná váltak. Másrészt pedig a keresőmotorok is egyre jobb teljesítményt nyújtottak. Viszont a gondosan szelektált katalógusokkal ellentétben a keresők olyan találatokat is gyakran listáztak, amelyek szerzőségével vagy megbízhatóságával kapcsolatban legalábbis erős kételyek merültek fel a kereső személyekben. Sajnos a különböző források megfelelőségének támogatása még eléggé hiányos a különböző webes keresőinterfészekben.

7.1.2. *Kiindulópontok online könyvtárkatalógusok esetében*

Éveken keresztül a különböző online könyvtári katalógusok szeparáltan álltak a felhasználók rendelkezésére, ezért mindegyikbe külön-külön kellett beírni a keresőkifejezést. Napjainkban azonban ezeket az interfészeket integrálták, azaz a felhasználónak csak egyszer kell beírnia a keresőkifejezést és a több forrásból érkező találatokat összevonja a rendszer. Ez azonban a kontextus elvesztésével járhat. Ezzel számolva egyre több könyvtári kereső kezdte alkalmazni a facettás keresőinterfészt, amelyben a különböző facetták szerint csoportosítva jelennek meg a találatok. Így a felhasználó könnyen választhat a számára megfelelőnek tűnő források találatai között. Egy frissebb innováció a Lexis-Nexis keresőinterfésze, amelyen a régi rendszerektől eltérően adatok nagy kollektívájában kereshetünk, azok címeiben és adatsoraiban.

7.1.3. *Interaktív dialógusok mint kiindulópontok*

A dialógus alapú interfészek az interfésztervezés kezdeteitől intenzív kutatás alatt állnak. Azonban használatuk nem terjedt el igazán, mert nem képesek nagy hatásokkal dolgozni.

7.2. A keresés történetének támogatása

A keresés folyamata támogatásának egy másik módja az, ha lehetővé tesszük a felhasználók számára, hogy megtartsák kereséseik kontextusát, forrását és találati listáját. A legegyszerűbb módja ennek a keresések és történetük rögzítése, illetve annak lehetővé tétele, hogy a felhasználó ezekhez hozzáférhessen. Ahogyan a 4. fejezetben láttuk, még a legkorábbi TTY-szerű rendszerek is lehetővé

tették a felhasználó számára, hogy kereséseit tárolja, azokhoz újból hozzáférjen és komplex keresőkérdéseket fogalmazzon meg.

Egy másik lehetséges módszer a webmorzsák (breadcumb) használata. Ezek kifejezetten gyakoriak e-commerce weboldalakon, illetve azokon az oldalakon, amelyek facettás navigációt is használnak. Megkülönböztethetjük az útvonal webmorzsákat, amelyek a felhasználó által kattintott linkeket mutatják meg, illetve a hely- vagy információs struktúra alapú webmorzsákat, amelyek az oldal struktúráját mutatják meg, jelölve abban a felhasználó helyzetét. Nielsen (2007) ez utóbbiak alkalmazását javasolja, hiszen ezek például akkor lehetnek különösen hasznosak, ha a felhasználó egy találati lista linkjére kattintva érkezett az információs struktúra egy mélyebb rétegébe. A felhasználóknak nem okoz problémát használatuk, nagyon kis helyet foglalnak el az oldalon.

7.3. A keresési folyamat egészének támogatása

Egy jóval kevésbé ismert terület a keresési folyamat egészének támogatása. Marchionini et al. (2000) bemutat egy keretrendszert, amelyben a „gyorsnézeteken” van a hangsúly. Ezek a következők:

elsődleges nézetek: dokumentumok, amelyekhez a felhasználó hozzáfért illetve a találati lista

áttekintések: lehetséges kiindulópontokat mutat, orientálja a további keresést

előnézetek: azt mutatja meg, hogy mi fog történni ha a felhasználó választ egyet a lehetséges utak közül, lehetővé téve a tudatos döntést a következő lépéssel kapcsolatban

felülvizsgálati nézetek: előző keresések áttekintése, a keresés történetének támogatása

külső nézetek: olyan információkat mutat meg a háttérben, amelyek a későbbiekben hasznosak lehetnek, de nincsenek a figyelem aktuális fókuszában

megosztott nézetek: mások által végrehajtott kereséseket mutatnak.

Nem sok interfész támogatja a keresési eredményekhez való hozzáférés ilyen alaposságát. Vannak azonban olyan szofisztikált könyvtári információs rendszerek, amelyek támogatják az ilyen és ehhez hasonló nézeteket. Például az Ariadne rendszerben a képernyő három részre van osztva, amelyben a kérés eredményeit bélyegképek formájában jelenítették meg. A kártyák három sorban voltak elrendezve. Az első sor metaaktivitást indikált (egy parancs kiválasztása a rendszer menüjéből) a második sor tartalmazta a konkrét találatokat és a legalsó a megtekintett vagy elmentett dokumentumokat.

7.4. A keresés integrálása az értelemalkotással

Ahogy a 3. fejezetben bemutatásra került, a keresés az értelemalkotásnak pusztán egy alkotóeleme. Melyek azok a jellegzetességek, amelyek ezt a folyamatot a keresés felől támogatják? Patterson et al. (2001) szerint ezek a keresési lista tartalmának áttekintése, a megtekintett dokumentumok listája, követése, javaslatok, ajánlások a megtekintendő dokumentumokkal kapcsolatban, hasonló dokumentumok keresése és bátorítás új kérések futtatására.

Az interfésznek támogatni kellene információs csoportok rugalmas újrendezését, rendezését, csoportosítását, elnevezését és átnevezését. Ezek az interfészek alkalmazzák továbbá a hétköznapi megfigyelt dokumentumrendezési szokásokat is (iratok rendezése az íróasztalon).

7.5. Konklúziók

A fejezet tárgyalta azokat az interfészeket, amelyek a keresés kiindulópontjával szolgálhatnak, általuk nyomkövethető a keresés folyamata, támogatják magát a keresés folyamatát és értelmes információ alkotására adnak lehetőséget.

8. A navigáció és a keresés integrálása

A keresés és a böngészés egy lehetséges megkülönböztetési szempontja az, hogy míg a keresés olyan kollektiókat produkál, amelyek még nem voltak abban a formában összegyűjtve, míg a böngészés egy már meglévő információs struktúrában segít eligazodni.

A kategóriarendszerek segítségével érhető el a leghatékonyabb információ mind a böngészés mind a keresés esetében. Egy rögzített kategóriarendszer értelmezhető címkék készlete oly módon, ami tükrözi a releváns tárgyerület fogalmait. Három alapvető típusát különböztethetjük meg: lapos, hierarchikus és facettás. Segítségükkel kiválasztható a dokumentumok egy részhalmaza, ezáltal a találatok szűkítése, dokumentumok csoportosítása, felosztása részhalmazokra. Rendezhetünk és sorba állíthatunk velük dokumentumokat. Fontos szerepük van a klaszterző algoritmusoknak, hiszen nagyobb dokumentumhalmazok esetében a kézi annotálás nem kifizetődő.

8.1. Kategóriák navigációhoz és szűkítéshez

Sok weboldal rendelkezik kifinomult információs struktúrával, amely kategóriákat használ, hogy segítse a felhasználó navigációját. Néhány speciális esetben ezeket standardizálják.

8.2. Kategóriák a keresési eredmények csoportosításához

Dumais et al. (2001) végeztek egy kísérletet egy osztályozó algoritmussal, amely a weboldalakat tíz felső szintű kategóriába sorolta. Hét különböző módon mutatták meg a keresési eredményeket a felhasználóknak, aszerint, hogy a kategóriák csoportjai, nevei, dokumentumcímek vagy dokumentum összefoglalók voltak láthatóak. Azt vizsgálták, hogy mely megjelenítés esetében találja meg a felhasználó az előre definiált kérésre a releváns dokumentumot. A legjobb eredményeket a kategóriacsoportok prezentálásával érték el. Azonban hozzá kell tenni, hogy a kísérletben szereplő keresőkérések nem a legrealisztikusabbak voltak.

8.3. Kategóriák az eredmények rendezéséhez és szűréséhez

A böngészésen kívül a kategóriák használhatóak a keresési eredmények rendezéséhez is. Ez az információ egyes kategóriák szerint (sorting) vagy több kategória szerinti rendezését (stable sorting) jelenti. A rendezés kimondottan jól használható olyan esetekben, ahol a kategória értelem szerűen rendezhető valamilyen szempont szerint (valós szám, egész szám, sorszám). Azonban nem működik jól, ha nincs valamilyen belső rendezési elve az információnak.

Nominális értékek rendezése helyett inkább azok szűrése javasolt esetlegesen összekapcsolva a rendezéssel. A rendezés új szempont szerint mutatja be az információt, de nem vesz el belőle, míg a szűrés csökkenti a találati halmazt a szűrési feltételnek megfelelően. Ez egy nagyon hasznos funkció, mindenképpen kellene majd használni.

8.4. Találati eredmények rendezése tartalomjegyzékhez hasonló nézet szerint

Egy másik gyakorta használt információs szervezési elv a hierarchikus fa-struktúrába szervezett információ prezentálása. Ennek egy egyszerűbb megvalósítása a tartalomjegyzék szerinti (TOC)

megjelenítés. Sok rendszer használja sikeresen ezt a módszert. Az egyik ezek közül a SuperBook (Egan et al. 1998 a,b; Landauer et al. 1993). A rendszer a találatot a bal oldalon egy kézikönyvszerű tartalomjegyzéket míg a jobb oldalon a dokumentumot mutatta. A fókusz-plusz-kontextus módszert alkalmazva ha a felhasználó a tartalomjegyzék egy eleme fölé vitte az egérmutatót, akkor a jobb oldali megjelenítés dinamikusan frissült a vonatkozó információval. Azok a részek, amelyek tartalmazták a keresőkifejezést nagyobb méretben jelentek meg a tartalomjegyzék többi részéhez képest. Mindenképpen javasolt egy hasonló interfész technológia alkalmazása. A DynaCat rendszerben a találatokat kategóriák alá rendezve prezentálták a felhasználónak a jobb oldalon egy TOC-hoz hasonló szervezésben, míg a bal oldalon az adott kategória alá tartozó találatok jelentek meg TOC-szerű ábrázolásban.

8.5. A webes tartalom hierarchikus navigációjának hanyatlása

A web korai időszakában kézzel épített kategóriarendszerek támogatták a leghatékonyabban a felhasználó keresőtevékenységét(Yahoo!, LookSmart, Open Directory Project). Azonban ahogyan a weboldalak száma nőtt és a keresőmotorok határfoka javult úgy alkonyult be ezeknek a kézzel épített kategóriarendszereknek. Továbbá ezeket a rendszereket olykor nehéz használni, hiszen elképzelhető, hogy egy adott információs igény kielégítéséhez néhány kategória összevonására is szükség lehet, amit ezek nem tettek lehetővé.

8.6. Facettás navigáció

Dokumentumok egyes kategóriákhoz rendelése azért problémás, mert a legtöbb dokumentumot több kategória alá is be lehet sorolni egyidejűleg. Egy jobb megoldás lehet, ha egyes dokumentumokhoz több kategóriát vagy attribútumot rendelünk. Ez facettázott metaadatokat használataként ismert, amely segítségével megoldódik a kötött hierarchiakon keresztül történő navigáció problémája. Az alapötlet egy olyan hierarchikus kategóriakészlet elkészítése ami egy facettának felel meg. Az ennek eredményeként létrejövő interfész a facettás navigációs rendszer.

8.6.1. A Flamenco facettás navigációs interfész

A facettás osztályozáson alapuló navigáció használata egyre elterjedtebb a weben. A Flamenco kutatási projektben a szerző megvizsgálta ezek használatát és leírta fontosságukat a weboldalak navigációjával kapcsolatban. A Flamenco egyik célja a facettás navigációs rendszerek vizsgálata volt. Hearst itt a könyvtártudománnyal kapcsolatos megalapozó viszonyról ír (Ranganathan természetesen, illetve Bates).

A metaadatok kategóriák orthogonális készletéből állnak, amelyek az információs objektum valamilyen egyedi aspektusát rögzítik. Például a festészet esetében olyan facettákat hozhatunk létre, mint például Hordozó (vászon, fa stb) Hely (Ázsia, Európa stb) Állatok és Növények, Ég és Föld, illetve információt a művészeti irányzatokról, művészekről, elkészítés időpontjáról stb. Egy facetta lehet lapos (Pablo Picasso művei), vagy hierarchikus (Bécs, Ausztria, Európa). Lehet egyértékű vagy többértékű, azaz legalább egy értéket rendelünk hozzá (36 cm magas) vagy többet (olajfestéket, tintát és vízfestéket használ). A facettán belüli részekre azok alkategóriáiként hivatkozhatunk.

Minél több facettába helyezzük az információs objektumot, a felhasználó annál könnyebben találja azt meg. A hatékony facettás navigáció egyik kulcskomponense a böngészés integrálása a kulcsszavas kereséssel. Azaz ha a felhasználó futtat egy keresőkifejezést, akkor a találati lista nemcsak a

kulcsszavakat tartalmazó dokumentumokat adja vissza, hanem a kulcsszavaknak megfelelő metaadatokkal ellátottakat is. Ez azt jelenti, hogy ha például a felhasználó a Flamencoban beírja, hogy shrine, akkor a facettás kategóriarendszer is újraszerveződik a találati lista facettáinak megfelelően. Ez minden keresőesemény(kulcsszavas vagy böngészett) lefutásakor megtörténik, így a felhasználó a találati lista szűkítésével egyre közelebb jut a releváns dokumentumhoz, de eközben megmarad a lehetőség, hogy a facettáknak megfelelően kiterjessze a keresését, ha éppen felfedezni szeretne egy bizonyos információs halmazt.

Nagyon sok apró tervezési döntést kell jól meghozni ahhoz, hogy egy, a Flamencóhoz hasonló rendszer jól használható legyen. A facettás navigációs rendszerben a webmorzsák alkalmazását különösen meg kell foglalni és nagyon ajánlott az egyes facettákon belüli útvonalat is rögzíteni. Nagyon jó megoldás található a Flamencoban.

Felmerül azonban a kérdés, hogy hogyan jelenítsünk meg hierarchikus kategóriarendszereket anélkül, hogy túlszűfölnánk a képernyőt vagy összezavarnánk a felhasználót. A Flamencoban a lépésről-lépésre lefelé haladást alkalmazzák az előrehaladásban. A felhasználó egy adott hierarchiaszint felsőbb szülőit és a konkrét kategóriák fölé menve az egerével az adott alkategória előnézeti képét mutatja meg. Így a felhasználó egyszerre három hierarchikus szintet lát át.

8.6.2. Facettás navigáció az eBay Express-szen

A facettás navigáció lett az a standard navigáció, amit információban gazdag weboldalak tervezésekor mindenképpen számításba vesznek és az esetek nagyon nagy részében integrálnak is. Ilyen az eBay Express navigációja is, amely lényegében a Flamenco rendszer adaptálása kereskedelmi környezetben.

8.7. Navigáció közösségi címkéken és közösségi könyvjelzőkön keresztül

A másik nagyon valószínű oka a kézzel épített webkategóriáknak az, hogy egyszerűen nem tudnak lépést tartani a web tartalmának növekedésével. Az annotáció meglehetősen erőforrásigényes feladat. Ennek egy fontos és egyre növekvő alternatívája lehet a közösségi oldalakon használt címkék alkalmazása (del.icio.us.), illetve a közösségi könyvjelzők megosztását lehetővé tevő oldalak elterjedése (dogear).

8.8. Klaszterizálás a keresőinterfészekben

A kategóriarendszerek egyik hátránya, hogy a kategóriákat vagy kézzel vagy valamilyen algoritmus segítségével rendelik a dokumentumokhoz. Kummamuru et al (2004) megkülönböztetik a monothetikus és polythetikus klaszterezést, ahol előbbi egy megosztott jellegzetességen utóbbi több megosztott jellegzetességen alapul.

8.8.1 Klaszterezés dokumentumok közötti hasonlóság alapján

A legismertebb és legkorábbi kutatások a dokumentumok klaszterezésével kapcsolatban a Scatter/Gather projektben manifesztálódtak. A cél egy automatikusan és dinamikusan generált tartalomjegyzék készítése volt egy dokumentumhalmaz navigációjához.

A kifejezéseket súlyozták és vektorként reprezentálták, a klaszterezés hierarchikus agglomeratív módon történt, ami egy polythetikus megközelítés. Az összefoglalók témák főkategóriáit tartalmazták, amelyek az egyes klasztereket általában meghatározták.

Mindazonáltal használhatósági tanulmányok azt mutatták, hogy az interfész alig teljesített jobban

egy hagyományos keresőmezős interfésznél. Úgy tűnik, hogy egy sokkal hatékonyabb megoldás, ha kombináljuk a klaszterezést a hagyományos keresőmezős interfésszel úgy, hogy a klaszterező algoritmust a találati listán futtatjuk le. Hasznos lehet, ha a felhasználókat egy bizonyos irányba akarjuk irányítani, megkímélve őket a kevésbé vagy egyáltalán nem releváns dokumentumok megtekintésétől. Fontos azonban, hogy törekedjünk a klaszterezés kapcsán a konzisztens szintek kialakítására a klasztereken belül.

Problémát jelenthet az, hogy a dokumentumok hasonlítanak egymásra – de sokféleképpen. Szóval a polythetikus klaszterezéssel vigyázni kell.

8.8.2. Klaszterezés megosztott kifejezés alapján

A monothetikus klaszterekkel végzett longitudinális kísérletek (Findex) során azt találták, hogy a felhasználók 26%-a használta a találati dokumentumok szűkítésére a klasztereket, továbbá azt, hogy a klaszter használata az idő előrehaladtával egyre nőtt.

A keresőkifejezések elemzése arra világított rá, hogy a klaszterezés segítségével a hagyományos találati sorrendben hátul végző dokumentumok feljebb kerülhetnek a találati listán, illetve nagy segítséget nyújthat még a keresőkifejezés egyértelműsítésében.

A Findex klaszterei felfedező jellegű feladatoknál bizonyultak nagyon hatásosnak, ahol a feladat nem egy konkrét dokumentum megtalálása volt, hanem több releváns dokumentum összegyűjtése. Egy fontos megválaszolandó kérdés a klaszterezést támogató interfészek kapcsán az, hogy pontosan hány csoportot kell megmutatni? Hasonlóképpen a keresőkifejezés ajánló rendszerekhez, itt is elmondható, hogy a kevesebb olykor több.

8.8.3. Klaszterezés a weben

Az évek során sok weboldal kezdte el támogatni valamilyen formában a klaszterezési eljárások valamelyikét. Jóllehet az algoritmusok nem publikusak, általában a monothetikus algoritmusok csoportjába tartoznak. Minőségükben igen eltérőek lehetnek.

8.9. Klaszterek vs. kategóriák a keresőinterfészekben

A fejezet a keresési eredmények szervezésének módszerét, illetve a navigáció és a keresés összekapcsolásának lehetőségeit tárgyalta. A két alapvető módszer a kategóriarendszerek és a dokumentumok klaszterezése volt.