

Az egységes keresőnyelv megalkotásának egy lehetséges menetrendje

BEVEZETÉS

IT és kulturális környezet

A tudásalapú társadalom kihívásai sokféleképpen megfogalmazódtak. Az elmúlt években (évezredekben) **a fő problémát alapvetően az információhiány jelentette**: az információk térben szeparáltak voltak, nehéz volt a hozzáférés és egyáltalán, az ismeretek katalogizálása is hiányzott, tehát a megszerezhető ismeretek listája sem volt megfelelően elérhető. Így ebben a fázisban a nagy problémát magának az információknak a megszerzése jelentette.

A XX. század végén az informatika és a kommunikáció forradalmával a helyzet alapvetően megváltozott: a fő problémát nem az információk megszerzése jelenti, hanem az információk szelekciója, értelmezése, azaz **a legfőbb gond a ránk zúduló információk közül a releváns információ kiszűrése**.

Természetes és kézenfekvő az a gondolat, amely szerint, ha az információrobbanást az informatika fejlődése hozta, akkor a technika fejlődése megoldja önmagától ezt a kérdést, hiszen ez adatkezelési kérdés, az pedig számítástechnika, amely dinamikusan fejlődik és az egységnyi teljesítmény egyre olcsóbb. Ez, sajnos, nem igaz: önmagában adatfeldolgozási kérdésként nem lehet a problémát relevánsan megoldani, az ismeretek szelektív elérése nem csupán hardverkérdés.

Mindez ismét felveti annak a szükségességét, hogy az információkban való eligazodást ne csak információtechnológiai kérdésként, azaz technikai-technológiai kérdésként kezeljük, de az információk megfelelő, szelektív elérése érdekében ismeretprezentációs oldalról is újra meg kell vizsgálni a kérdést.

Ugyan ez a megállapítás természetesen egyértelműen érvényes az egészségügy szakirodalmának feltárására, kereshetőségére is. Az elektronikus adattárolás és az ehhez kapcsolható távhozzáférés átalakította az irodalomkeresést. A könyvtárból, a szaksegítségől lassacskán kiszorul a feladat: a felhasználó maga fér hozzá – a könyvtár fizikai kihagyásával – az információkhoz. A nagy kérdés a szakma (szakkönyvtárak) számára az, hogy ez a kimaradás tényleges lesz vagy csak virtuális, azaz a könyvtárak alapozzák-e meg a későbbiekben a szakemberek könyvtáron kívüli munkáját.

Mindez szükségessé teszi annak a kérdésnek a felvetését, hogy milyen irányban kell továbbfejlődnie az irodalomfeltáró, illetve az ehhez kapcsolt leíró rendszereknek.

Várható tendenciák

Vélhetően két ellentétes, de egyaránt jelenlevő tendencia fog érvényesülni.

Az egyik várható tendencia leértékeli az információkeresés szabályozását, szakmai előkészítését. Az internettechnika, a WEB-használat az általános kultúra részévé teszi a keresőfelületek alkalmazását, és miután a szakirodalom tárolása, elérhetősége is egyre nagyobb mértékben full text formában lesz a világhálón elérhető, az irodalmat keresők egy része nem fog speciális keresőrendszereket használni, hanem a megszokott WEB-es keresőrendszerét fogja itt is alkalmazni. Ezek a rendszerek a teljes szöveg feltárásán alapulnak, nincs mögöttük semmilyen kategorizált, strukturált leírónyelv. Vélhetően ez a módszer lesz elterjedt a laikus felhasználók számára. A módszer pontatlan, nehézkes, több keresést tesz szükségessé, de a WEB-felületek terjedésével hozzászokunk ehhez a hátrányhoz, hiszen az élet egyéb területein sem áll rendelkezésre strukturált keresőindexálás. Ugyanakkor a WEB-keresők is folyamatosan fejlődnek, ma már csak látszólag hiányzik mögülik az intelligencia: a háttérben – a felhasználó elől rejtve – számos szinonimaszótár, illetve csoportképző, „lásd még ...” funkciójú adatbázis működik.

A másik tendencia várhatóan az lesz, hogy a (speciális, szakmai) keresőrendszerekben a hozzáadott érték, a mögöttes tudás lesz a meghatározó. Miután a speciális, szakmai keresőrendszerekben speciális tudásra van szükség, várhatóan megmaradnak, illetve továbbfejlődnek a szakmai keresőrendszerek. A szakmai keresőrendszerek titka a mögöttes tudás. Ezek a mögöttes tudások ugyanakkor ágatszpecifikusak, mindez azt vetíti elő, hogy mégis szükség lesz egy korszerű keresőnyelv megalkotására.

ELVI MEGKÖZELÍTÉS

A bevezetés értelmében szükség van egy korszerű irodalomleíró, irodalomfeltáró nyelv létrehozására vagy legalábbis a meglévő korszerűsítésére. Ez utóbbi kifejezés nem pontos: ma Magyarországon nincs elfogadott szótárkészlet, amely alapját képezhetne ennek a munkának, így korszerűsítésről – alap híján – nem lehet beszélni.

Lehetséges alternatívák – döntési kényszerek

A keresőnyelv megalkotására több lehetőség van, és ezeket az alternatívapárokat célszerű külön-külön megvizsgálni.

A keresőnyelv megalkotásában logikailag két jól elkülöníthető alapelv szerint lehet elindulni:

- **keresőnyelv** (leírnyelv) megalkotása,
- keresőnyelv feladatát is betöltő **fogalomtár** alkotása.

KERESŐNYELV (LEÍRÓNYELV)

A keresőnyelv fogalma ugyan használt, de nem eléggé egyértelmű a meghatározása. Ezért – legalább e tanulmány keretére érvényesen – saját definíciót adok. **Keresőnyelv** alatt egy olyan

- rögzített elemkészletű,
- hierarchikus szerkezetű,
- egyértelmű meghatározású,
- egyértelmű szinonimákat tartalmazó

„szótárt” értek, amely alkalmas egy dokumentum tartalmi feltárására, ennek kategorizált azonosítására és leírására. **A keresőnyelv ennek megfelelően szavakat, kifejezéseket, szóalakokat tartalmaz.**

A keresőnyelv ugyan tartalmazhat más tengelyek felé utalásokat, elágazásokat, de alapvetően síkban írja le a fogalomrendszert.

KERESŐNYELV FELADATÁT IS BETÖLTŐ FOGALOMTÁR

A **fogalomtár** ezzel szemben

- szintén rögzített elemkészletű
- hierarchikus szerkezetű
- egyértelmű meghatározású
- egyértelmű kapcsolatleírású

fogalomrendszer, amelyben az elemkészlet nem a fogalom lehetséges megnevezéseit, hanem magukat a fogalmakat tartalmazza. A fő különbség tehát abban áll, hogy a szótár (keresőnyelv) megnevezés-központú, a fogalomtár pedig objektumorientált. A fogalomtár természetesen kiegészítendő hozzákapcsolt szótárakkal, és ezek a szótárak funkcionálnak keresőnyelvként. A fogalomtár-alapú keresőnyelv ugyanakkor rendelkezik azzal az előnnyel, hogy, miután a fogalmakat azonosítja és ahhoz rendel szótárakat, alkalmas arra, hogy különböző szótárakat használjon a felhasználó és a keresőrendszer, azaz lehet az egyik nyelven végezni a keresést, míg a keresőnyelvet használó programok képesek más nyelveken is keresni a fogalom megfelelőjét.

Orvosi vagy egészségügyi fogalomtár készítése

A keresőnyelv megalkotásánál a következő releváns kérdés az,

hogy hagyományos orvosi (a klinikai tudományokat és az azokat megalapozó tudományágakat) tartalmazza-e, vagy egységes szerkezetben kell tartalmaznia az egészségügy új területeit, amelyek az egészségügyi tervezés, szervezés, köz- és népegészségügyi területei (egészségügyi informatika, egészségpolitika, egészségügyi finanszírozás, egészségügyi management, egészségügyi minőségbiztosítás stb.).

Ma a fent felsorolt új szakterületek részint megerősödtek, másrészt nem csupán a szakterületek szűk szakembergárdájának érdeklődésére tartanak számot. Ma ezek a területek az egészségügy széles rétegeit érintik, így ezen szakterületek egységes feldolgozhatóságát, kereshetőségét biztosítani kell.

Szerkezet – sík vagy térleíró fogalomtár létrehozása

A leírnyelv/fogalomtár logikai szerkezetében nem mindegy, hogy a fogalmi hierarchiát és kapcsolatrendszerét síkban vagy térben kívánjuk ábrázolni. A sík ábrázolás a fő nézet szerinti kapcsolatokat írja le. A térbeni ábrázolás a fogalmak lehetséges kapcsolatrendszerét írja le, és különböző nézetek (szempontrendszerek) szerint csoportosítható módon adja vissza a fogalom lehetséges struktúráit.

Leírás alapja

El kell dönteni azt is, hogy mi legyen a leírás alapja, kiinduló pontja.

Itt a döntésnek két tényezőre kell kiterjednie:

- egyrészt, hogy a leírás alapjául mit tekintünk,
- másrészt, hogy milyen IT támogatással, milyen eszközben, programnyelven készül maga a leírás.

A leírás alapjául az alábbiak szolgálhatnak:

- a legjobb fogalomtár fordítása,
- a legjobb keresőnyelv fordítása,
- a legjobb keresőnyelvet alapul véve önálló fogalomtár megalkotása.

Ma technikai és mentális okok miatt ez utóbbi megoldás tűnik a legrealisabbnak. Várhatóan egy létező, elismert keresőnyelv (tezaurusz) fogalomkészletét és alapkategória-rendszerét kell kiindulásnak tekinteni, és ennek alapján kell egy fogalomtárat elkészíteni.

A FOGALOMTÁR ALAPJAI

A fentiek alapján megkísérlem összefoglalni a fogalomtár javasolt tulajdonságait.

A fogalomtár javaslatunk szerint térben ábrázolható szerke-

zetű, hierarchizált és a fogalmak kapcsolatait leíró fogalomrendszer.

A fogalmak belső azonosítása

A fogalomtár belső azonosítását nem valamely ismert szótár elemkészlete adja, hanem a fogalmak belső azonosítására egy erre a célra képzett azonosítórendszer való. A felhasználónak ugyanakkor nem feltétlenül kell ezzel a fogalmi azonosítóval találkoznia. Ezt az elvet szemléletes példával úgy érzékeltethetjük, mintha a BNO-kód csak a számítógépes rendszerek belső programjában lenne, ugyanakkor a beírt szöveget a számítógép a kóddal azonosítaná, majd a végprocesszálsnál a kódhoz rendelt szöveg jelenik meg ismét a felhasználó előtt.

A fogalmak, szótárak, szinonimák

A fogalmak különböző tereihez szótárak, azon belül szinonimák rendelendők.

SZÓTÁRAK

A szótárak alatt a fogalomrendszernek egy nyelvkészlethez rendelését értem, azaz a valamilyen szempontból szelektált elemkészlet azonos nyelvi megfeleltetését. A nyelv lehet természetes (élő vagy holt nyelv, mint a magyar vagy a latin) és lehet mesterséges: lásd kódrendszerek.

A szótárak alapvetően három csoportba sorolhatók.

- Természetes nyelvi szótárak (angol, magyar, latin, orvosi nyelv)
- Szinonimák:
 - Megengedett (elfogadott) szinonimák
 - Javító szinonimák
- Ismert és elfogadott kódrendszerek (BNO, OENO, SNOMED, stb.), azaz mesterséges nyelvek

Természetes nyelvi szótárak

A fogalomtárnak szüksége van alapszótárakra, amelyek a fogalom egy-egy ismert nyelvén tartalmazzák a fogalom megnevezését. Javasolt szótárak minimum: az angol, a magyar (köznyelv), a latin, az orvosi nyelvi.

Szinonimák

A rendszerben az adott szótárakhoz kapcsoltnak szükség van szinonimákra is. A szinonimaszótárakat is két módon kell elkészíteni: megengedett (elfogadott) szinonimák, javító szinonimák.

A megengedett (elfogadott) szinonimák az egyes fogalmaknak az ekvivalens értékű, nyelvileg, ortográfiailag elfogadott változatai, alakváltozatai. Természetes, hogy ezek a szinonimák csak az adott nyelven belül értelmezhetők.

A javító szinonimák ezzel szemben a fogalomnak az elterjedt, de a helyesírásában, nyelvhasználatában meg nem engedett alakjai. A javító szinonimák sohasem jelennek meg a rendszerben választható alakként vagy outputként. Csupán arra szolgálnak, hogy az adatbeviteli hibák esetén automatikus javítások legyenek felajánlhatók.

Ismert és elfogadott kódrendszerek (BNO, OENO, SNOMED stb.)

A szótárak nem csupán a természetes nyelvek elemkészleteivel értelmezhetők. E tekintetben a fogalomtárnak ugyanúgy egy szótárja a BNO-, az OENO-, a SNOMED-kód, csupán itt a fogalom nem minden eleméhez lehet ebből a szótárból elemet rendelni.

Ezek a lehetőségek együttesen biztosítják azt, hogy a fogalom bármilyen elnevezését inputként megadva eljuthatunk a fogalom azonosításához, majd a fogalom a processzálsban, illetve az outputlistákban tetszőleges „nyelven” használható, megjeleníthető.

JAVASOLT MENETREND

A fogalomtár megalkotása vélhetően az alábbi menetrend szerint lehetséges a legcélszerűbben.

- Fogalomtár és leírónyelv elveinek, szerkezetének, technológiájának meghatározása a vitafórum alapján
- Tesztadatbázis készítése (egy-egy terület leírása)
- A teszt adatbázis kipróbálása
- Javaslat készítése a leírónyelv szerkezetére, elveire
- A javaslat megvitatása a szakmai kollégiumokkal és szakmai fórumokkal
- Egy alapstruktúra előkészítése (fordítás alapján)
- Az egyes szakterületek revíziója a szakmai kollégiumok által
- Az akkreditált folyóirat fogalmának definiálása és bevezetése
- IT segédletek készítése
- rendszer bevezetése

Fogalomtár és leírónyelv elveinek, szerkezetének, technológiájának meghatározása a vitafórum alapján

A vitafórum arra szolgál, hogy az itt felvetett kérésekre a szakma nyilvánosságának bevonásával választ találjon. Ennek alapját részint a MEDINFO által felkért szerzők vitaindító tanulmányai, másrészt a vitafórum hozzászólásai adják.

Tesztadatbázis készítése (egy-egy terület leírása)

Az így megállapított alapelvek szerint kell elkészíteni egy-egy (szűkebb) szakterület leírónyelvét, illetve a mögöttes fogalomtárat. A két tesztadatbázisnak eltérő területről kell származnia, azaz az egy klinikai és egy egészségpolitikai terület kidolgozására van szükség.

A tesztadatbázis kipróbálása

Az így kialakított leírónyelvmintát „élesben” kell tesztelni, azaz meg kell vizsgálni, hogy alkalmas-e a szakirodalom feltárására és leírására.

Javaslat készítése a leírónyelv szerkezetére, elveire

Ezután kell véglegesíteni a javaslatot, amely két forrásból táplálkozik: egyrészt a vitafórumból, másrészt a tesztállomány használatából.

A javaslat megvitatása a szakmai kollégiumokkal és szakmai fórumokkal

Az így előkészített javaslatot széles szakmai vitára kell bocsátani, ahol elsősorban a szakmai kollégiumoké a főszerep.

Egy alapstruktúra előkészítése (fordítás alapján)

Ezután kell egy alapanyagot elkészíteni, amelynek az alapja a fordítás adaptációja.

Az egyes szakterületek revíziója

Az így előkészített anyagokat az illetékes szakmai kollégiumoknak kell kiegészíteni, véleményezni, korrigálni.

Az akkreditált folyóirat fogalmának definiálása és bevezetése

Ezzel párhuzamosan célszerű megkísérelni a magyar egészségügyi-tudományos folyóiratok akkreditációs rendszerének kialakítását. Itt el kell érni, hogy a minősített folyóiratok cikkeinek kulcs- és tárgyszavazása e rendszer szerint készüljön.

IT segédletek készítése

Mindehhez megfelelő IT segédletek szükségesek, mind az elkészítéshez, mind az intelligens használathoz.

A rendszer bevezetése

Ez a legvégső fázis, amely természetesen legfeljebb csak az akkreditáció szintjén lehet kötelező.