

Automatikus tartalmi osztályozás és társítás kidolgozása az Igazságügyi Minisztériumba beérkező állampolgári levelekre

Kabai Dóra¹, Bigazzi Sára², László János^{1,2}

¹ MTA Pszichológiai Kutatóintézet, 1132 Budapest, Victor Hugo u. 18-22.
{kabaidora, laszlo}mtapi.hu

² PTE BTK Pszichológiai Intézet, 7624 Pécs, Ifjúság útja 6.
bigazzisara@hotmail.com

Kivonat: Kutatásunk⁸⁵ célja egy olyan automatikus tartalomelemző módszer kidolgozása, amely az IM ügyfélszolgálatának munkáját segíti. A munkafolyamat összetettsége miatt két részfeladatot választottuk ki kutatásunk tárgyául, amelyek elképzelhetőek automatikus működésben. A levelek elsődleges szétosztása a főbb jogi területek mentén történik, a minisztérium szerkezeti felépítését követve. A jog logikájára támaszkodva kidolgoztunk egy kategóriarendszert, amiben kódoltuk a leveleket, majd csoportosítással vizsgáltuk a tematikus kategóriák differenciálhatóságát. A téma szerinti osztályozást tanulóprogramok segítségével végeztük. Mélyebb struktúrák feltárásához egy új módszertant dolgozunk ki. A természetes nyelven írt levelekben a témára és az eljárás stádiumára vonatkozó információk, események sorozata lefordítható a közigazgatás nyelvére. A szűk jogi területet érintő ügyek kifutási lehetőségei definiálhatók döntési utakként. Az utak leírásához sematikus forgatókönyvek alkalmazását javasoljuk, amelyek események sorozatát és köztük lévő relációkat is képesek magukban foglalni [2], [3], [4]. A levélben megjelenő események láncolata szűkíti a releváns forgatókönyvek halmazát, így lehetővé válik a megoldási lehetőségek automatikus társítása.

1 Levelek elosztása

Az ügyfélszolgálat komplex munkafolyamatának feltérképezése után elkülönítettünk olyan feladatrészeket, amelyek elképzelhetőek emberi jelenlét nélkül. A levelek beérkezésük után egy elsődleges osztályozáson esnek át, ahol az érintett jogi területek szerint továbbküldik őket a megfelelő osztályra. Az ügyfélszolgálat tapasztalataira és a jogszabályokra támaszkodva rögzítettünk egy kategóriarendszert, amelybe a leveleket besoroltuk. Klaszterező eljárással vizsgáltuk a korpusz klaszterekbe sorolhatóságának mértékét. Eddigi eredményeink azt mutatják, hogy sem az entrópia, sem a tisztaság értéke nem tér el jelentősen az 50%-tól. További futtatások szükségese-

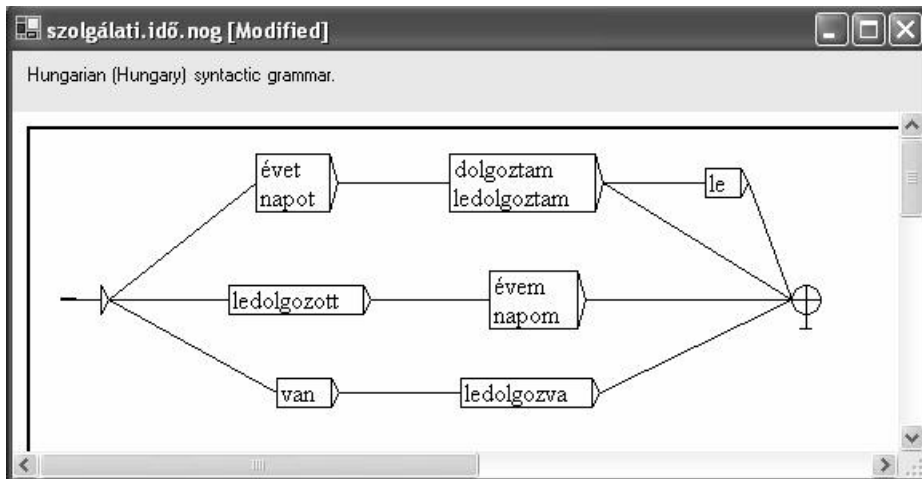
⁸⁵ Kutatásunkat az NKFP 6/074/2005 pályázat támogatja.

ahhoz, hogy végső következtetéseket vonjunk le, például a minta növelésével javulhatnak az eredmények.

A klasszifikáció eddigi eredményei eltérést mutatnak a téma és más dimenziók esetében, ami alátámasztja azt az elgondolásunkat, hogy a tematikus jegyek elemzése szavak, kifejezések szintjén elégséges lehet. A találati arány javuló tendenciát mutatott a tanuló-, tesztadat arány változtatásának függvényében, azonban további fejlesztések, a korpusz növelése szükséges, hogy a találati arány elfogadható hibaszázalékkal működjön.

2 Társítás

A egyedi esetkehez való megoldási lehetőségek automatikus társításának feltétele a nyelvek átjárhatósága és a probléma pontos beazonosítása. A természetes nyelven írt levelekben megjelenő információk, események lokális szintaktikai, szemantikai szabályok segítségével lefordíthatók rögzített közigazgatási kifejezésekre⁸⁶. Az 1. ábrán láthatunk egy gráfot, ami a szolgálati időre vonatkozó kifejezéseket kódolja. Ha a levélben elmesélt történet minden pontját lefordítjuk jogi nyelvezetre, tulajdonképpen egy jegyzetet készítünk a levélből, amely a jog fogalmi rendszerében értelmezhető.



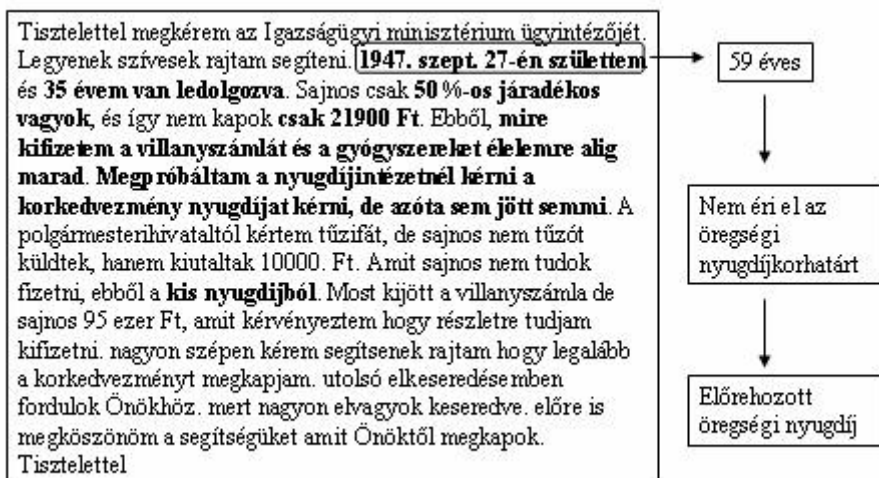
1. ábra. A szolgálati időre vonatkozó kifejezések gráfja.

A probléma pontos beazonosításához az érintett jogi területet döntési utak formájában felvázoljuk. Az utak események kötött sorrendjéből állnak, amelyekhez egyértelműen társíthatók kifizetési lehetőségek. Minden út leírásához egy szemantikus forgatókönyvet használunk, amely jogi kifejezéseket tartalmaz. A levél fordításából készült eseménysor összevethető a forgatókönyvekkel, amelyek az egyezés mértéke szerint sorrendbe

⁸⁶ A Nyelvtudományi Intézet fejlesztése és együttműködése révén nyílt lehetőségünk a Nooj nevű program használatára.

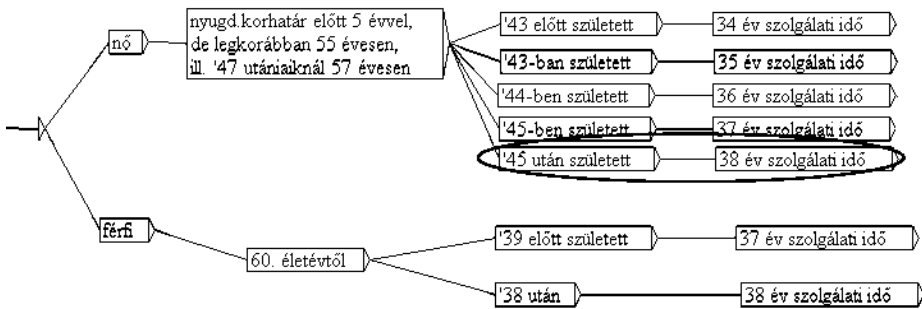
rendezhető. Ezzel a módszerrel a megoldási javaslatok listája automatikusan társítható az állampolgárok leveleihez.

A következőkben egy példán keresztül szemléltetjük a társítás folyamatát. A nyugdíjjal kapcsolatos panaszok, kérések úgy tűnik jól elhatárolható csoportot alkotnak. A beérkező levél nyelvi markereinek detektálása révén szűkíthetjük a vonatkozó jogi területet és a releváns forгатókönyvek halmazát. Az 2. ábra egy panaszos levelet tartalmaz, amelyben nyelvi markerek utalnak a nyugdíj típusára („1947. szept. 27-én születtem”, „35 évem van ledolgozva”, „50%-os járadékos vagyok”), az eljárás folyamatának állapotára („Megpróbáltam a nyugdíjintézetnél kérni a karkedvezmény nyugdíjat kérni, de azóta sem jött semmi”).



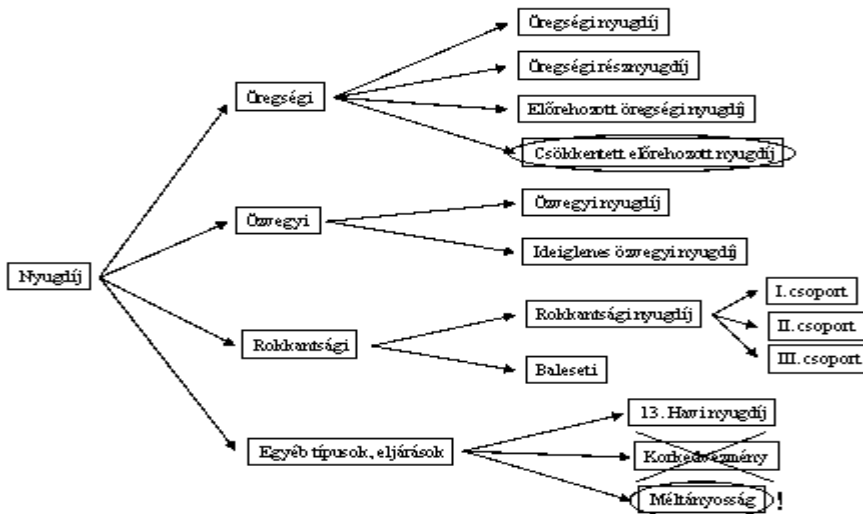
2. ábra. Egy nyugdíjjal kapcsolatos állampolgári levél. A nyelvi markerek detektálása szűkíti a vonatkozó jogi területet.

A születési évszám arra vonatkozó információt tartalmaz, hogy az állampolgár elérte-e a nyugdíjkorhatárt. A jogszabályok egyértelmű feltételeket szabnak az életkor és a szolgálati idő tekintetében. A 3. ábrán látható gráf mutatja az előrehozott öregségi nyugdíj feltételi hálózatát. A példában szereplő eset adatai alapján nem jogosult előrehozott öregségi nyugdíjra, mert nem rendelkezik az előírt minimális szolgálati idővel.



3. ábra. Az előrehozott öregségi nyugdíj feltételei egy döntési fával modellálva.

A levélből kinyert információk és a jogi adatbázis oda-visszaható szűkítése és bővítése révén zárhatunk ki, illetve vehetünk fel megoldásként újabb kifutási utakat. A 4. ábra szemlélteti, hogy a példában felmerültek alapján az előrehozott öregségi nyugdíj kizárásával a nyugdíj típusát illetően meghatározható a csökkentett előrehozott nyugdíj. Míg az egyéb eljárások tekintetében kizárható a karkedvezmény, azonban felmerül a méltányosság kérelem beadásának lehetősége.



4. ábra. A nyugdíjhoz kapcsolódó jogi terület ábrázolása egy döntési fa formájában.

Bibliográfia

1. Abonyi J.: Adatbányászat. A hatékonyság eszköze. Gyakorlati útmutató kezdőknek és haladóknak. ComputerBooks, Budapest (2006).

2. Graesser, A.C., Pomeroy, V., & Craig, S. (2001). Psychological and computational research on theme comprehension. In W. van Peer and M.M. Louwse (Eds.), *Thematics in psychology and literary studies* (pp. 19-34). Amsterdam: Benjamins.
3. László J.: *A történetek tudománya. Bevezetés a narratív pszichológiába*. Budapest: Új Mandátum Könyvkiadó (2005)
4. Schank, R. C., Abelson, R.: *Scripts Plans Goals and Understanding – An Inquiry into Human Knowledge Structures*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 1977.
5. Schank, R. C.: *Dinamikus emlékezet. A forgatókönyv-elmélet újraértelmezése*. Budapest: Vince Kiadó. (2004)