

FÓRIS ÁGOTA¹ – SOMOGYI ZOLTÁN^{2,1} – B. PAPP ESZTER^{3,1}

¹Károli Gáspár Református Egyetem, TERMIK (Terminológiai és Kommunikációs Kutatócsoport)

²Oracle Global Services Hungary

³Translators without Borders

foris.agota@kre.hu

<https://orcid.org/0000-0002-7599-0323>

zoltan.somogyi@protonmail.com

<https://orcid.org/0000-0002-3721-6513>

eszter.papp@gmail.com

<https://orcid.org/0009-0001-5577-9897>

Fóris Ágota – Somogyi Zoltán – B. Papp Eszter:

Magyar Nemzeti Terminológiai Adatbázis tervezése. Általános kérdések

Alkalmazott Nyelvtudomány, XXIV. évfolyam, 2024/2. szám, 21–41.

doi:<http://dx.doi.org/10.18460/ANY.2024.2.002>

Magyar Nemzeti Terminológiai Adatbázis tervezése Általános kérdések¹

Planning the Hungarian National Terminology Database. General questions

The need for a national terminology database in Hungary has been repeatedly raised in the last decade, especially as most European countries already have their national term base or a terminology portal. Planning the Hungarian National Terminology Database and developing its financial, human resource, and software requirements were listed among the general recommendations for creating the Hungarian terminology strategy (Fóris & Bölskei, 2019: 157). Based on these recommendations, in 2021–2023, the Hungarian Research Centre for Linguistics did the preparations required before launching the work on Hungarian terminology strategy (such as surveying relevant standards and studying existing term bases in Europe; for details, see the articles published in the green paper: Prószéky et al., 2023). The plans are to be implemented in the framework of the MTA TMNP2023-1/2023. "Magyar terminológiastratégia" project from 2023 to 2027, in which one sub-project is tasked with planning and launching the Hungarian National Terminology Database (MTA web).

This paper examines the general questions relevant to designing the Hungarian National Terminology Database. Section one offers a background and a brief overview of existing literature on terminology management and terminology strategy. Section two focuses on the primary purpose of such a national termbase and the general aspects to consider. The termbase needs to serve a wide range of target audiences, including human users (editors, experts, and the general public) and machine learning applications alike. During the planning phase of the work, special attention has to be paid to developing the structure of the database and the data categories. The database needs to be able to handle any subsequent changes, such as adding or removing data categories and restructuring existing ones. Section three takes the future users (both lay users and experts) into account and then focuses on the needs of a terminologist user, mostly on managing terminology projects. One key requirement to consider is the flexibility of the tool so that it can be tailored to the specific needs of the Hungarian-speaking language community, such as the possibility to set the user interface to Hungarian, or to add regional variants as new languages to the database. It must host an extensive and freely expandable list of metadata to accommodate all kinds of linguistic and terminological information, including multimedia. When considering the actual editing work, it is beneficial if the software can facilitate correct data entry with prompts, preset rules, picklists, and automated functions, such as duplicate screening. The tool should

¹ A kutatásokat a KRE BTK Terminológiai és Kommunikációs Kutatócsoport (TERMIK) keretében végeztük. A szerzők köszönik az MTA TMNP2023-1/2023. projektje támogatását, melyben az első szerző vezető tanácsadóként vesz részt.

also offer validation functions for quality assurance purposes. Given that compiling a national terminology database is a long-term project involving several terminologists and subject matter expert editors, the tool must incorporate project management functions to regulate user rights, enable assigning tasks to participants, track progress, and so on. A dashboard providing statistical information about the editing work and, later on, usage information once the national termbase is deployed is also a useful feature. Software that meets these technological requirements needs to have a secure IT background. Finally, the paper concludes with a brief summary in section four.

Keywords: Hungarian National Terminology Database, terminology strategy, database planning, structure of the database, structure of data categories

1. Bevezetés

A magyar terminológiastratégia kialakítására vonatkozó általános ajánlások között szerepelt a „*Magyar Nemzeti Terminológiai Adatbázis*² megtervezése, költségigényének, személyi és eszközfeltételeinek kidolgozása” (Fóris & Bölcskei, 2019: 157). Ezekre az előzetes ajánlásokra támaszkodva a Nyelvtudományi Kutatóközpont munkatársai az „ELKH Kiemelt pályázat SA52/2021: Terminológia” című elnyert pályázata keretében 2021 és 2023 között előkészítették a szükséges magyar terminológiastratégiai munkák elindítását (részletesen lásd a zöld könyvet és a benne foglalt tanulmányokat: Prószéky et al., 2023). A tervek megvalósítása az MTA TMNP2023-1/2023. „Magyar terminológiastratégia” című projektje keretében történik 2023 és 2027 között, ahol az egyik alprojekt célkitűzése a magyar nemzeti terminológiai adatbázis tervezése és elindítása.³

Mivel más tanulmányokban és könyvekben korábban részletesen bemutattuk és vizsgáltuk a témával kapcsolatos nemzetközi és magyar szakirodalmat, e tanulmánynak nem a szakirodalmi áttekintés a célja. A témához kapcsolódó szakirodalom kiterjed a terminológiastratégia kérdéseire, a terminológiai adatbázisok építésének részleteire (nyelvészeti, technológiai, informatikai szempontok, szabványok), és a projektmenedzsment szempontjaira is, a tervezés kérdései közül több is a folyamatok menedzselésével kapcsolatos; továbbá különböző szabványokra lehet támaszkodni. A projektmenedzsment szempontjait projektmenedzsment kézikönyvek és módszertanok tartalmazzák (pl. PMBOK Guide, Six Sigma, Scrum), emellett léteznek kifejezetten terminológiai projektekre vonatkozó szabványok is (pl. ISO 15188:2001), ezeknek az ismeretét és használatát alapvetésként kezeljük. A terminológiamenedzsment alapfogalmai és folyamatai nemzetközi szabványokban vannak szabályozva (pl. ISO 1087:2019, ISO 26162-1:2019, lásd még Sermann, 2023), továbbá több kézikönyv tárgyalja ezeket (pl. Wright & Budin, 1997, 2001; Drewer & Schmitz, 2017; Warburton, 2021). A nemzeti terminológiai adatbázisok célja különbözik a kereskedelmi célra vagy egyéni

² Az adatbázis szerzők által javasolt elnevezése: Magyar Nemzeti Terminológiai Adatbázis (MaNTA).

³ Ez a projekt az MTA „A Tudomány a Magyar Nyelvért Nemzeti Program” egyik alprogramja (l. MTA web), a HUN-REN Nyelvtudományi Kutatóközpont projektje. Ezzel a programmal az MTA elsődlegesen a magyar nyelv digitális térben való kutatását kívánja támogatni.

célra készült adatbázisoktól; ennek részleteibe nem bocsátkozunk e tanulmányban, azért sem, mert e részletek rögzítése az adatbázis tervezési fázisában a projektstratégiai dokumentumban szükséges.

Egy magyar nemzeti terminológiai adatbázis megtervezésekor, technológiai, anyagi és személyi feltételeinek felmérésekor támaszkodhatunk a már létező nemzeti terminológiai adatbázisok tapasztalataira. Az Európában működő terminológiai adatbázisok és terminológiai portálok jellemzői más írásokban olvashatók (lásd pl. Nilsson, 2017; Fóris & Bölcskei, 2019; Fóris & B. Papp, 2020; B. Papp, 2023; Tamás, 2023), az adataikat és elérhetőségüket közzétettük a zöld könyvben (Fóris et al., 2023); jelen tanulmányban kizárólag a tervezett magyar nemzeti terminológiai adatbázis technológiai hátterének vizsgálatát tűztük ki célul. Többen is megállapították, hogy a több országban használt magyar nyelvű terminológiát harmonizálni (egyesíteni) szükséges egymással, vagyis össze kell gyűjteni és közzé kell tenni a különböző országokban használatos magyar nyelvű terminusokat (pl. Tolcsvai Nagy, 2017, 2023; Benő & Péntek, 2023; Lanstyák 2023), ehhez magyar nemzeti terminológiai adatbázis elkészítésére van szükség. A magyar nemzeti terminológiai adatbázis bejegyzéseinek tervezett struktúráját Fóris & B. Papp (2020) tette közzé. Az adatbázis kezelésére alkalmas szoftveres keretrendszer kiválasztásának szempontjait egy másik tanulmányban vizsgáltuk, ugyanitt két adatbázis-kezelő szoftver, az ingyenesen elérhető *Terminologue* és a kereskedelmi forgalomban kapható *Kalcium quickTerm* előnyeit és hátrányait elemeztük és értékeltük (Fóris & Somogyi, 2024).

A magyar nemzeti terminológiai adatbázis megtervezése és elindítása nagy erőforrásokat, tervezést, előkészítést igényel. A magyar szakszókincs összegyűjtése, terminológiai elveknek megfelelő feldolgozása, egységes struktúrájú adatbázisban történő közzététele sok lépésből álló, nyelvészek és szakemberek együttműködését megkövetelő feladat. Várható haszna, hogy az adatbázis hozzájárul a magyar szakszókincs egyesítéséhez (harmonizációjához), ezzel megalapozza a későbbi egységesítési (szabványosítási) lehetőségeket, támogatja a fordítást és elősegíti a magyar szakszókincs kutatását. Jelentős alapot és segítséget nyújt a lexikára vonatkozó nyelvészeti és szaktudományos kutatások előmeneteléhez, az intézményi, vállalati, kereskedelmi célú szakszótárak és terminológiai adatbázisok készítéséhez, illetve a tankönyvek írásához.

Jelen tanulmány célja a magyar nemzeti terminológiai adatbázis tervezéséhez szükséges általános kérdések mérlegelése, elsősorban a technológiai háttér kialakításakor fontos szempontok áttekintése. A tanulmány szerkezetét ezeknek a szempontoknak megfelelően strukturáltuk. A második fejezetben az adatbázis-tervezés során tárgyalandó általános szempontokat tárgyaljuk. A harmadik fejezetben először a felhasználó szempontjait vizsgáljuk, különös tekintettel az adatbázis nyelvére és a funkcionalitásra; ezt követően a terminológus szempontjait tekintjük át; majd a terminológiai projekt technológiai leképződését,

konkrétan az adatbázis- és adattípusokat, a projekt- és feladatmenedzsment modult, valamint a statisztikai adatokat tartalmazó irányítópult szerepét; végezetül az informatikai háttértechnológiával foglalkozunk. A negyedik fejezet az összefoglalás.

Az itt bemutatott keretrendszer koncepciója és a vizsgálati szempontok összeállítása során több évtizedes, valamint a terminológiai adatbázisokkal szerzett napi szintű tapasztalatainkra támaszkodtunk, illetve a meglévő (fent hivatkozott) szakirodalmi ajánlásokra.

2. Általános megfontolások

Egy terminológiai projektben kulcsszerepet játszik annak meghatározása, hogy milyen célközönség számára készítjük az adatbázist (erről Heinisch, 2023 ír részletesen az UniVieTerm adatbázis készítése kapcsán). Egy nemzeti terminológiai adatbázis esetén egészen más, sokkal szélesebb felhasználói körrel kell terveznünk, mint más projektek esetében, pl. céges vagy intézményi terminológiai adatbázis összeállításakor. Ezek ugyan szerkezetükben és felépítésükben hasonlóak lehetnek, céljuk viszont eltérő: a nemzeti terminológiai adatbázis a korábban már meghatározott terminológiatervezési és nemzeti terminológiapolitikai törekvések megvalósulása (Infoterm, 2019: 179). A majdani felhasználók heterogenitását nemcsak a terminológiához és általában a nyelvhez való viszonyuk, de a technológiával szembeni elvárásaik, illetve az azzal kapcsolatos tudásuk is nagyban meghatározza (Sermann & Tamás, 2020). Mivel az olvasó felhasználók a hagyományos, papíralapú referenciaművekkel (szótárak, enciklopédiák, lexikonok) szemben elektronikus formában használják az adatbázist, a tervezési és később a fenntartási fázisban alapvető jelentősége van a technológiai háttérnek. A célközönség igényeit és elvárásait többféle módon lehetséges meghatározni. Egyrészt, elképzelhető egy idealizált célközönség, amelyet valamilyen tudás és attitűd alapján határozzunk meg, ez alapján szólhat az adatbázis kizárólag szakembereknek, erősítve és bizonyos szempontból kodifikálva a szakember–szakember közötti kommunikációt az adott szakterület (illetve domén) feldolgozása nyomán. Egy másik lehetőségként felmerülhet, sőt, bizonyos szintig és megkötésekkel fel is kell merülnie, hogy az adatbázis a szakember–laikus közönség közötti kommunikáció színtere is legyen. Mindennek van technológiai vonatkozása is, hiszen a laikus felhasználó nagy valószínűség szerint a felhasználói felület egyszerűségét és intuitív jellegét keresi, míg a szakértő felhasználó adott esetben igényli és hajlandó is finomhangolni a keresését, illetve a keresési találatokat további szempontok szerint is képes feldolgozni. Másrészt – a projekt anyagi és erőforrásbeli kereteitől függően – irányított, előzetes felhasználói interjúkat is lehet készíteni, amelyek során reprezentatív minta alapján, előre kidolgozott kérdésekkel lehet felmérni, hogy egy-egy alcsoportnak milyen elvárásai és előzetes tudása,

tapasztalata van általában az adatbázisokkal és kifejezetten egy nemzeti terminológiai adatbázissal kapcsolatban.

A 21. század rendkívül gyors nyelvtechnológiai fejlődésének és a mindennapi életben megjelenő alkalmazásainak köszönhetően azonban nemcsak az merül fel, hogy miként működik az adatbázis az emberi felhasználók számára, hanem az is, hogy a különböző mesterséges intelligenciára, illetve gépi tanulásra épülő rendszerek miként tudják hasznosítani az adatokat. Ugyanis jó esetben a terminológiai adatbázisok nemcsak tisztán lexikális tudást reprezentálnak és adnak át az emberi felhasználók számára, hanem egy adott domén és aldomén kiterjedt vizsgálatával a fogalmak közötti összefüggéseket is feltárják és rögzítik. Amennyiben ez a típusú indirekt ismeret a gépek számára is feldolgozható módon kinyerhető a rendszerből, akkor azt fel lehet használni olyan helyzetekben is, amikor az adott gépi modellt nem pusztán strukturálatlan adatokkal kívánják betanítani, hanem cél az adatok közötti összefüggések átadása is. Ez a szempont jól szemléltethető a nagy nyelvi modelleknél is, ahol megfigyelhető, hogy bizonyos esetekben a modell nem vagy nem pontosan képes visszaadni fogalmi összefüggéseket, ami arra vezethető vissza, hogy a modell betanítása főként nem strukturált adatokon történt, illetve az adott domén fogalmi hálójának a modell által is értelmezhető formája nem része a tanítóadatnak. A strukturált adatokkal való betanítás növeli a modell által adott válaszok pontosságát (vö. Faith, 2024). Éppen ezért a terminológiai adatbázisok jó kiindulópontot képezhetnek ontológiák és tudásgráfok kiépítésére is, ami által a nemzeti terminológiai adatbázis kialakítása és az ott felhalmozott tudás és tapasztalat közvetlenül kapcsolható más, feltörekvőben lévő kutatási területekhez is. Ennek pedig szintén kihatása van arra a technológiai környezetre, amelyben mind a terminológus felhasználók, mind pedig a humán és gépi felhasználók mozognak, hiszen ideális esetben mindenkinek ugyanabban az architektúrában kellene dolgoznia, együttműködni és a rögzített ismereteket felhasználnia.

Ez utóbbi rámutat arra, hogy mennyire lényeges az, hogy a szoftveres környezet a lehető legrugalmasabb legyen. Ha egy pillanatra elrugaszkodunk a végterméktől és közelebbről szemléljük a terminológiai munkafolyamat egyes lépéseit, akkor láthatjuk, hogy minden egyes mozzanatában fontos a flexibilitás. Előfordulhat, hogy a projekttervhez képest a konkrét munka során nem lehet egy-egy domént vagy aldomént olyan részletesen kidolgozni, mint ahogyan azt a kezdetkor eltervezték, illetve a munka során olyan új szempontok kerülnek elő, amelyekre az előkészítési fázisban nem gondoltak. Ezért lényeges, hogy a szoftveres környezet képes legyen tárolni a már elvégzett munkát, nemcsak a konkrét bejegyzések formájában, hanem olyan módon is, hogy látsszon, hogy egy-egy adatmezőben szereplő információ miért és miért úgy került be abba a mezőbe. Egyrészt makroszinten erre magyarázatot tud adni a szerkesztési kézikönyv, illetve más (akár technikai) felhasználói dokumentáció, másrészt fontos, hogy mikroszinten is lehessen metaadatokkal ellátni egy-egy döntést, ami

történhet külön erre szolgáló adatkategóriákkal (pl. megjegyzés) vagy akár a szoftver azon képességével, hogy egy adott bejegyzés „élettörténetét” is vissza lehessen követni, valamint olyan kiegészítésekkel ellátni, amelyek a végfelhasználó számára esetlegesen nem is láthatók. Ez főként akkor döntő jelentőségű, ha egy-egy terminológiai projekt több évre vagy akár több évtizedre is átnyúlik: megfelelő makro- és mikroszintű dokumentáció hiányában még egy konkrét felhasználó döntései sem követhetők vissza teljes mértékben. Az időben való elhúzódnak másik vetülete, hogy a szoftveres keretrendszernek képesnek kell lennie annak biztosítására, hogy az adatbázist egy későbbi munkafolyamatban (akár külön projekt keretén belül) bővíteni lehessen: nemcsak új adatokkal, hanem az adatbázis struktúrájának, a meglévő bejegyzések adatkategóriáinak és az ezekhez kapcsolódó logikai-szerkezeti tulajdonságainak bővítésével vagy módosításával is.

Akármilyen adatbázis kialakításánál, legyen ez terminológiai vagy más felhasználású, meghatározóak a kezdeti lépések. Egy nemzeti terminológiai adatbázist jellegéből, valamint a szerkesztő felhasználók és olvasó felhasználók heterogenitásából adódóan is a lehető legszélesebb körű felhasználásra kell kialakítani. Ezért a tervezési fázisban fel kell mérni a piacon kapható keretrendszerek képességeit, illetve azt, hogy a kezdeti lépések során meghatározott igényeket milyen módon lehet technológiailag kiszolgálni. Már ebben a korai szakaszban sem érdemes kicsiben gondolkodni, hiszen a projekt kezdetekor nem látható előre, hogy későbbi lépésekben milyen olyan új szereplők és új igények merülhetnek fel, amelyekre a technológiának válaszokat kell adnia. Ezért érdemes olyan keretrendszert választani, amely lehetőleg minden eshetőségre fel van készülve, így például képes a terminológiai bejegyzéseket hatékonyan összekapcsolni, benne a felhasználók saját maguk határozzák meg a relációk típusát, illetve mindezt az emberi és gépi felhasználók számára is feldolgozható formában tudja megjeleníteni. A bejegyzések összekapcsolása mellett persze elvárható az adatkategóriák és az adatbázis szerkezetének kellő rugalmassága is, hiszen, ha a szoftveres környezet pusztán a kezdeti elképzeléseket tudja megvalósítani, akkor előfordulhat, hogy elvesz a későbbi bővíthetőség lehetősége. Ugyanakkor, ha mégis lehetőség olyan rendszer használatára, amely kellően rugalmas lenne, akkor nagyobb hangsúlyt kell fektetni az adatok rendszerek közötti kompatibilitására, vagyis arra, hogy az egyik rendszerben létrehozott adatot a másik rendszerben újra lehessen hasznosítani, ha úgy adódna, hogy a szoftveres környezet változna a későbbi fázisokban a kezdethez képest. Ennek számos megvalósulási módja lehet, az egyik legelterjedtebb és legegyszerűbb az adatbázis olyan fájlformátumba való exportálása, amelyet egy másik, szintén terminológiai adatok kezelésére létrehozott rendszer kezelni képes. Ilyen fájlformátum lehet például a TBX (lásd ISO 30042:2019).

3. Technológiai követelmények

Ahogy arról a korábbiakban is szó volt, a terminológiai projekt elindulásakor az előkészítési fázis egyik kiemelt lépése a **technológiai háttér meghatározása** (Drewer & Schmitz, 2017: 140). Fontos több szempontot megvizsgálni, figyelembe venni, mert egy adott rendszer melletti elköteleződés alapvetően meghatározza a szerkesztési munkát, valamint az adatbázis használhatóságát. Az alábbi szempontok egyfajta ellenőrzőlistaként is funkcionálhatnak mindazok számára, akik intézményként egy terminológiai projekt keretén belül terminológiai adatbázis bevezetését tervezik.

3.1. A felhasználó szempontjai

3.1.1. Az adatbázis nyelve

A felhasználóknak az adatbázisról szerzett első benyomása nagymértékben kihat a későbbi felhasználói élményre. Ennek a tapasztalásnak lehet vizuális vetülete (mennyire letisztult a felhasználói felület stílusa, milyen a színkezelés, megtalálja-e elsőre a keresési mezőt), de talán még fontosabb a nyelvi aspektus, azaz, hogy milyen nyelven használhatja az adatbázist. Egy nemzeti terminológiai adatbázisnál elengedhetetlen, hogy a felhasználó a saját anyanyelvén tudja használni az adatbázist, illetve megjelenjen a saját nyelvváltozata is, amennyiben az az adott nyelvközösségben meghatározó (pl. a környező országokban beszélt magyar nyelvváltozatok). Ez a szempont technológiai szinten ott jelenik meg, hogy a terminológiai keretrendszer képes-e kezelni a nemzeti adatbázis fő nyelvét (ez általában nem probléma, hiszen a piacon elérhető rendszerekben az elérhető nyelvek köre kiterjedt), valamint van-e lehetőség további, a beépített listától eltérő nyelveket, illetve nyelvváltozatokat létrehozni és megfelelő módon megjeleníteni. Ugyanakkor a szerkesztési módszerek, az adatbázis felépítése és az adatmezőkben lévő információk a nemzetközi közönség számára is érdekesek vagy relevánsak lehetnek, így ideális esetben az adatbázisnak legalább angolul (illetve angol közvetítő nyelven) is elérhetőnek kell lennie.

A nyelvi kérdés nemcsak a felhasználói felületben, vagyis a gombok, keresési beállítások, szűrők tekintetében jelenik meg, hanem az adatkategóriák nyelvében is. Így tehát egy ideális szoftveres környezetben, ha a felhasználó nyelvet vált, akkor nemcsak a felhasználói felülethez tartozó szövegek, de maguk az adatkategóriák szövegei is megváltoznak. Mindez a konkrét adatmező tartalmától is független: például, még ha maga a definíció csak magyarul szerepel, akkor is lehetősége van egy angol nyelvű felhasználónak arra, hogy lássa, hogy ez milyen adatkategória. A felhasználói felület és az adatkategóriák két- vagy többnyelvűségének mindenképpen érdemes kiterjedni a felhasználás módját és mikéntjét taglaló súgószövegekre, hiszen a rendszer összetettségétől függően minden felhasználónak szüksége lehet arra, hogy tájékozódjon arról, hogy pontosan miként tudja megjeleníteni a kívánt keresési eredményt és azt miként értelmezze, és milyen további beállításokat tud tenni.

3.1.2. Funkcionalitás – felhasználói szempontok

A vizuális és nyelvi megjelenésen túl hangsúlyos szerepe van a funkcionalitásnak. Itt az elsődleges kérdés, hogy mennyire intuitív az adatbázis használata. Ez a kérdés nagyban függ a szoftveres környezet adta lehetőségektől, hiszen, ha az csupán egyszerű, szöveges keresések végrehajtására képes, akkor a felhasználónak sem esik nehezére átlátni a lehetőségeket. Egyszerű keresés alatt adott adatkategóriára (leginkább a terminusra) való keresés értendő, teljes szóalak szerint (teljes egyezés) vagy szótöredékek szerint (részleges egyezés). Viszont, ha összetettebb keresésre is képes a rendszer, akkor lényeges, hogy a beállítások milyen formában jelennek meg a végfelhasználó számára. Összetett keresés alatt elsősorban olyan finomhangolt beállításokat értünk, mint bizonyos adatkategóriák mentén való szűrés lehetősége, keresési operátorok használata: úgymint „minden, ami tartalmazza az alábbi”, ennek ellenkezője: „minden, ami nem tartalmazza az alábbi”, ÉS, illetve VAGY típusú kapcsolatok (ami magában foglalja a keresési feltételek összefűzhetőségét és egymással való szabad kombinációját), a szavak elejének vagy végének kereshetőségét, valamint szabadon meghatározott szöveges mintázatok keresését reguláris kifejezésekkel (Drewer & Schmitz, 2017: 169). Mindezeket pedig ideális esetben akár menteni is lehet későbbi újrafelhasználás céljából. Akármilyen összetettséggű is a rendszer, a jó felhasználói dokumentáció (akár szöveges, de még jobb a rövid videós forma) ebben is tud segíteni. Ezen a ponton azonban érdemes elkülöníteni az **olvasó felhasználó** (akiket még tovább oszthatunk laikus és szakértő felhasználókra) és a **szerkesztő felhasználó** által támasztott keresési igényeket. Alapvetően elmondható, hogy utóbbi felhasználói csoportnak legalább azokat a kereséseket le kell tudnia futtatnia, mint a végfelhasználónak, és ugyanolyan módon, hozzáteve, hogy az eltérő igények miatt fontos, hogy a szoftveres környezetben mindkét oldalon pontosan lehessen szabályozni, hogy ki, mit és milyen módon tud keresni. Például a végfelhasználó szempontjából nem érdekes, hogy az elmúlt egy héten pontosan mely felhasználók módosítottak akármilyen bejegyzést vagy hogy egy bejegyzésnek milyen munkafolyamatbeli állapota van, míg a szerkesztő felhasználónak szüksége lehet ezekre az információkra.

Utóbbi szemponttal szorosán összefügg annak a kérdése is, hogy melyik adatkategória releváns az olvasó felhasználó és melyik a szerkesztő felhasználó számára. Ez szintén visszavezethető a felhasználók igényeire, elvárásaira és technológiai képzettségére is. Ezért alapvető kérdés annak meghatározása is, hogy az idealizált (vagy interjúkon keresztül reprezentatív módon felmért) felhasználói csoport mennyi adatot tart soknak vagy kevésnek. Alapvetően elmondható, hogy a laikus felhasználó esetén a kevesebb több, ezért a terminológiai adatbázis struktúrájának és felépítésének meghatározásakor arra is ki kell térni, hogy az adott információhalmazból mit kell látnia a laikus felhasználónak, a szakértő felhasználónak és a szerkesztő felhasználónak (Drewer & Schmitz, 2017: 152). Az igazán jó szoftveres környezet e tekintetben

is rugalmas, vagyis lehetőséget biztosít arra, hogy a felhasználó maga döntse el, hogy mennyi adatot lát. Ezt technológiailag akár felhasználói profilokkal, akár szerkezeti szinten, egymásból lenyíló táblákkal lehet megoldani. A fő szempont az, hogy valamilyen választási lehetőséget kínáljunk a felhasználók számára, illetve a különböző csoportok – a szoftver korlátai miatt – ne csak ugyanazt legyenek képesek látni.

Fontos, hogy a végfelhasználói élmény szempontjából lehetőleg ne csak száraz keresést és szűrést lehessen megvalósítani, hanem lehetséges legyen „barangolni” a bejegyzések között. Ezt technológiailag a bejegyzések összekötésével lehet megvalósítani, ami történhet hiperhivatkozásokon keresztül, így a végfelhasználó a Wikipédiához hasonlatosan azon kaphatja magát, hogy az egymásba kapcsolódó bejegyzéseken keresztül „bejárta” az adott domént vagy éppen áttévedt más doménokra is. Ennek még látványosabb megoldása pedig a már említett gráfos megjelenítés, ahol a bejegyzések közötti kapcsolatok hálózatos formában vannak ábrázolva, így a felhasználó nemcsak pusztán a kapcsolat tényét látja maga előtt, hanem a kapcsolat típusát is, így téve szert extra információra, amit a megfelelő vizualizációs dizájn mélyít el még jobban.

A passzív felhasználói élményen túl a célközönség bevonása is cél lehet. Az adatbázis kialakításakor, illetve a keretrendszer kiválasztásakor mérlegelni kell a lehetőségét annak, hogy a végfelhasználó képes legyen-e visszajelzést adni az adatbázis tartalmáról. Az adatbázis népszerűségét, ismertségét és jó hírnevének elterjedését nagyban segítheti, ha erre van lehetőség, hiszen akkor a végfelhasználó úgy érezheti, hogy maga is része az adatbázis bővítésének, elmélyítésének, így végső soron bizonyos keretek között egyfajta kollaboratív terminológiai munka is megvalósulhat. Technológiai szempontból ez egyszerűen és összetetten is megoldható: előbbi esetében csupán egy központi e-mail-cím megadása szükséges, ahol visszajelzést lehet küldeni, amire valaki válaszol, utóbbi esetében viszont a szoftveres keretrendszer lehetőséget adhat, hogy adott bejegyzésnél, pontosan meghatározható módon lehessen visszajelzést adni, így a fogadó fél oldalán rögtön látható, hogy melyik bejegyzésről van szó, sőt, ha erre a rendszer ad lehetőséget, a háttérben még azt is meg lehet adni, hogy ne egy közös címre érkezzon a visszajelzés, hanem ahhoz a konkrét terminológushoz, szakértőhöz vagy egy kijelölt (al)csoporthoz, akik az adott bejegyzésen dolgoztak.

Az aktív felhasználói élmény másik aspektusa az **adatok letölthetősége**, illetve az adatbázisban található **adatok újrafelhasználhatóságának lehetősége**. Egy nemzeti terminológiai adatbázisnak az egyik legfontosabb jellemzője, hogy bárki számára elérhető ingyenesen. Ideális esetben a felhasználók számára az is lehetséges, hogy keresési eredményeiket az általuk preferált formátumban tudják letölteni, természetesen tiszteletben tartva a vonatkozó szerzői jogi megkötéseket. A felhasználói csoport heterogenitására jó bizonyíték, hogy ezeket a letöltött formátumokat akár fordítók is hasznosíthatják munkájuk során, vagy akár

egyetemi tananyagok kiegészítéseként is fel lehet használni. Ha ezt a lehetőséget is biztosítani szeretnénk, akkor elvárható, hogy a szoftveres keretrendszer mind a gépi, mind pedig az emberi felhasználású formátumokat támogassa, így biztosítva a felhasználási kör lehető legnagyobb szélességét.

A felhasználók szempontjának semmi esetre sem elhanyagolható pontja az **akadálymentesség** kérdése. A felhasználói igények és a technológiai lehetőségek ezen a ponton is dominánsak, illetve erre a kérdésre is ki kell, hogy terjedjenek. Ezen a területen is számos nemzetközi szabvány született (pl. WCAG, 2008), amely szabályozza, hogy milyen technikai feltételeknek kell teljesülniük, hogy az adott rendszert a látás- és hallássérült felhasználók is gond nélkül tudják használni, illetve ajánlásokat ad ki ezekkel kapcsolatban. A siketek vonatkozásában nem is pusztán csak a felhasználói felület megjelenése a meghatározó, hanem az is, hogy a szoftveres környezet képes legyen olyan multimédiás adatkategória-típust kezelni, ahol a terminus jelnyelvi változatát is meg lehet jeleníteni. Ez történhet extra rövid videós formátumokkal, GIF formátummal, illetve képpel a jelnyelvi jelek megjelenítésére.

3.2. A terminológus szempontjai

A végfelhasználó igényeivel szembeállítva, de ezzel együtt, annak kiegészítéseként érdemes meghatározni a projekt elején, hogy milyen szempontok relevánsak a terminológus számára. Az adatbázis építőelemeit adó adatkategóriák tekintetében az észszerűségekre kell törekedni, vagyis nem érdemes se túl kevés, se túl sok adatkategóriát meghatározni. A tapasztalatok azt mutatják (Fóris & B. Papp, 2020), hogy nem érdemes sokat meghatározni, mert a szerkesztők nem töltik ki mindet, illetve az adott doméntól és aldoméntól függően nem is biztos, hogy minden adatkategória releváns. Egészen más jellegű adatkategóriák lehetnek fontosak és hasznosak egy jogi adatbázisban, mint egy orvosiban (még ha lehetnek is kapcsolódási pontok). A nemzeti terminológiai adatbázis feltöltése során, ahogy újabb és újabb szakterületek terminológiája kerül feldolgozásra, egyre több adatkategória hozzáadása válhat szükségessé. Ezt látjuk például a horvát Struna adatbázisban (<http://struna.ihjj.hr/en/>), amely 2013-ban 23 adatkategóriát tartalmazott (Bratanić & Ostroški Anić, 2013: 678), 2023-ban viszont már 46 adatmezővel dolgoztak (de ezek közül csak a terminus és definíciója, valamint a doménbe való besorolás kötelezően kitöltendő, továbbá néhány adminisztratív kategória automatikusan rögzít adatokat). Éppen ezért az előkészületeknél meg kell határozni egy minimális adatkategória-listát (lásd a DatCatInfo erre vonatkozó ajánlásait és az adatkategóriákat rendszerező adatbázist), amelyet minden szakterülettől függetlenül alapként lehet kezelni és amihez majd az adott doménnek megfelelően további adatkategóriákat lehet hozzárendelni.

Technikai szempontból a szoftveres környezetnek tudnia kell kezelni ezt a potenciális sokszínűséget, ami megjelenhet egyrészt olyan módon, hogy a

szoftveres környezetben belül aladatbázisokat hozunk létre szakterületenként (vagy akár a kidolgozás mögött álló csoportonként, intézményenként, szervezetenként), hozzátevé, hogy a végfelhasználónak biztosítani kell azt a lehetőséget, hogy minden egyes kereséskor pontosan tudja, hogy melyik aladatbázisban keres, és be tudja állítani, hogy melyikben akar és melyikben nem. Hátránya a több aladatbázist használó megközelítésnek, hogy ha egy terminus több doménban is szerepel, ezekben más-más információkat tartalmazhat. Ez bizonyos esetekben még akár indokolt is lehet, hiszen a fogalmak meghatározó jegyei különböző doménekben más hangsúlyt kapnak, ami visszatükröződhet a szakértők szemléletében és a definíció megfogalmazásában is. Az leszögezhető, hogy nem lehet univerzalitásra törekedni, hiszen a fogalomrendszerek és maga a nyelv sem univerzalizálható. Éppen ezért kulcskérdés, hogy a szoftveres környezet lehetővé tegye azt, hogy ha valamilyen okból kifolyólag az adatkategóriákat vagy azok értékeit (pl. doménba vagy aldoménba sorolás) újra kell gondolni az előkészítési fázis után, akkor azt a rendszer úgy tudja kezelni, hogy minimalizálja a manuális munkát és az adatvesztést, illetve lehetőséget adjon az adatbázis minőségének biztosításával kapcsolatos feladatokra, úgymint duplikátumok szűrése, bejegyzések összevonása vagy szétválasztása.

A végfelhasználó mellett a terminológus és a doménszakértő felhasználói élménye is mérvadó. Ez a gyakorlatban azt jelenti, hogy a szoftveres környezetet kellően intuitív módon lehessen használni, hiszen a végfelhasználói csoportok tudásbeli és tapasztalatbeli sokszínűsége a másik oldalra is jellemző. Fontos szempont, hogy a rendszer automatizációs megoldásokkal segítse a terminológust és a doménszakértőt a szerkesztéssel kapcsolatos fáradságos manuális munka elkerülésében, valamint biztosítsa a szabályozhatóságot, vagyis, hogy az egyéni túl csoportos szinten is ugyanolyan módon, ugyanolyan hatékonysággal tudják végezni a munkájukat. Ennek egyik megnyilvánulási formája lehet, hogy az adatmezők kitöltésére vonatkozó szabályokat ne kelljen mindig fejben tartani, hanem a szoftver segítse a pontos és helyes adatbevitelt, például hibaüzenetekkel, amennyiben nem az előre meghatározott formátumú vagy tartalmú, értékű adat kerül a mezőbe.

3.3. A terminológiai projekt technológiai leképeződése

A terminológiai munkafolyamat lényegében **projektmunka**. Jó esetben terminológusok, doménszakértők, nyelvi szerkesztők, technikai szakértők és projektvezetők vesznek benne részt. Az ideális szoftveres keretrendszernek lehetőséget kell biztosítania arra, hogy a különböző felhasználókat megfelelő módon lehessen csoportba rendezni, elkülöníteni, illetve az általuk végzett munkát kiosztani, amit ők a rendszeren belül el is tudnak végezni, illetve követni lehessen a projekt előrehaladását, és mindezt számszerűsíteni is képes legyen a szoftver (Drewer & Schmitz, 2017: 158). Az alábbiakban annak ismertetése

következik, hogy az ideális keretrendszer milyen olyan modulokból áll, amelyek a fentieket támogatni képesek.

3.3.1. Az adatbázis és az adattípusok

A terminológiai keretrendszer központi eleme maga az **adatbázis**, amelyben a terminológus és doménszakértő felhasználók dolgoznak és amelyben a végfelhasználó keres. A már sokat emlegetett adatkategóriákon és azok egymáshoz képesti viszonyán túl megvizsgálandó az is, hogy a szoftveres környezet milyen lehetőséget biztosít az adatkategóriákhoz rendelt adatmezőkben lévő adattípusok kezelésére.

Az első és legfontosabb adattípus a **szöveg**. Ennek kapcsán érdemes megjegyezni, hogy ideális esetben a szoftveres környezet nem csupán sima szöveg bevitelére alkalmas, hanem lehetőséget biztosít formázások alkalmazására is. Egyes szakterületeknél elengedhetetlen a szövegszintű formázás, kiemelendő példa a matematikában az alsó és felső indexek megfelelő kezelése vagy akár bizonyos karakterek dőlttel szedése a kémiai nomenklatúrában. Ezenfelül nélkülözhetetlen a megfelelő karakterkódolás: az ideális szoftveres környezetnek legalább az UTF-8-as karakterkészletet kell támogatnia, felkészítve az adatbázist arra, hogy a különböző nyelvek írásrendszereinek karaktereit megfelelő módon tudja kezelni.

A következő adattípus a **hiperhivatkozás**. Ez nemcsak a korábbiakban említett, bejegyzések közötti kapcsolódás egyik technikai megvalósítója, de ez biztosítja az adatbázis összekapcsolódását a külvilággal, vagyis az internettel. Ebben az esetben nem is biztos, hogy a szabad internetre kell gondolni, hanem ezalatt lehet olyan egyéb nyelvi rendszereket is érteni, mint például a Magyar Nemzeti Szövegtár vagy egy szakszöveg-adatbázis, azaz kívánatos, hogy közvetlenül a terminológiai adatbázisból lehessen benne keresést futtatni az adott terminusra.

További adattípusként említhetők még a **számok**, illetve a **dátumok**. Utóbbi már az adatbázis azon részeihez tartozik, amelyet ideális esetben nem a felhasználónak kell kézzel bevinnie, hanem a rendszer automatikusan generálja. Mind a dátum, mind pedig a felhasználó neve mint automatikus érték kulcsfontosságú az adott bejegyzés élettörténetének nyomon követhetőségében, éppen ezért nem ajánlott a felhasználóra bízni, hogy új bejegyzés létrehozásakor, illetve egy már meglévő módosításakor beviszi-e ezeket a metaadatokat. A visszakövethetőség egyik következő lépése, amire az ideális keretrendszer szintén lehetőséget biztosít, hogy nem csupán meg lehet nézni, mi volt a változás, hanem könnyedén vissza is lehet állítani az előző állapotot. Erre akkor lehet szükség, ha valamilyen tömeges művelet előre nem látható hibát eredményezett, így nem kézzel kell visszaállítani az előző állapotot, illetve a funkció hiányában nem kell adatvesztéssel szembenézni. A felhasználó és a dátum metaadatain túl itt még hasznos a további megjegyzés hozzáfűzésének lehetősége is, hiszen akár

évek múltán visszatekintve már nem biztos, hogy egyértelmű, hogy egy-egy változásnak mi volt az indoka.

A pusztá adatbevitelen túl különösen hasznos az egyes bejegyzések különböző rendszerek és logika szerinti osztályozása, csoportosítása. Az adattípusok szintjén ezt az előre meghatározott listaértékek biztosítják. Ez elsősorban a doménba és aldoménba való besorolásnál jelentős. E tekintetben is van szerepe az ideális keretrendszernek: jó, ha lehetőséget tud biztosítani az egyes fölé- és alárendelt kategóriák közötti viszony meghatározására és leképezésére. A gyakorlatban ez azt jelenti, hogy meg lehet határozni, hogy adott kategóriához milyen lehetséges alkategóriák tartoznak, és a rendszer csakis azokat engedni bevinni. Ez egyrészt biztosítja, hogy a szerkesztő felhasználó véletlenül se párosítson olyan kategóriát olyan alkategóriával, ami eltér az előre meghatározottól, továbbá átláthatóvá teszi a kategória–alkategória-rendszert. Egy nagyobb adatbázisnál, ami nem is feltétlenül terjed ki több doménra, a kategóriarendszer olyan komplexitású lehet, hogy egy idő után mind a szerkesztő felhasználó, mind pedig az olvasó felhasználó számára átláthatatlanná válik, ezért praktikus a korlátozás és a függőségi viszony meghatározásának lehetősége. A felhasználói oldalról a kategóriákat úgy lehet segíteni átlátni az ideális keretrendszerben, hogy lehetőséget biztosítunk a fastruktúrába rendezett kategóriák és alkategóriák közötti navigációra, így megvalósítva a facettás keresési módot, ahol egymásból lenyíló listák segítik az olvasó felhasználót a tájékozódásban, illetve a keresési találatok szűrésében.

Olyan egyéb adattípusokra is szükség lehet, mint a **logikai értékek** (igaz vagy hamis), amely a felhasználó számára egy jelölőnégyzet formájában jelenthet meg és azt jelzi, hogy egy adott adatkategória érvényes vagy nem érvényes az adott bejegyzésre, vagy akár a **multimédiás adattípus**, amely az adott bejegyzés ismeretanyagát szemlélteti vagy egészíti ki képpel, videóval vagy hanggal.

Akármiilyen adatkategóriákkal, azon belül adatmezőkkel és adattípusokkal dolgozunk, a tervezési fázis során meg kell határozni, hogy ezek közül melyek kötelezőek. Ez elsősorban a szerkesztő felhasználót érinti, aki a konkrét adattal foglalkozik. A kötelezően kitöltendő adatkategóriák megléte vagy hiánya (valamint az abban található adatok milyensége) jelenti a választóvonalat ahhoz, hogy egy adatbázist egyáltalán terminológiai adatbázisnak lehet-e nevezni és hogy milyen típusúnak (Sermann & Tamás, 2020). A korábban említett minimális adatkategória-listával kapcsolatban evidens lehet azt gondolni, hogy ezek legyenek a kötelező elemek, így lehet biztosítani, hogy doméntól függetlenül minden egyes bejegyzésben ugyanazok az adatkategóriák szerepeljenek.

A kötelezőség, nem kötelezőség kérdése elvezet a következő nélkülözhetetlen szemponthoz, amelyet az ideális keretrendszernek tudnia kell kezelni, ez pedig a **validáció** kérdése (Drewer & Schmitz, 2017: 180). Ez alatt olyan mechanizmust kell érteni, amely biztosítja, hogy az egyes adatok az előzetesen kidolgozott alapelveknek feleljenek meg, és mindehhez valamilyen technikai segítséget

nyújt. Ebben egyrészt szerepet játszanak maguk az adattípusok is, hiszen ideális esetben egy olyan adatmezőbe, ahol csak dátum szerepelhet, a rendszer már nem is enged más típusú adatot bevinni, például szöveget vagy képet. De ilyen mechanizmus lehet a már korábban említett domén–aldomén-elrendezés is, amennyiben a rendszer nem engedélyezi az olyan kombinációkat, amelyek eltérnek az előre meghatározottaktól. A validáció egyébként a szoftveres keretrendszerrel függően lehet emberi szerkesztői feladat is, de ha lehet, meg kell előzni a problémát már az adatbevitel során. Amennyiben a szoftveres környezet a konkrét beviteli validációt nem is tudja kezelni, akkor legalább lehessen a defektív kategóriákra, illetve jelenségekre előre definiált szűrőket létrehozni és időről-időre egyfajta nagytakarítást végezni. **Az adatbevitel pontossága, illetve az alapelveknek való megfelelés sokszereplős terminológiai projekteknél kulcskérdés.** Ahol sok ember dolgozik, ott hiába van kidolgozva bármilyen szerkesztési elv, mindig lesznek eltérések, ezért ennek detektálását és orvoslását a technológiai eszközökkel kell tudni támogatni.

Az adatbázis építőelemein túl fontos annak meghatározása is, hogy az elemek egymáshoz képest milyen viszonyban vannak, ez maga az adatbázis struktúrájának meghatározása. A gyakorlatban ez azt jelenti, hogy **el kell dönteni a projekt kezdetekor, hogy mely adatkategóriák kerüljenek a bejegyzés szintjére, a nyelvi indexálás szintjére, illetve a terminus szintjére** (Drewer & Schmitz, 2017: 132). Technológiai oldalról lényeges tényező lehet, hogy a keretrendszer biztosít-e lehetőséget ennek a struktúrájának a módosítására akkor, amikor már van valamennyi adat az adatbázisban. Az előkészítési fázisban nem lehet mindenre gondolni, ezért az ideális rendszer képes kezelni akár még a radikálisabb irányváltásokat is anélkül, hogy ez nagy mennyiségű manuális munkával vagy az adatvesztés veszélyével fenyegetne. Az adatstruktúra mellett az adatkategóriák közötti függőségi viszony szabályozása is fontos lehet az ideális rendszerben. A gyakorlatban ez azt jelenti, hogy meg lehet határozni, hogy ha a felhasználó képet tölt fel az egyik erre kialakított adatmezőbe, akkor a kép forrását rögzítő adatmező automatikusan megjelenjen és csak akkor legyen kötelező, ha van az előző adatmezőben valamilyen adat.

3.3.2. Projekt- és feladatmenedzsment modul

Az ideális terminológiai keretrendszernek nem csupán magát az adatbázist kell tartalmaznia, hanem olyan funkciókat is biztosítania kell, amellyel technológiailag támogatni lehet a teljes projektfolyamatot (Drewer & Schmitz, 2017: 158). Minél több szereplő vesz részt a terminológiai adatbázis szerkesztésében, annál nehezebb nyomon követni, hogy ki, mit, milyen módon és milyen határidőre kell, hogy befejezzon. A követhetőség hiánya nemcsak mikroszinten jelentkezik, hanem kihatással van a projekt egészére: különálló rendszerek nélkül nem lehet meghatározni, hogy adott időpillanatban hol tart a projekt, milyen mérföldkövek teljesültek, és milyen messze van még a kitűzött

cél. Ennek hosszú távon az is a következménye lehet, hogy a projekt közben nem lehet irányt váltani, vagy más módon módosítani az előzetes tervekhez képest.

A projektben részt vevő egyes szereplők technológiai vetülete kapcsán előtérbe kerül, hogy az adott rendszer milyen mértékben képes a felhasználó- és jogosultságkezelésre. A felhasználókezelés elsősorban abban jelenik meg, hogy az új felhasználók miként kerülnek be a rendszerbe. Tisztázni kell, hogy milyen módon lehet szerkesztő felhasználóként és olvasó felhasználóként regisztrálni. Utóbbi felhasználói csoport esetén talán kerülendő a regisztrálási kötelezettség, hiszen az esetleg eltántoríthatja az olvasót a rendszer használatától. Ennek a folyamatnak a komplexitása az adott keretrendszerhez mérten különböző lehet.

A jogosultságkezelés vonatkozásában, alapvetően minden rendszer esetén elmondható, hogy nem ideális, ha minden felhasználónak a lehető legszélesebb jogköre van. Ennek gyakorlati jelentősége az, hogy a jogkörök korlátozásával elkerülhetővé válik az esetleges véletlen adattörlés vagy az, hogy egy adott csoporttag olyan bejegyzéseket legyen képes szerkeszteni, amihez egyébként nincs kompetenciája, illetve korlátozhatók a nem kívánt, több bejegyzésre kiterjedő tömeges szerkesztési jogok, mint például a tömeges törlés, exportálás vagy importálás. Ugyanez a szemlélet jelenik meg a feladatkezelésben is az egyéni felhasználók szintjén: az ideális rendszerben jól körvonalazhatóan el kell tudni különíteni, hogy milyen feladatok vannak a felhasználóhoz rendelve, mivel kell foglalkoznia és miként tudja jelezni, ha készen van.

Felmerülhet az a kérdés, hogy miért is kellene ilyen komplex, dedikált rendszert erre fenntartani, amikor a kapcsolattartás céljára ki lehet használni a már meglévő technológiai lehetőséget, konkrétan az e-mailt, illetve más üzenetküldő alkalmazásokat, valamint az adatbázis táblaszerű jellegéből adódóan az Excelt mint kiegészítő szerkesztési felületet. Habár a terminológiai munkában rendszeresen használjuk az Excelt (vagy más táblázatkezelő alkalmazásokat) ideiglenes adatgyűjtésre, de tapasztalataink alapján a résztvevők, bejegyzések és adatmezők bizonyos száma felett hamar átláthatatlanná válnak a folyamatok, valamint nagy a hibázási lehetőség. Egy nemzeti adatbázis esetén különösen fontos az adatok megfelelő kezelhetősége, azért is, mert – egy cégben folyó munkával szemben – a bedolgozó csoportok összetétele nagyon különböző lehet, és rövid időn belül változhat. A keretrendszernek biztosítania kell azt, hogy mindenki akár egyszerre dolgozhasson az adatbázisban és ki tudja használni a rendszer adta lehetőségeket, amikre a különböző táblázatkezelő programok nem alkalmasak, hiszen nem elsősorban szövegszerkesztésre lettek kitalálva. Továbbá az Excelben történő munkavégzéskor haladó technikai tudást igényel a magukat a fájlokat és a fájlon belül bizonyos lényeges tartalmi részeket írásvédetté tenni, megakadályozva a felhasználókat a szándékolt vagy szándékolatlan szerkesztési vagy fájlműveletek elvégzésében. Ezen alkalmazások használata magában hordozza a felesleges munka, sőt az adatvesztés kockázatát is; sok felhasználó esetén nagyon nehézkes bennük a verziókövetés, ezen kívül hiányoznak belőlük

a projektvezetési és validálási funkciók. Egyszerű táblázatkezelő rendszer helyett érdemes olyan keretrendszert választani, amely képes többdimenziós referenciák kezelésére és megfelelően hatékony.

A projekt-, illetve feladatkezelő modulnak le kell tudnia követni a terminológiai projekt minden lépését. Mindamellet, hogy más projekthez hasonlóan a terminológiai jellegűek is nagyon sokszínűek lehetnek, meg lehet határozni bizonyos lépéseket, ahol a technológiának is nagy szerepe van. Az adott domén feltérképezése és felmérése nyomán az azonosított fogalmakat jelölő terminusok gyűjtésével kezdődik a munka, ami történhet megfelelő szövegekből való kivonatolással, illetve az adott szakterület szakértőjének bevonásával. Ideális esetben ezek valóban terminusok és nemcsak terminusjelöltek, illetve a gyűjtés során már lehet azonosítani olyan adatokat, amelyeket aztán az adatbázisban is lehet rögzíteni. Ebben a kezdeti fázisban nem biztos, hogy rögtön a szoftveres környezetben kell dolgozni, hiszen a terminuslista és néhány adat rögzítésére más eszközök is alkalmasak. Ehhez kapcsolódik az az elvárás, hogy a későbbiekben lehetőség legyen ennek az előzetes listának az importálására a szoftveres keretrendszerbe további szerkesztés céljából. Ennek kapcsán felmerül az a technológiai igény a keretrendszerrel kapcsolatban, hogy lehessen bele importálni különböző fájlformátumokból, illetve meg lehessen határozni, hogy a kiinduló fájl egyes részei az adatbázis mely részeibe kerüljenek (adatleképezés).

Az importálást követően a munkafolyamat következő lépése annak meghatározása, hogy ki dolgozza ki a bejegyzéseket. A kiinduló lista méretétől, összetettségétől és szakterületi sokszínűségétől függően ez egy vagy több embert jelent. A szoftveres keretrendszerrel szemben támasztott követelmények itt is jelentős szerepet játszanak: lehetővé kell tennie, hogy előre meghatározott felhasználói csoportokhoz lehessen feladatokat rendelni, sőt akár a csoporton belül egyénekhez is. A munkafolyamat következő lépése maga a kidolgozás, illetve a gépi és humán validáció, amelyekhez kapcsolódóan az előbbiekben már ismertettük a legfontosabb technikai szempontokat és követelményeket.

Az egyes feladatok közötti mozgást a rendszernek tudnia kell követni nem pusztán felülnézetből, vagyis, hogy egy adott bejegyzés éppen hol tart, hanem ezeket a lépéseket az adatkategóriák mentén is meg kell tudnia jeleníteni. A legkézenfekvőbb technikai megoldás erre egy **státusz** (vagy **állapot**) nevű adatkategória létrehozása, amelynek előre definiált elemei jelzik, hogy a bejegyzés éppen melyik fázisban tart. Ennek az adatnak a bejegyzéshez való rendelése ideális esetben automatikusan történik, azaz, amikor a munkafolyamat átlép az egyik fázisból a másikba, akkor a rendszer magától megváltoztatja a státuszra vonatkozó adatkategória értékét. Ez a funkció két további lehetőséget is biztosít: egyrészt, mérhetővé teszi, hogy az adatbázisban adott időpillanatban mennyi munka van folyamatban, melyek azok a bejegyzések, amelyek még jóváhagyásra várnak és hány van belőlük, másrészt pedig képesek szabályozni, hogy a szerkesztési rendszerből pontosan mi is kerüljön ki a végfelhasználók által

is látható rendszerbe, amennyiben a munkafolyamat utolsó lépése után a bejegyzés jóváhagyott státuszt kap és automatikusan elérhető lesz az olvasó számára.

A projekt- és feladatkezelés, valamint az adatbázis szemléletének határán vannak a tömeges szerkesztés lehetőségeinek technikai feltételei. Ahogy arról korábban szó volt, előfordulhat, hogy a projekt során valamilyen szemléletbeli változtatásra van szükség, aminek az lesz a technológiai vonatkozása, hogy át kell strukturálni az adatkategóriákat, újakat kell hozzáadni, át kell nevezni, vagy törölni kell őket. Különösen igaz lehet ez az olyan adatkategóriákra (pl. domén, aldómén), ahol a munka kezdetén bizonyos számú elemet meg kell határozni értéklisák formájában. Ennek nyomán előfordulhat, hogy nagy mennyiségű, már meglévő adatot kell módosítani a bejegyzésekben. Ideális esetben a szoftveres keretrendszer biztosít olyan mechanizmusokat és funkciókat, amellyel ezt nem manuálisan, hanem automatizált módon lehet elvégezni. A legfontosabb ilyen funkció a tömeges keresés és csere lehetősége, amelynek működése nagyban összefügg az alap keresési lehetőségekkel. A csere kapcsán nem pusztán szöveges változtatásokra kell gondolni, hanem akár átkategorizálásokra is, vagyis a cserének minden adatkategóriára és adattípusra ki kell terjednie ideális esetben. Ehhez kapcsolódik még további ellenőrzési munkafolyamatok megvalósításának lehetősége. Ilyen például az adatok konzisztenciájának ellenőrzése, ami megjelenik a duplikátumok szűrésének és kezelésének lehetőségénél vagy akár olyan esetekben is, amikor van egy terminusjelölt-listám és ellenőrizni szeretném, hogy ezekből mi szerepel már az adatbázisban. Ez utóbbi nagy számú bejegyzéseknél elkerülhetetlen igényként léphet fel.

3.3.3. Statisztikai adatokat tartalmazó irányítópult

Ezen igény felmerülése és technológiai kiszolgálásának lehetősége a terminológiai adatbázis sajátos jellegéből és a korábban említett munkafolyamatok nyomon követhetősége kapcsán jelenik meg. A megfelelő szerkesztési elvek betartása (pl. granularitás, azaz egy adatmezőben csak egy adat szerepel [vö. ISO 26162-1:2019]) lehetővé teszi ezek számszerűsítését is. Ennek nyomán az ideális szoftveres keretrendszernek rendelkeznie kell számítási funkcióval is. Ez a számítás egyrészt kiterjedhet mennyiségi kérdésekre is, azaz, hogy a lekérdezés pillanatában hány bejegyzés van összesen, az egyes bejegyzések milyen státuszban vannak, doménonként, illetve aldóménonként hány bejegyzés van, milyen az adatbázisban szereplő nyelvek megoszlása stb. Másrészt nemcsak az állapot felmérésére lehet használni, hanem minőségbiztosítási célokra is, amennyiben a rendszer képes kimutatást készíteni bizonyos adatok megléte vagy hiánya alapján vagy éppen az adatok integritásának megőrzéséről (pl. van-e olyan bejegyzések közötti hivatkozás, amely már nem él, mert a két összekötött bejegyzés közül az egyik törlésre került). Az adatokon túl célszerű az is, hogy számszerűsíteni lehessen a

terminológiai munkacsoport teherelosztását is, vagyis azt mérni, hogy ki, mennyit dolgozott, mely csoportoknak lehet segítségre szüksége egy adott alfeladat elvégzéséhez, mely doménokra vagy aldoménokra kellene koncentrálni, hol van elmaradás. A statisztikai adatok a végfelhasználó oldalán is fontosak lehetnek: hasznos lehet mérni az adatbázis látogatottságát, illetve azt, hogy az egyes olvasók mely bejegyzésekre keresnek, és milyen szűrési beállításokat használnak. Emellett érdekes lehet még az olvasó felhasználók és a szerkesztő felhasználók közötti interakciók számszerűsítése, vagyis hány terminusjavaslat vagy javítási kérelem érkezik be kívülről, és ez milyen mértékű interakciót eredményez.

3.4. Az informatikai háttértechnológia

Akármilyen adatbázist is használunk, nem szabad figyelmen kívül hagyni, hogy alapvetően informatikai rendszerekről beszélünk, ezért a keretrendszer kiválasztásánál ilyen szempontokat is figyelembe kell vennünk (Drewer & Schmitz, 2017: 142).

Az első és legfontosabb kérdés ezzel kapcsolatban, hogy **hol, milyen szerveren működik a terminológiai keretrendszer**. A piacon számos lehetőség van erre, ugyanakkor a nemzeti terminológiai adatbázis szempontjából elengedhetetlen, hogy egy megbízható intézet saját szerveréről legyen elérhető az adatbázis, amely több évre vagy akár több évtizedre képes biztosítani a megfelelő technikai infrastruktúrát. A hosszútávúság kérdése olyan szempontból is felmerül, hogy tudni kell biztosítani az adatbázis folyamatos elérhetőségét, amellyel kapcsolatban kiemelendő a biztonsági mentés folyamatos biztosítása olyan esetekre, amikor valamilyen hardverhiba vagy más infrastrukturális probléma miatt a szerver hirtelen leáll. Ebben az esetben nemcsak az elérhetőség a fő kérdés, hanem az adatvesztés kockázata is megnőhet. A másik alapvető szempont a megfelelő kibervédelem biztosítása, különösen olyan esetben, ha az adatbázis a végfelhasználók számára is hozzáférhető hozzászólások vagy terminusok beküldésének formájában.

Részben az informatikai architektúra, részben pedig az adatbázis funkcióinak kérdéskörébe tartozik az **interoperabilitás** is, vagyis az, hogy az adatbázisban tárolt adatok miként férhetők hozzá más rendszerek (tehát nem humán felhasználók) számára és azok hogyan tudják feldolgozni az adatokat. Terminológiai szempontból hasznos lehet például más, külső terminológiaiakivonatoló rendszerek csatolásának lehetősége (amennyiben ez a funkció nem található meg a szoftveres keretrendszerben), más, a nyelvhez (szakkorpuszok) vagy adott doménhoz tartozó szakadatbázisok, illetve akár forrásgyűjteményekkel foglalkozó adatbázisok (pl. könyvtári rendszerek) csatlakoztatása, amelyek segítik a bejegyzések kidolgozását, illetve amelyekkel megvalósulhat a kétirányú adatcsere is. Végfelhasználói szempontból pedig a különböző fordítástámogató, illetve tartalomkezelő rendszerekkel való kapcsolódás lehet hasznos. Az ideális keretrendszernek ezeket kell tudnia

biztosítani, valamint lehetőséget kell adnia a kapcsolódási beállítások finomhangolására is. Mindezek kapcsán azonban figyelembe kell venni, hogy az egyes külső kapcsolatok milyen infrastrukturális terhet róhatnak az adatbázisra, tisztában kell lenni azzal, hogy a szoftveres környezet képes-e adott mennyiségű forgalmat kezelni, valamint ehhez igazodva kell a megfelelő hardveres környezetet biztosítani.

4. Összefoglalás

Ebben a tanulmányban áttekintettük a magyar terminológiastratégia kialakítására vonatkozó általános ajánlások között szereplő magyar nemzeti terminológiai adatbázis tervezéséhez szükséges általános kérdéseket, elsősorban azt, hogy milyen technológiai követelményeknek kell megfelelnie és milyen informatikai háttérstruktúrára van szükség az adatbázis elkészítéséhez. Fontos szempont, hogy az ideális terminológiakezelő rendszernek mind az olvasó felhasználó, mind a szerkesztő felhasználó igényeit figyelembe kell vennie, valamint segítenie kell őket abban, hogy hatékonyan tudják használni a felületet. Előbbi szemszögéből lényeges, hogy hogyan látja a terminológiai adatokat, illetve milyen módon tud keresést végezni, az utóbbi számára pedig fontos a pontos és konzisztens adatbevitel, amit az adatkategóriák megfelelő kidolgozottsága, a szerkesztési alapelvek technológiai támogatása, valamint a rendszer által nyújtott automatizálási lehetőségek tesznek lehetővé. Mindemellett pedig szükséges az is, hogy a terminológiai munkát mint projekt folyamatot az ideális terminológiakezelő rendszer képes legyen hatékonyan támogatni.

Az általános szempontok és igények átgondolása után a következő lépés az elérhető rendszerek megvizsgálása és funkcióik, valamint lehetőségeik felmérése. Más nyelvtechnológiai alkalmazásokhoz hasonlóan, ma már a terminológiakezelő rendszerek széles palettája érhető el. Ezeket a szoftvereket egy másik írásunkban (Fóris & Somogyi, 2024) vizsgáljuk meg részletesen.

Irodalom

- B. Papp Eszter (2023). A terminológiastratégia kérdései Európában. In Prószéky Gábor (főszerk.), Fóris Ágota, Bölcskei Andrea, B. Papp Eszter & Lipp Veronika (szerk.), *A magyar terminológiastratégia kialakítása. Zöld könyv* (39–56). Budapest: Nyelvtudományi Kutatóközpont. doi: 10.18135/term.2023.4
- Benő Attila & Péntek János (2023). A terminológiastratégia szintjei és feltételei Erdélyben. In Prószéky Gábor (főszerk.), Fóris Ágota, B. Papp Eszter, Bölcskei Andrea & Lipp Veronika (szerk.), *A magyar terminológiastratégia kialakítása. Zöld könyv* (237–252). Budapest: Nyelvtudományi Kutatóközpont. doi: 10.18135/term.2023.12
- Bratanić, M. & Ostroški Anić, A. (2013). The Croatian National Termbank STRUNA. *Coll. Antropol.*, 37/3, 677–683.
- Drewer, P. & Schmitz, K. D. (2017). *Terminologiemangement. Grundlagen, Methoden, Werkzeuge*. Heidelberg: Springer.
- Faith, A. (2024). *Knowledge Graphs for Generative AI Use Cases*. LinkedIn Learning. <https://www.linkedin.com/learning/knowledge-graphs-for-generative-ai-use-cases> (2024. szeptember 27.)

- Fóris Ágota & B. Papp Eszter (2020). A preliminary study for the information structure of the terminology entry in the Hungarian (national) terminology database. In Brač, I. & Ostroški Anić, A. (eds.), *Svijet od riječi. Terminološki i leksikografski ogledi* [The world of words: Terminological and terminographic discussions] (325–342). Zagreb: Institut za hrvatski jezik i jezikoslovlje.
- Fóris Ágota, B. Papp Eszter & Bölcskei Andrea (összeáll.) (2023). Válogatott terminológiastratégiai bibliográfia. In Prószéky Gábor (főszerk.), Fóris Ágota, B. Papp Eszter, Bölcskei Andrea & Lipp Veronika (szerk.), *A magyar terminológiastratégia kialakítása. Zöld könyv* (331–348). Budapest: Nyelvtudományi Kutatóközpont. doi: 10.18135/term.2023.16
- Fóris Ágota & Bölcskei Andrea (2019). Ajánlások a magyar terminológiastratégiához. In Fóris Ágota & Bölcskei Andrea (szerk.), *Terminológiastratégiai kihívások a magyar nyelvterületen* (140–164). Budapest: L'Harmattan, OFFI Zrt.. Letöltés: <https://www.offi.hu/offi-akademia/kiadvanyok/terminologiastrategiai-kihivasok-a-magyar-nyelvterületen> [Utánközlés a Zöld könyvből, 2023, doi: 10.18135/term.2023.6]
- Fóris Ágota & Somogyi Zoltán (2024). A magyar terminológiastratégia megvalósíthatósága. A szoftveres keretrendszer kiválasztásának szempontjai a Magyar Nemzeti Terminológiai Adatbázis tervezése során. *Glossa Iuridica*, 11(1-2), 277–297. doi: 1055194/GI.2024.1-2.15.
- Heinisch, B. (2023). Terminological usability – adapting terminological databases to different user groups according to usability principles: The case of UniVieTerm. *Terminology Science & Research / Terminologie: Science et Recherche*, 26, 24–44. Letöltés: <https://journal-eaft-aet.net/index.php/tsr/article/download/3692/4582>
- Infoterm (2019). Terminológiapolitikai irányelvek. A terminológiapolitika kialakítása és megvalósítása különböző nyelvközösségekben. In Fóris Ágota & Bölcskei Andrea (szerk.), *Terminológiastratégiai kihívások a magyar nyelvterületen* (Ford.: Somogyi Zoltán, lektorálta: B. Papp Eszter) (167–218). Budapest: L'Harmattan Kiadó, OFFI Zrt. (Eredeti, angol nyelvű kiadás: UNESCO, 2005). Letöltés: <https://www.offi.hu/offi-akademia/kiadvanyok/terminologiastrategiai-kihivasok-a-magyar-nyelvterületen>
- Lanstyák István (2023). Nyelvmenedzselés-elmélet és terminológia. In Prószéky Gábor (főszerk.), Fóris Ágota, B. Papp Eszter, Bölcskei Andrea & Lipp Veronika (szerk.), *A magyar terminológiastratégia kialakítása. Zöld könyv* (253–276). Budapest: Nyelvtudományi Kutatóközpont. doi: 10.18135/term.2023.13
- Nilsson, H. (2017). *Trendencies in terminology*. Elhangzott: A SLOVKO 2017 – 9th International Conference: NLP, Corpus Linguistics, Terminology, e-Terminology konferencián 2017. október 26-án, Pozsonyban. Letöltés: <https://korpus.juls.savba.sk/~slovko/2017/Nilsson-Henrik-TNC-SLOVKO2017.pdf> (Hozzáférés: 2024. március 18.)
- Prószéky Gábor (főszerk.), Fóris Ágota, B. Papp Eszter, Bölcskei Andrea & Lipp Veronika (szerk.) (2023). *A magyar terminológiastratégia kialakítása. Zöld könyv*. Budapest: Nyelvtudományi Kutatóközpont. doi: 10.18135/term.2023
- Sermann Eszter & Tamás Dóra Mária (2020). Vizsgálati szempontrendszer online terminológiai adatbázisokhoz. In Fóris Ágota; Bölcskei, Andrea (főszerk.), *Nyelv, kultúra, identitás. Alkalmazott nyelvészeti kutatások a 21. századi információs térben: I. Terminológia, lexikográfia, fordítás* (84–94). Budapest: Akadémiai Kiadó.
- Sermann Eszter (2023). A terminológiai munkafolyamatok szempontjából releváns ISO-szabványok. In Prószéky Gábor (főszerk.), Fóris Ágota, Bölcskei Andrea, B. Papp Eszter & Lipp Veronika (szerk.), *A magyar terminológiastratégia kialakítása. Zöld könyv* (279–300). Budapest: Nyelvtudományi Kutatóközpont. doi: 10.18135/term.2023.14
- Tamás Dóra Mária (2023). A terminológiai adatbázisok mint a terminológiastratégia eszközei. In Prószéky Gábor (főszerk.), Fóris Ágota, B. Papp Eszter, Bölcskei Andrea & Lipp Veronika (szerk.), *A magyar terminológiastratégia kialakítása. Zöld könyv* (57–74). Budapest: Nyelvtudományi Kutatóközpont.. doi: 10.18135/term.2023.5
- Tolcsvai Nagy Gábor (2017). Bevezetés. In Tolcsvai Nagy Gábor (szerk.), *A magyar nyelv jelene és jövője* (19–31). Budapest: Gondolat Kiadó.

- Tolcsvai Nagy Gábor (2023). Nyelvstratégia és terminológia. In Prószéky Gábor (főszerk.), Fóris Ágota, B. Papp Eszter, Bölcskei Andrea & Lipp Veronika (szerk.), *A magyar terminológiastratégia kialakítása. Zöld könyv* (23–37). Budapest: Nyelvtudományi Kutatóközpont. doi: 10.18135/term.2023.3
- Warburton, K. (2021). *The Corporate Terminologist*. Philadelphia, Amsterdam: John Benjamins.
- Wright, S. E. & Budin, G. (eds.) (1997). *Handbook of Terminology Management. Basic Aspects of Terminology Management*. Vol. 1. Amsterdam, Philadelphia: John Benjamins.
- Wright, S. E. & Budin, G. (eds.) (2001). *Handbook of Terminology Management. Application-Oriented Terminology Management*. Vol. 2. Philadelphia, Amsterdam: John Benjamins.

Források

- DatCatInfo, <https://datcatinfo.net/> (2024. szeptember 26.)
- ISO 1087:2019 Terminology work and terminology science – Vocabulary
- ISO 26162-1:2019 Management of terminology resources — Terminology databases. Part 1: Design
- ISO 15188:2001 Project management guidelines for terminology standardization
- ISO 30042:2019 Management of terminology resources — TermBase eXchange (TBX)
- MTA web = Bemutatkozik az MTA Tudomány a Magyar Nyelvért Nemzeti Programja – Videón a közgyűlési tanácskozás. 2024. május 27. Letöltés: <https://mta.hu/kozgyules2024/az-mta-nemzeti-kutatasi-programok-bolcseszettudomanyi-alprogramjai-videon-a-kozgyulesi-tanacskozás-113706> (2024. szeptember 25.)
- Projektmenedzsment útmutató* (2023). PMBOK Guide Hetedik kiadás és A projektmenedzsment szabványa. Project Management Institute – PMI magyar Tagozat, Budapest.
- STRUNA, <http://struna.ihjj.hr/en/> (2024. április 3.)
- WCAG 2008 = Web Content Accessibility Guidelines (WCAG) 2.0. <https://www.w3.org/TR/2008/REC-WCAG20-20081211/> (2024. április 3.)