

TÖBB NYELVŰ OSZTÁLYOZÁSI RENDSZEREK KIDOLGOZÁSA ÉS HASZNOSÍTÁSA

H.-D. Lohmann

Wissenschaftlich-technisches Zentrum Automobilbau,
Karl-Marx-Stadt

1. Bevezetés

A tudományos és műszaki információk biztosítása és cseréje nemcsak nemzeti, hanem nemzetközi szinten is egyre nagyobb jelentőségre tesz szert. A nemzeti határon kívüli információcserére számos szervezési és módszertani segédeszközt igényel. Ez utóbbiak közé tartoznak az e célokra alkalmas kétnyelvű, illetve több nyelvű teauruszok formájában ismeretes osztályozási rendszerek. Ezek teszik lehetővé a több oldalú munkamegosztás keretében feldolgozott információk közös tárolását, visszakeresését és szolgáltatását. Csak ilyen alkalmazási célok esetében térülnek meg a több nyelvű teaurusz kidolgozásának ráfordításai. Ha egyedi kérdést akarunk feltenni külföldön található, illetve idegen nyelven szervezett teaurusznak (értsd: egy ilyen teauruszt alkalmazó információkereső rendszernek), akkor elég a kereső kérdés lefordítása. Másfelől viszont, ha élénk nemzetközi információcserére törekszünk, igen nagy a ráfordítás a megfelelő szervezési és módszertani segédeszközök alkalmazása nélkül. E ráfordítás az alábbi *költségtényezőkből* állhat:

- a forrás (dokumentum) tanulmányozása eredeti nyelven,
- a forrás fordítása a referáló nyelvére,
- a referátum elkészítése a referáló nyelven,
- a referátum fordítása a felhasználók nyelvére,
- a forrás beszerzése a felhasználók részére,
- a forrás fordítása a felhasználók nyelvére,
- a forrás tulajdonképeni tartalmi hasznosítása.

Ezt a munkáigényt már a több nyelvű teaurusz léte is nagymértékben csökkentheti.

2. A több nyelvű teaurusz jellemzői

A több nyelvű teaurusz a következők szerint jellemezhető [1]:

- tárgyszavak, illetve tárgyszócsoportok betűrendes és szisztematikus jegyzéke több nyelven;
- segédeszköz a dokumentumok automatizált visszakereséséhez;

a teaurusz segítségével az egyik nyelven megfogalmazott kereső kérdéseket és a dokumentum másik nyelven megfogalmazott „tartalmi súlypontjait” (keresőképét – szerk. megj.) egy és ugyanazon információkereső nyelvre lehet lefordítani;

a több nyelvű teauruszban az összes alkotó nyelven megadott tárgyszavakat meghatározott lexikográfiai szabályok szerint kezelik;

azokat a tárgyszavakat, amelyek fogalmi tartalma kielégítően egyezik, közös ekvivalenciaosztályba fogják össze;

minden ekvivalenciaosztályt egy deskriptor, vagy egy fogalmi kódszám reprezentál;

a kielégítő mértékben egyező ekvivalenciaosztályok több nyelvű reprezentáns deskriptorait egyértelműen egymáshoz rendelik: ez többnyire a közös azonosító kód segítségével történik;

a fogalmak fölé-, alá- és mellérendelését (szemantikai összefüggéseit, a szerk. megj.) a nyelvek mindegyikén a megfelelő deskriptorok közötti hivatkozások segítségével ábrázolják.

A szocialista országokban célszerű a több nyelvű teauruszt az orosz nyelv által összekötött kétnyelvű teauruszok komplexumaként felépíteni.

3. A több nyelvű teaurusz kidolgozása

3.1. A kidolgozás módjai

a) *A különböző országokban egymástól függetlenül jöttek létre nemzeti teauruszok. Szerkezetük nem egyezik, az általuk felölelt szakterületek csak részben egyeznek.*

Ebben az esetben nincs akadálya annak, hogy az egymástól függetlenül létrejött teauruszokat közös kódok segítségével összekapcsoljuk. Az egymással kapcsolatba hozott teauruszok alapszerkezete emellett változatlan marad.

Ilyen módon azonban a kompatibilitás alapfeltételei (ld. alább) nem tarthatók be. Ebből súlyos problémák adódnak, amelyek az ilyen viszonylag egyszerű alapokon létrejött teaurusz működőképességét kérdésessé teszik.

b) A különböző országokban egymástól függetlenül nemzeti teauruszok jöttek létre. Szerkezetük nem egyezik, az általuk felölelt szakterületek csak részben egyeznek.

Egy nemzetközi összetételű munkacsoport több nyelvű, egységes alapszerkezetű teauruszt dolgoz ki. Ez esetben a nemzeti sajátosságoktól messzemenő elvonatkoztatásra van szükség. A kidolgozást olyan általános alapelvek szerint kell végezni (a szerkezet, a fogalmak terjedelme, a fogalmak tartalma stb. tekintetében), amelyeket az összes együttműködő partner el tud fogadni. Ennél a megoldásnál a ráfordított munka túlnyomó része az egyes országok közötti hosszadalmas egyeztetés lesz. Ez csak számos centralizált munkafázis során oldható meg.

c) Az egyik érdekelt országban van nemzeti keretek között bevált teaurusz, amelynek szerkezete és fogalmi terjedelme nemzeti keretek között használhatónak bizonyult.

Az ezek szerint egy országban kipróbált teauruszt lefordítják n célnyelvre, s így n -nyelvű teaurusz jön létre. Ez esetben az a probléma merül fel, hogy a fordítás nem biztosítja magától értetődően a fogalmi tartalom azonosságát az egyes nyelveken. Egyes fogalmak esetében a más nyelvre fordítás olykor lehetetlennek is bizonyulhat. A teaurusz tárgyszavai között E ekvivalenciájától ($0 \leq E \leq 1$) függően (ld. alább) a nemzetközi egyeztetés kisebb, illetve nagyobb mértékű ráfordítást követel.

d) A feldolgozandó területnek az érdekelt országok egyikében sincs teaurusza.

Egy nemzetközi munkacsoport a b) pontnak megfelelő egységes alapszerkezetű, n -nyelvű teauruszt dolgoz ki. A b) szerinti esettel ellentétben itt egységes irányelvekből lehet kiindulni. Ezek képezik az alapot az n -nyelvű kompatibilis teauruszok előállításához. A munka ez esetben arányosan oszlik meg a tárgyszógyűjtemények nemzeti keretekben történő kidolgozása és a nemzetközi egyeztetés között.

3.2. A kidolgozás elvei

A több nyelvű teaurusz kidolgozásakor rendeltetésének és feladatainak megfelelően a következőkből kell kiindulni:

3.2.1. Az információfeldolgozás követelményeiből

A teaurusz a gépesített, részben automatizált, illetve automatizált dokumentum-visszakeresés szükséges feltétele. Ezért nem szabad öncélúan összeállítani minél nagyobb méretű tárgyszógyűjtemény formájában, hanem elsősorban a korszerű számítástechnika követelményeivel

kell számot vetni. A szocialista országokban a számítástechnika az ESZR alapján fejlődik. Ez kedvező feltételeket teremt ahhoz, hogy a több nyelvű teauruszok kidolgozásakor az ESZR technika számítógépi programrendszerét az összes partner érdekében útbaigazításként (orientációként) használhassuk.

Ezzel kapcsolatban az is nyilvánvaló, hogy a több nyelvű teaurusz kidolgozásával egyidejűleg létre kell hozni a résztvevő országokban egyaránt kötelező indexelési irányelveket. Ezekre az elvekre az illető országokban alkalmazott számítástechnika rányomja bélyegét, és ezeknek a maguk részéről is szavatolniuk kell az információfeldolgozásának kompatibilitását a partnereknél.

3.2.2. A kompatibilitás (illeszthetőség) feltételeiből

Kompatibilitásra kell törekedni:

a különböző nemzetközi ágazati információs rendszerek több nyelvű teauruszai között;

nemzetközi ágazati információs rendszerek több nyelvű teauruszai és a forrásra (dokumentumtípusra) orientált nemzetközi információs rendszerek több nyelvű teauruszai között;

a több nyelvű teauruszok nemzeti változatai és a kötelező* egy nyelvű nemzeti teauruszok között.

A kompatibilitás biztosítása érdekében az alábbi követelményeknek kell eleget tenni:

a) Egy több nyelvű teauruszon belül ugyanazon ekvivalenciaosztályok különböző nyelvű deskriptorainak egymással egyértelműen ekvivalenseknek kell lenniük.

Más szavakkal: egy és ugyanazon fogalmi tartalom kifejezésére a több nyelvű teaurusz egyes nemzeti komponenseiben azonos megnevezéseket (deskriptorokat), (értsd: azonos értelmű deskriptorokat, szerk. megj.) kell alkalmazni.**

b) Ugyanazon ekvivalenciaosztályok vonatkozásában a több nyelvű teaurusz deskriptorai feleljenek meg a kötelező nemzeti teauruszok deskriptorainak.

c) A több nyelvű teaurusz nemzeti komponensei alapszerkezetüket tekintve legyenek azonosak. Ez egyúttal azt is megköveteli, hogy a teauruszokat egyeztetett módszertani elvek szerint dolgozzák ki.

* A „kötelező” teauruszokról szólva a szerző az NDK erősen koordinált teaurusz-szerkesztési és -használati rendjéből indul ki (szerk. megj.)

** A gyakorlati tapasztalatok szerint ez az ekvivalencia (fogalmi egybeesés) gyakran nem biztosítható deskriptor és deskriptor, csupán deskriptor és deskriptor-kombináció között, (szerk. megj.)

3.3. Több nyelvű teaurusz kidolgozásánál felmerülő problémák

3.3.1. Az ekvivalencia problémái

Az ekvivalenciának a több nyelvű teauruszon belüli elérése érdekében a következő feltételeket kell betartani [2]:

sokoldalúan (minden partnernél) álljanak rendelkezésre azok a deskriptorok, amelyek segítségével az információcsere történik;

a minden partner birtokában meglévő, de nyelvileg különböző deskriptorok fogalmi tartalma messzemenő egybevághósága a tartalmilag egyező ekvivalenciaosztályok alapján;

az egymáshoz kapcsolódó több nyelvű deskriptorok egymással ekvivalens voltának egyértelmű kimutatását; olyan egységes osztályozási alapszerkezet, amelynek alapján az összes nyelven rendezik a deskriptorállományt.

A több nyelvű teaurusz kidolgozása során különböző ekvivalenciatípusok fordulhatnak elő. Ezek a következők:

a) Teljes ekvivalencia

Akkor jön létre, ha egy deskriptor különböző nyelvű komponenseinek fogalmi tartalma teljes mértékben egyezik.

német	orosz
-------	-------

pl. Motordrehzahl	Число оборотов двигателя (motor fordulatszám)
-------------------	--

b) Részekvivalencia

Akkor jön létre, ha egy deskriptor különböző komponenseinek fogalmi tartalma csak részben egyezik. Ez elsősorban akkor fordul elő, ha az egyik nyelv egyik deskriptorához rendelt tárgyszó szemantikai tartalma a másik nyelven eltérő. Ekkor a két komponens között a következő viszony áll fenn a tartalmi egyezés tekintetében.

1 : több, több : 1, 1 : rész, stb.

német	orosz
-------	-------

Pl. Motorrad Vibration Schwingung Kraftfahrzeug (gépjármű) (nem sínhez kötött motorhajtású szállítóeszköz)	мотоцикл, мотоциклет (motorkerékpár, moped) вибрация (rezgés, lengés) автомобиль (gépkocsi)
--	--

Ilyen részekvivalenciák felléptekor meg kell vizsgálni, hogy tudomásul lehet-e venni a szemantikai tartalomban meglévő különbséget, vagy pedig a résztvevő nyelvek megfelelő tárgyszavainak fogalomtartalmát (-terjedelmét) határozottan megváltoztatva lehet-e egyezést elérni.

c) Nincs ekvivalencia („null-ekvivalencia”)

Ez akkor fordul elő, ha valamely deskriptor fogalmi tartalma csak a résztvevő nyelvek egyikén, vagy egy részében létezik.

német	orosz
-------	-------

Pl. Kirsche

Ha valamely fölérendelt fogalomnak nincs megfelelője a partneryelven, gyakran lehet a hozzárendelhető alfogalmakat alkalmazni. Ilyen módszerrel az ekvivalencia hiányát („null-ekvivalenciát”) ki lehet küszöbölni.

A tapasztalatok szerint az orosz-német nyelvkombináció vonatkozásában a fogalmak kb. 80%-ánál áll fenn szemantikai megegyezés.

3.3.2. Az osztályozási alapszerkezet problémái

A teaurusz kidolgozásához vezető út meghatározásával egyidejűleg döntést kell hozni az osztályozási alapszerkezetről, vagyis a fogalmak rendezésére alkalmazandó rendszerről is. E rendszernek megfelelően kell felosztani az elkészülő tárgyszógyűjteményt.

Egy több nyelvű teaurusz szempontjából rendkívül fontos, hogy a fogalmak rendezésére szolgáló rendszer a népgazdasági nemzeti szerkezetétől függetlenül alakuljon ki. Ezért a deskriptorok szisztematikus elrendezése tudományszaktól függetlenül történjék. Ez a szempont különösen akkor vehető figyelembe már a kidolgozás kezdetétől fogva, ha a több nyelvű teauruszt újonnan építik fel és nem egy már meglévő teauruszból indulnak ki.

3.4. A kidolgozás szakaszai

A több nyelvű teauruszok kidolgozásának eddigi tapasztalatai alapján a következő munkaszakaszokat állapíthatjuk meg [1]:

- a) Irányelveinek kidolgozása a teaurusz létrehozására, ideértve
 - a kidolgozás módszereinek meghatározását;
 - a több nyelvű teaurusz tematikai határainak és alapszerkezetének rögzítését;
 - a teauruszba felveendő fogalmak és az azonosítási kód meghatározását.

b) A több nyelvű deskriptorjegyzék (vagy az ugynevezett makrotezaurusz) kidolgozása, az ekvivalenciaosztályok megadásával.

c) A több nyelvű és sokoldalúan egyeztetett indexelési irányelvek kidolgozása, valamint az alkalmazandó számítástechnikai eszközök és eljárások meghatározása az a) és b) ponttal párhuzamosan.

d) Nemzetközi indexelési és visszakeresési kísérlet végrehajtása a következő célból

a deskriptorok és nem-deszkriptorok teljességének vizsgálata;

hierarchikus és asszociatív kapcsolatok meghatározása;

a több nyelvű koordinált indexelés kipróbálása.

e) A hierarchikus és asszociatív kapcsolatok kidolgozása.

f) A több nyelvű tezaurusz gyakorlati alkalmazása és a módosításokat átvezető szolgálat megszervezése.

Irodalom

- [1] ROSENBAUM, M.-D.: Methodische Grundsätze für die Erarbeitung mehrsprachiger Thesaurien der Mitgliedsländer des RGW. (Módszertani alapelvek a KGST tagországok több nyelvű tezauruszainak kidolgozására.) = Informatik, 21. k. 2. sz. 1974. p. 17–21.
- [2] Vorläufige methodische Richtlinie für die Erarbeitung russisch-deutsch, deutsch-russischsprachiger Thesaurien. (Ideiglenes módszertani irányelvek orosz-német és német-orosz tezauruszok kidolgozására.) ZIID, Berlin, 1974.
- [3] BRENDLER, G.: Der mehrsprachige Thesaurus – ein Instrument zur Rationalisierung des Informationsflusses im internationalen Rahmen. (A több nyelvű tezaurusz, mint a nemzetközi információáramlás ésszerűsítésének eszköze.) Beiträge zur maschinellen Sprachdatenverarbeitung. Berlin, 1973.
- [4] MATER, E.: Elementarbegriffe als Grundlage maschineller Indexierung und Konvertierung. (Elemi fogalmak, mint a gépi indexelés és konvertálás alapjai.) = Dokumentation/Information, Schriftenreihe des Institutes INER der TH Ilmenau, 25. sz. p. 47–60.
- [4] SCHIMPFERMANN, M.: Erfahrungen bei der Entwicklung eines internationalen Bauinformationssystemes. (A nemzetközi építésügyi információs rendszer fejlesztésekor szerzett tapasztalatok.) = Informatik, 21. k. 1. sz. 1974. p. 8.
- [5] BRENDLER, G.: Probleme beim Aufbau eines zweisprachigen Thesaurus im Sprachenpaar deutsch-russisch. (Német-orosz tezaurusz készítésekor felmerülő problémák.) = Dokumentation/Information, Schriftenreihe des Institutes INER der TH Ilmenau, 24. sz. p. 57–72.

LOHMANN, H.-D.: Több nyelvű osztályozási rendszerek kidolgozása és hasznosítása

Tudományos és műszaki információk nemzetközi cseréje módszertani segédeszközként két- vagy több nyelvű tezauruszok kidolgozását igényli a nemzetközi munkamegosztással feldolgozott információk és dokumentumok közös tárolása, visszakeresése és szolgáltatása céljából.

A több nyelvű tezaurusz kidolgozásának technikája megköveteli a fogalmak ekvivalenciaosztályainak létrehozását. A fogalmi osztályok tartalmát a komponens nyelveken kifejező tárgyszavak a több nyelvű tezaurusz deskriptoraiként használhatók. A cikk felvázolja a több nyelvű tezauruszok készítési módját, a különféle nyelvű deskriptorok teljes és részleges ekvivalenciájának illetve az ekvivalens deskriptorok hiányának problematikáját és felsorolja a kidolgozás szakaszait.

LOHMANN, H.-D.: Devising and using multilingual systems of classification

The international exchange of information in science and technology requires methodical aids in form of bilingual or multilingual thesauri for storage, retrieval and dissemination of information processed in multilateral shared labour.

Techniques of elaboration of multilingual thesauri require the determination of classes of concept equivalencies. Terms of the component languages explaining the conceptual content of these classes may be chosen as descriptors of the multilingual thesaurus.

The method of the elaboration of multilingual thesauri under various starting conditions, as well as problems of total, partial and zero equivalencies are treated with, and steps of the elaboration are described.

ЛОХМАНН, Х. Д.: Разработка и функционирование многоязычных систем классификации

Международный обмен научно-технической информацией требует пособий в форме двух- или многоязычных тезаурусов для совместных хранения, поиска в выдаче обрабатываемых информационных материалов.

Техника разработки многоязычных тезаурусов требует создания классов условной эквивалентности понятий. Ключевые слова разных языков-компонентов, которые выражают смысловое со-

держание этих классов условной эквивалентности, могут служить в качестве дескрипторов многоязычного тезауруса.

Различные методы построения многоязычного тезауруса, в случае наличия различных исходных материалов, а также проблемы полной, частной и нулевой эквивалентностей дескрипторов различных языков рассматриваются подробно, и представляются этапы разработки.

LOHMANN, H.—D.: Erarbeitung und Nutzung mehrsprachiger Klassifikationssysteme

Der internationale Austausch von wissenschaftlichen und technischen Informationen erfordert methodische

Hilfsmittel in der Form von zwei- oder mehrsprachiger Thesauri für das gemeinsame Speichern, Recherchieren und die Herausgabe von mehrseitig erarbeiteten Informationen.

Die Technik der Erarbeitung eines mehrsprachigen Thesaurus erfordert die Aufstellung von Äquivalenzklassen der Begriffe. Jene Sachwörter (sprachlichen Komponente) der Thesauri, die den Begriffsinhalt der Äquivalenzklasse repräsentieren, können als Deskriptoren des mehrsprachigen Thesaurus benützt werden. Die Erarbeitungsweise der mehrsprachigen Thesauri und die Problematik der totalen Äquivalenzen, Teiläquivalenzen und Nulläquivalenzen von Deskriptoren in mehreren Sprachen werden eingehend behandelt und die Etappen der Erarbeitung angegeben.



**A
MAGYAR KÖNYVTÁROSOK EGYESÜLETE
MŰSZAKI KÖNYVTÁROS SEKCIÓ
HÍREI**

Tájékoztató

A Műszaki Könyvtáros Szekció február 19-én tartott összejövetelén „Mikromásolatok szerepe a könyvtári és tájékoztatási munkában” címmel tartott előadást DEZSŐ László (NIMDOK) és TÓKÉS László (MTA Könyvtára).

A téma bevezetőjében DEZSŐ László összefoglalta a mikrofilmtechnika elterjedésének okait; alkalmazásának jelentőségét. Ismertette a VINITI-ben használatos mikrofilm-tárolási és kereső rendszert és a hozzá tartozó berendezéseket.

Elmondta, hogy a KGST tagországok intézkedéseket tettek a mikromásolatok méreteinek szabványosítására.

Az előadás második részének bevezetőjében TÓKÉS László elemezte a jelenleg alkalmazott mikroformákat. Részletesen ismertette a felhasználási területeket, illetve a felhasználás szempontjainak változását. Kitért az előállítás és a tárolás költségeire is. Előadását gazdag mintaanyag tette szemléletessé.

A szekció a későbbiekben írásos tájékoztatót küld tagjainak, összefoglalva és kiegészítve az elhangzottakat, különös tekintettel a mikrofilmek beszerzési, feldolgozási és szolgáltatási problémáira.