

---

# Többnyelvű tezaurusz építése és szolgáltatása webes környezetben

---

Förhécz András, [fand\\_lev@freemail.hu](mailto:fand_lev@freemail.hu)

Mészáros Tamás, [meszaros@mit.bme.hu](mailto:meszaros@mit.bme.hu)

BME Méréstechnika és Információs Rendszerek Tanszék

---

# Áttekintés

- n DISCO projekt
  - n Támogatott eszközök
  - n Források
  - n Tezaurusz szerkezete
  - n Fogalomtárak összevonása
  - n Munkafázisok
  - n Tezaurusz bíráló rendszer
  - n Nyilvános portál
-

---

# Probléma

- n közös munkaerőpiac az Európai Unióban
    - q sokan tanulnának, dolgoznának külföldön
  - n képességek, elvárások kommunikációja
    - q Curriculum Vitae, álláshirdetések
  - n kormányzati oldalról egységes követelmények
  - n EU kínál eszközöket (*transparency tools*)
    - q például Europass CV
  - n *de nincs egységes terminológia*
-

---

# DISCO projekt

- n European **D**ictionary of **S**kills and **C**ompetencies
  - n Célok:
    - q közös terminológia lefektetése a szakmai képességekről és képesítésekről
    - q többnyelvű, automatikus fordítás
    - q példa alkalmazás a használatára
  - n Nyelvek (résztevő partnerek):
    - q angol, cseh, francia, litván, magyar, német
-

---

# Támogatott eszközök

- n European transparency tools
    - q átláthatóság és átjárhatóság támogatása
    - q pl. Europass CV
  - n átfogó szaknyelvi támogatás
  - n tárgykör: szakmai készségek és képesítések
  - n környezetfüggetlen
    - q egy alkalmazás felé sem elkötelezett
  - n kísérleti alkalmazás
-

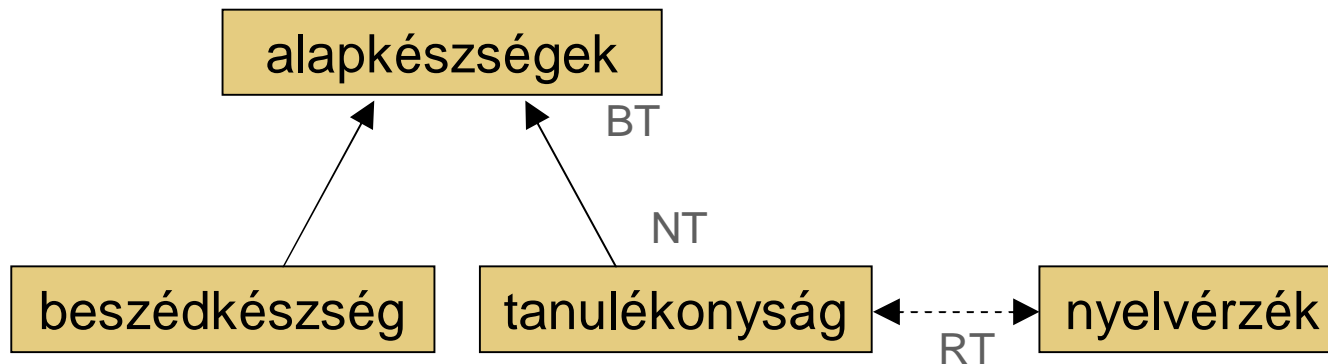
---

# Források

- n országos adatbázisok
    - q jogi szabályozás megköveteli
  - n ISCED, 1975
    - q UNESCO, oktatás elemzésére szolgáló tanulmány
    - q kilenc átfogó szakterület
  - n Europass CV
    - q nagyobb témák, mint nyelvtudás, jogosítvány
  - n nincs szabványos osztályozás
  - n nemzetközileg nehezen összeegyeztethetők
-

# Tezaurusz

- n fogalmak és szemantikus relációik leírása
- n hierarchiába rendezett fogalmak
  - q *broader term (BT)* , *narrower term (NT)*
- n kapcsolódó fogalmak: *related term (RT)*
- n szinonimák, forrás, magyarázat stb.



---

# Összekapcsolás vagy egyesítés?

- n átfogó tezaurusz elkészítése
    - q összekapcsolás (*mapping*)
    - q egyesítés (*merging*)
  - n többnyelvűség
    - q egységes struktúra: szimmetrikus tezaurusz
    - q részleges fordítás – nem minden fordítható le
    - q eltérő rendszerezés, pl. nyomda: ipar vagy média?
  - n egységes jelölésrendszer (közös terminológia) kell
    - ž **egyesítés**
-



---

# Elosztott fejlesztés

- n konzorcium végzi, nem lehet teljesen központosított
  - n tudásbázisok elosztott fejlesztése
    - q ontológia szerkesztők: Ontolingua, KAON
    - q komplexitás miatt nem célszerű
  - n megbízott partner végzi az egyesítést
  - n partneri véleményezés, majd javítás
    - q a véleményezéshez saját szoftvert fejlesztettünk
-

---

# Munkafázisok

- n már létező források összegyűjtése
  - n egyesítés angol nyelven
  - n partneri véleményezés és javítás
  - n nyelvi változatok fordítása
    - q részleges fordítás
  - n kísérleti alkalmazás
- } szektoronként iterálva
-

---

# Tezaurusz bíráló rendszer

- n importálás, változások követése
  - n áttekinthető fogalmi struktúra (hierarchia)
  - n kereshető, szűrhető tartalom
  - n minden fogalomhoz fórum tartozik
    - q bírálók megjegyzései
    - q változások automatikus bejegyzése
-

# Tezaurusz bíráló rendszer

### Browse hierarchy

Broader	active	Narrower																
TT1 <b>job-related skills and competencies</b>	TT15 <b>arts</b>	<table style="width: 100%; border-collapse: collapse;"> <tr><td style="border: 1px solid black; padding: 2px;">TT153</td><td style="border: 1px solid black; padding: 2px;"><b>ability to read music</b></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">TT178</td><td style="border: 1px solid black; padding: 2px;"><b>aesthetic sensitivity</b></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">TT180</td><td style="border: 1px solid black; padding: 2px;"><b>creativity</b></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">TT160</td><td style="border: 1px solid black; padding: 2px;"><b>draftsmanship</b></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">TT182</td><td style="border: 1px solid black; padding: 2px;"><b>fashion sense</b></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">TT171</td><td style="border: 1px solid black; padding: 2px;"><b>imitative ability</b></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">TT185</td><td style="border: 1px solid black; padding: 2px;"><b>musical talent</b></td></tr> <tr><td style="border: 1px solid black; padding: 2px;">TT169</td><td style="border: 1px solid black; padding: 2px;"><b>translation and interpretation skills</b></td></tr> </table>	TT153	<b>ability to read music</b>	TT178	<b>aesthetic sensitivity</b>	TT180	<b>creativity</b>	TT160	<b>draftsmanship</b>	TT182	<b>fashion sense</b>	TT171	<b>imitative ability</b>	TT185	<b>musical talent</b>	TT169	<b>translation and interpretation skills</b>
TT153	<b>ability to read music</b>																	
TT178	<b>aesthetic sensitivity</b>																	
TT180	<b>creativity</b>																	
TT160	<b>draftsmanship</b>																	
TT182	<b>fashion sense</b>																	
TT171	<b>imitative ability</b>																	
TT185	<b>musical talent</b>																	
TT169	<b>translation and interpretation skills</b>																	

**Term name:**

Match:  any  start of name  exact

**Reference:**  ▼

is code #  [click to select](#)

**Source:**

code #  [click to select](#)

**Subject field:**

code #  [click to select](#)

**Created:**

-

**Modified:**

-

}

and  
 or

### Term view

TT15	<b>arts</b>	[draft]
Broader terms:	<b>job-related skills and competencies</b>	
Narrower terms:	<b>ability to read music, draftsmanship, translation and interpretation skills, imitative ability, aesthetic sensitivity, creativity, fashion sense, musical talent</b>	
Use for terms:	<b>artistic interests</b>	
Scope note:	Use for job-related skills in the area of fine arts: drawing, painting, sculpture; performing arts: music, drama, dance, circus; graphic and audio-visual arts: photography, cinematography, music production, radio and TV production, printing and publishing; design; craft skills.	
Subject field:	21-Arts	
Created:	2005-11-21	
Modified:	2005-11-29	

[post] #2 [reply] system

New 'use for' term: **artistic interests**

#1 [reply] system

Scope note changed from:

to:

Use for job-related skills in the area of fine arts: music, drama, dance, circus; graphic cinematography, music production, radio craft skills.

Created: 2005-11-21 .

Last modified: 2005-11-29.

New broader term: **job-related skills and competencies**

New narrower term: **ability to read music**

New narrower term: **draftsmanship**

New narrower term: **translation and interpretation skills**

New narrower term: **imitative ability**

---

# DISCO Online Tool

- n adatbázis alapú webrendszer
  - n publikus web portál
    - q bármelyik nyelven bejárható a hierarchia
    - q fogalmak lefordíthatóak
    - q személyes kompetencia profil építhető
  - n webszolgáltatás
    - q Európai Unió portálok, pl. Europass CV
-

# DISCO Online Tool

<a href="#">Home</a>   <a href="#">Browse</a>   <a href="#">Search</a>   <a href="#">Translate</a>   <a href="#">CV profiler</a>	
<b>Navigate this page</b> <a href="#">&gt; Sectors</a> <a href="#">&gt; Soft skills</a>	<b>Thesaurus language:</b> Hungarian <input type="button" value="OK"/>
<b>Tip:</b> Click on a flag and see this page in that language!	<h2>Számítástechnikai ismeretek</h2>
<b>Database Info</b> Pilot version for "health" and "architecture". <a href="#">1145 Terms</a> <a href="#">8 languages</a>	<b>description:</b> Használható olyan alapvető számítástechnikai műveltségre, melyek szükségesek a legtöbb irodai munkánál, pl. szövegszerkesztés, táblázatkezelés, stb.
	<b>synonyms:</b> informatikatudás, számítástechnikai írástudás
	<b>broader terms:</b> <a href="#">kulcsképessegek</a>
	<b>narrower terms:</b> <a href="#">email-használat</a> <a href="#">internethasználat</a> <a href="#">képesség új szoftverek elsajátítására</a> <a href="#">speciális szoftverek ismerete</a>
	<b>related terms:</b> <a href="#">számítástechnika</a>

---

**Köszönöm a figyelmet!**

---