

# TÉMATÉRKÉP TECHNOLÓGIA BEMUTATÁSA, A TM4L SZERKESZTŐ MAGYAROSÍTÁSÁNAK FOLYAMATÁN KERESZTÜL

*Kovács Zoltán Csaba*  
*Fejlesztési igazgató*  
KRIPTO Kutatásfejlesztési Kft.  
[kovacs.zoltan@kripto.hu](mailto:kovacs.zoltan@kripto.hu)

*Tóth Máté*  
*Könyvtáros*  
OSZK – Könyvtári Intézet.  
[thmate@oszk.hu](mailto:thmate@oszk.hu)

## ABSZTRAKT

Tanulmányunkban, a TM4L tématerkép szerkesztő és vizualizáló szoftver magyarosításának folyamatán keresztül szeretnénk bemutatni, a tématerképek felhasználási lehetőségeit. A tématerképek használatát elsősorban a tudás megosztásában tartjuk fontosnak. Olyan technológiát jelent, melynek használatával a kommunikációnk egyértelműbbé válik azáltal, hogy a tárgyiasult információt ugyanúgy értelmezi az a személy aki „leírta”, mint az aki „olvassa”.

**Kulcsszavak:** topic map, tématerkép, téma, asszociáció, előfordulás, erőforrás, tudásmegosztás, szemantikus web, ontológia

## 1. BEVEZETÉS

A tématerképek használata nagyon szűk körű. Különösen igaz ez a magyarországi használatára. Ez ösztönzött bennünket arra, hogy egy meglévő, ingyenes tématerkép (topic map, rövidítése: TM) szerkesztő (editor) és vizualizáló (visualization) szoftvert ajánljunk a magyarországi potenciális felhasználóknak. Fontosnak tartottuk, hogy a tudásmegosztás technológiájában magyar nyelvű legyen az alkalmazás, hiszen anyanyelvünkön gondolkozunk, így hatékonyabbak lehetünk.

2007. novemberében találtunk rá egy észak-Karolinai egyetem, a Winston-Salem State University honlapjára: <http://compsci.wssu.edu/iis/nsdl/download.html>, amelyről letölthettünk egy TM szerkesztőt és vizualizálót. A szoftver két projekt (TM4J és TM4L) végterméke volt. A projektek vezetője Darina Dicheva kutatónő felajánlotta, hogy ha magyarosítjuk a programot, akkor a magyar verziót is felteszi a letölthető programjai közé. <http://myweb.wssu.edu/dichevad/>

Erről a munkánkról szeretnénk jelen publikációnkban beszámolni. A beszámolónk alkalmat ad arra, hogy a tématerképek használatára kitérjünk, és ezzel is gyarapítsuk a tématerképekről szóló magyar nyelvű írások körét.

Cikkünkben, a magyar kifejezéseket használjuk, és zárójelben megadjuk az angol megfelelőjét.

## 2. HÁTTÉR ÉS MOTIVÁCIÓ

A szerzők korábbi kutatásuk folytatásának szánták jelen cikküket. Kutatási területük a tudásmenedzsment témakör (theme). Kovács Zoltán a menedzsment területen végez kutatásokat, Tóth Máté a könyvtári szolgáltatásban bekövetkező paradigmaváltással, a szemantikus web technológia terjedésével foglalkozik.

Mindkét szerző fontosnak tartja, hogy a felmerült fogalmakat példákkal világítsák meg. A tématerkép szerkesztés megköveteli a fogalmak (subject) egyszerű ugyanakkor egyértelmű leírását, és ez is közrejátszott abban, hogy a szerzők érdeklődése a tématerképek felé fordult. Ösztönzőleg hatott ránk az is, hogy a Topic Maps támogatja a vizualizációt.

Mint minden új módszer ez is magában hordja azt a veszélyt, hogy nem terjed el, illetve széles körben elterjed, majd elhal. A döntési táblák használatánál figyelhettük meg a széleskörű használatot, majd az elhalását.

## 3. A TÉMATÉRKÉPEKRŐL RÖVIDEN

A tématerkép egy elektronikus dokumentum, melyben gondolatainkat tárgyasítjuk. Hasonlíthatnánk jegyzetfüzethez, noteszhez, glosszáriumhoz, tartalomjegyzékhez, indexhez, iktatókönyvhöz, tárgymutatóhoz, szinonimaszótárhoz, és még sok másához, hiszen mindezeket a szerepeket betöltheti.

A Tématerkép három építőköve: téma (topic), asszociáció (association) és az előfordulás (occurrence). A TM-ben a témák és asszociációk mellett rövidebb adatelemek (feljegyzések) is tárolódnak, amelyeket belső előfordulásoknak is neveznek (internal occurrence). A nagyobb állományok külön tárolódnak, és a TM-ben csak az állományok elérési címe található; egy hiperlink segítségével hívhatjuk be a külön helyen lévő tartalmat. Ez az állomány lehet egy file szerveren, egy iratkönyvtárba és azon belül egy irattárban, vagy lehet a Weben. A TM-en belül, a topic-ok között is hiperlinkek segítségével közlekedhetünk, illetve az alkalmazáshoz írt kezelőprogram szerint.

A tématerképben a **téma** (topic) áll a középpontban, és a témához kapcsoltn írjuk be tudásunkat, gondolatainkat. A téma fogalmát úgy érthetjük meg, hogy kiindulunk a gondolatainkból, melyek a valós világ dolgainak (objektumainak, tárgyainak), eseményeinek és egyéb jelenségeinek megjelenése az agyunkban. Gondolatainkhoz érzelmek, érzetek is kötődnek, amelyeket ma még nem tudunk tárgyasítani. Agyunkban a gondolatok között asszociálunk, és ezeket, az **asszociációkat** (associations) már leírjuk a tématerképben.

Gondolataink elemeit nevezzük fogalmaknak (subject). Fogalom bármi lehet, amiről beszélünk, vagy az agyunkban megfogant. A fogalom lehet címezhető (addressable subject), vagy nem címezhető (non-addressable subject). Azokat a fogalmakat tekintjük címezhetőnek, melyek tárgyasíthatók a számítógépes rendszerben, az úgynevezett elektronikus térben, és be tudjuk azonosítani őket egy azonosítóval (identifier, rövidítése: id). A tématerképben belül egy témát megtalálhatunk az azonosítója alapján, például a Winston-Salem State University témát: <topic id="wssu"> azonosítóval. Ilyenkor nem adjuk meg a helyét, hanem rá

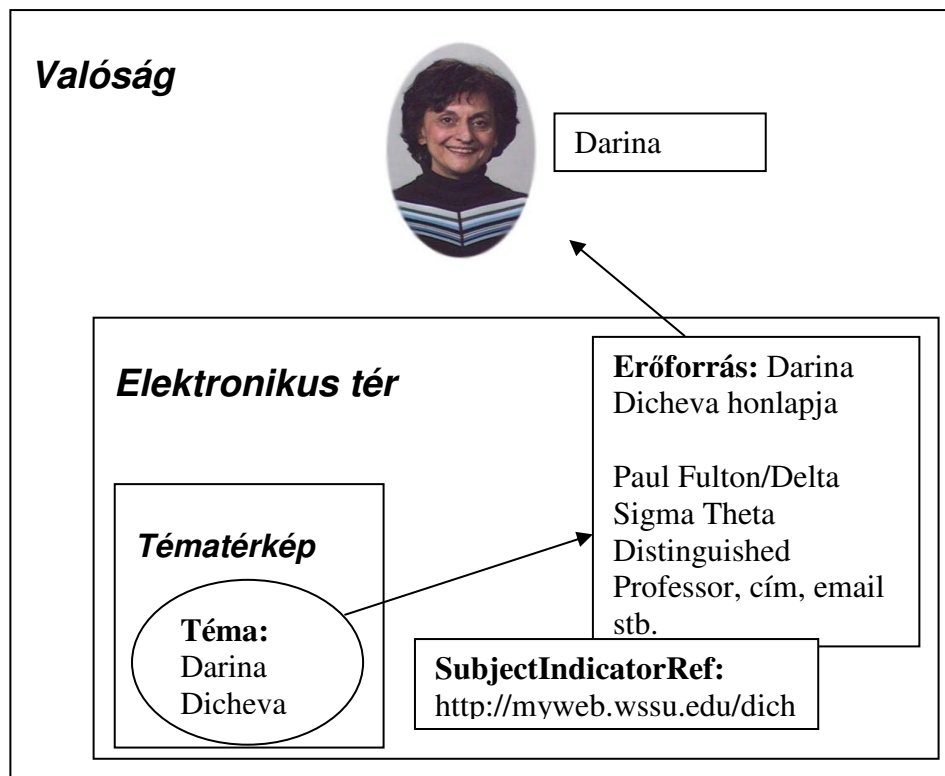
kell keresni az azonosítóra. Ekkor úgynevezett **URN (Uniform Resource Name)**, egységes forrás név megadásával jelöltük a témát. Amennyiben a témához, az elektronikus térben lévő külső erőforrás tartozik, például egy honlap – <http://myweb.wssu.edu> –, úgy azt **URL (Uniform Resource Locator)**, egységes forrás-azonosítóval címzünk, mely az erőforrás webcímét, az erőforrás helyét is megmutatja. Az URN és az URL összefoglaló neve: **URI (Uniform Resource Identifier)**, egységes erőforrás-azonosító.

Az elektronikus térben tárgyiasult fogalmakat – témákat – **erőforrásnak** (resources) nevezzük. A TM elsősorban az erőforrásokkal foglalkozik. Pontosabban bevehetünk a témák közé olyan valós világban létező fogalmakat, melyek nincsenek az elektronikus térben, például: Darina Dicheva, de ezt a TM4L szerkesztőben jelezni kell egy tárgyi azonosító (Subject Indicator) megadásával. A tématerképbe, az XTM fájlba, bekerül egy subjectIdentity XTM elem, illetve ezen belül a subjectIndicatorRef tartalmazza Darina honlapját:

xlink:href="<http://myweb.wssu.edu/dichevad/>"

Az egységes tárgyi azonosítók használatával lehetővé válik az azonos fogalmakat tartalmazó tématerképek összekapcsolása (merge). A tárgyi azonosítók egységesítésére tett törekvések egyike az Ontopia cég azonosító keresője (<http://psi.ontopedia.net>), de ezt a szerepet betöltheti egy intézményi honlap, vagy egy Wikipédia oldal is. (Dichev, C. – Dicheva, D. – Fischer, J. 2007)

### *A tárgyak identitása a tématerképekben*



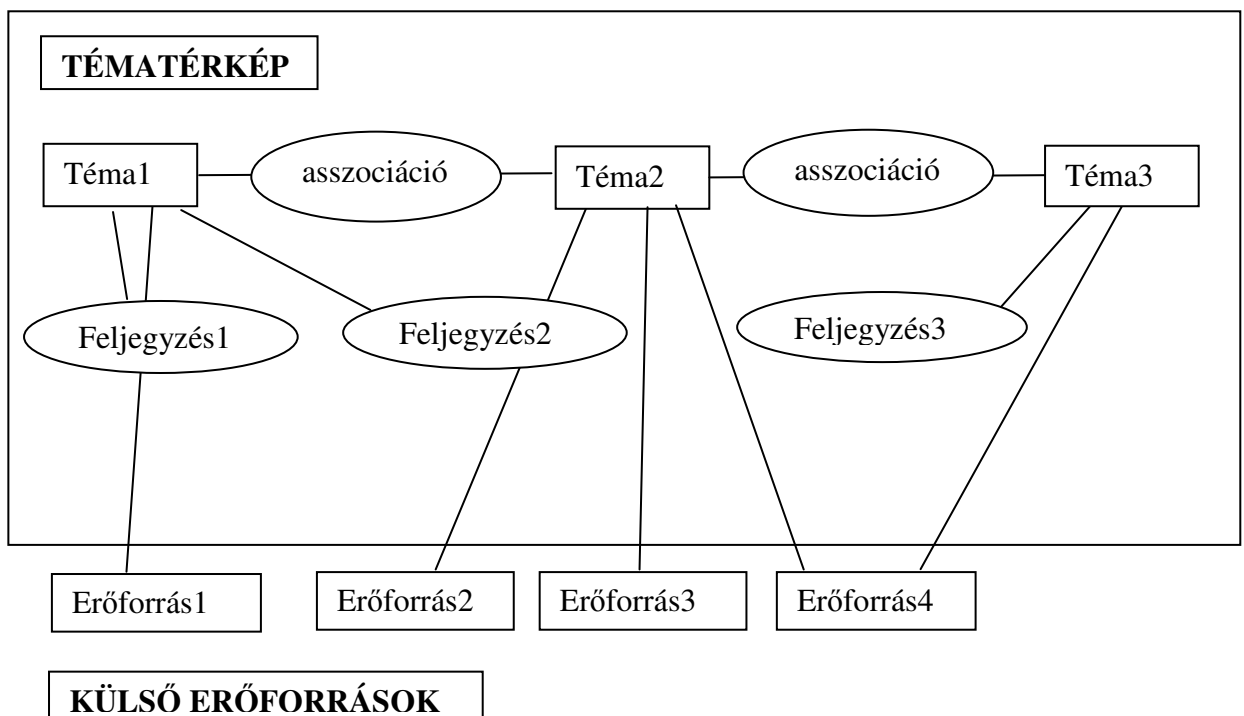
1. ábra

Visszatérve a tématerképek másik építőkövére, az asszociációra, azt mondhatjuk, hogy az asszociáció két téma kapcsolatát mutatja. Pl. Kovács Zoltán (az egyik téma) illetve a KRIPTO Kft (a másik téma) munkaviszonyban van egymással (az asszociáció típusának neve: munkaviszony). A munkaviszonynak két szerepét különböztethetjük meg: egyik szerep a munkaadói, másik szerep a munkavállalói. Ha a KRIPTO Kft van előtérbe állítva, akkor azt mondjuk: A KRIPTO Kft munkaadója Kovács Zoltánnak. Ezt úgy is mondhatjuk: Kovács Zoltán munkaadója a KRIPTO Kft.

Ha Kovács Zoltánt állítjuk előtérbe, akkor azt mondjuk: Kovács Zoltán munkavállalója a KRIPTO Kft-nek. Ezt úgy is mondhatjuk: A KRIPTO Kft munkavállalója Kovács Zoltán. A TM4L szerkesztőben az associations szó helyett a relationships szó található, de a fordítás során visszatértünk az eredeti szóra, az asszociációkra.

A Tématerképek harmadik építőköve, az előfordulás (occurrence), a téma kapcsolódását mutatja a címezhető információforráshoz, röviden erőforráshoz. Az erőforrás lehet egy karakterlánc (string), vagy egy állomány. A TM4L szerkesztőben nem jelenik meg az előfordulás (occurrence) kifejezés, hanem az erőforrás szó, illetve az erőforrás típusának neve kifejezést használja. A szerkesztő által képzett XTM fájlban természetesen az occurrence szó jelöli a téma és az erőforrás közötti kapcsolódást.

### *A tématerkép technológia általános modellje*



2. ábra

## 4. A SZEMANTIKUS ADATOK MEGJELENÉSE AZ ADATKEZELÉSBEN

Az adatkezelés az elektronikus adatfeldolgozás elindulásakor, az 1950-es években, jelentés nélküli számsorok, karaktorsorok bevitelét és kezelését jelentette a számítógépben. Kezdetben az adatok számok voltak, majd megjelentek a karakterláncok, a logikai rekordok és a fájlok. A számítógépben lévő adatok jelentése a programozó fejében volt. Az adatok jelentését idegen szóval az adatok szemantikájának nevezzük. A programozó tudta, hogy például az input adatként bevitt 168-as szám, Szabó Péter 1969. november havi, ledolgozott óráinak a számát jelenti. A programozó tudta azt is, hogy arról a Szabó Péterről van szó, aki a szervezési osztály dolgozója. A 168-as szám értelmezéséhez szükséges, úgynevezett szemantikus adatok, a programozó fejében voltak. Az adatfeldolgozás során képzett kimutatások fejlécében, az értelmezés megkönnyítése érdekében néhány szemantikus adat kiíratásra került. A táblázatkezelők megjelenésével, néhány szemantikus adat bekerült a táblázat első soraiba, az úgynevezett fejlécekbe, illetve a táblázat első oszlopába. Ezáltal az elektronikus térbe bekerültek szemantikus adatok, például mezőnevek.

Ezen túlmegegy az adatbázis séma azáltal, hogy az adatok, adattáblák jelentéseit is beírhatjuk. A szemantikus adatok uralkodóvá válása a tématerképekben történt meg. Egy jelentés nélküli karaktorsorhoz (string), például egy e-mail címhez: [thmate@oszk.hu](mailto:thmate@oszk.hu) megadjuk, hogy ez egy „e-mail cím”. Megadjuk, hogy ez az e-mail cím Tóth Mátéé, aki a Személy topic típusba tartozik. Tóth Mátéhoz kapcsolatosan megadjuk a honlapjának címét:

[http://www.ki.oszk.hu/107/e107\\_plugins/deptdir/deptdir.php?0.show.2.46](http://www.ki.oszk.hu/107/e107_plugins/deptdir/deptdir.php?0.show.2.46), megadhatjuk a munkahelyét stb.

A szemantikus adatszerkezetek iránti igény az elektronikus adatfeldolgozás kezdetétől jelen volt, de a kutatásoknak a weben lévő nagy tömegű de gyenge felhasználhatóságú adat megjelenése adott lendületet, és így jutottunk el a szemantikus webig. (Passin, T. B. 2004)

## 4. A TM4L SZERKESZTŐ ÉS VIZUALIZÁLÓ

Ebben a fejezetben mutatjuk be a TM4L szerkesztőt, és a vizualizáló programját, egy mintapéldán keresztül. A mintapélda egy strukturálatlan adatokat tartalmazó menedzseri notesz tématerképben. (Blumberg, R. – Atre, S: 2003) A menedzseri notesz, a menedzser által fontosnak tartott feljegyzéseket tartalmazza. Ilyen feljegyzések például:

- Iktatási feljegyzések fontosabb dokumentumokról,
- Elérhetőségi adatok személyekről és szervezetekről (név, cím, telefon, stb.),
- Azonosító adatok illetve adat sűrítmények kiemelt objektumokról (vagyontárgyakról, partnerekről, projektekről),
- Feljegyzések eseményekről, ezen belül például gazdasági eseményekről,
- Feljegyzések határidős feladatokról és tevékenységekről, kiindulva a személyes illetve szervezeti célokból és lezárva a napi tervekkel,

- Ötletek feljegyzése,
- Szakterületi fogalmak feljegyzése. (Kovács – Lencse, 2007)

A TM4L szerkesztőbe első lépésben a szerkesztendő tématerképről, esetünkben a Kovács Zoltán menedzseri noteszéről (KZnotesz) írunk be adatokat: tématerkép neve, rövid leírás, a tématerkép készítője, létrehozás dátuma, stb. Ezt követően bevételre kerülnek a témák: Kovács Zoltán, KRIPTO Kft, a KRIPTO Kft. irattára, és ezen belül minden folder név, vevők, szállítók és ügyintézőik neve, vagyontárgyak neve, projektek neve, stb. Fontos témák az úgynevezett téma típusok, melyeknek a példányai is témaként kerülnek be. Például az Esemény, melynek egy példánya: Évzáró értekezlet, vagy a Feladat, melynek egyik példánya: Felkészülés az évzáró értekezletre.

A témák bevételét követően az asszociációkat visszük be. Esetünkben a kiemelt asszociációk: (Garshol, L. M. 2004)

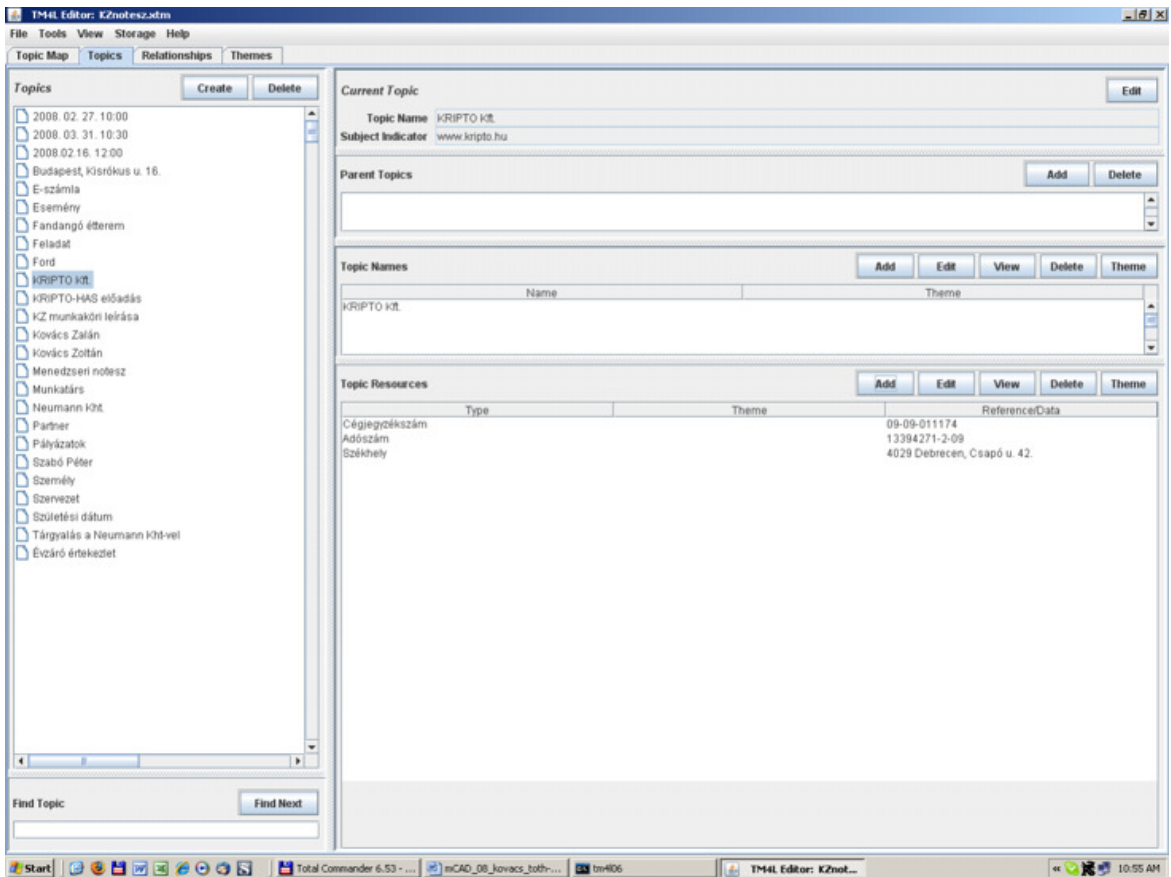
- Osztály-példány viszony, szerepei: osztálya illetve példánya. Az egyes osztályokhoz tartozó példányok száma nagy, ezért nem általában beszélünk osztály-példány viszonyról, hanem osztályonként külön-külön képzünk egy asszociációt. Például. Esemény-példány viszonyról beszélünk, melynek szerepei: esemény típusa és esemény példánya.
- Egész-rész viszony, szerepei: egésze illetve része. Ezt az asszociációt használtuk a KRIPTO Kft. irattárának felbontására. A legalsó szintű irattári elem, az úgynevezett dosszié, vagy levélfolder, osztályként tartalmazza a dokumentumokat, mint példányokat.
- Szuperosztály-alosztály viszony, szerepei: szuperosztálya illetve alosztálya.
- Munkaviszony, szerepei: munkaadó illetve munkavállaló.
- Tulajdonviszony, szerepei: tulajdonos illetve tulajdon.
- Esemény-időpont viszony, szerepei: eseménye illetve időpontja.
- Apa-fiú viszony, szerepei: apja illetve fia.

Az asszociációk bevételével párhuzamosan bevételre kerülnek a szerepek, és a szerepekhez tartozó konkrét témák. Például, a munkaadóhoz kötjük a KRIPTO Kft-t, és a munkavállalóhoz kötjük Kovács Zoltánt. (Pepper, S. 2002)

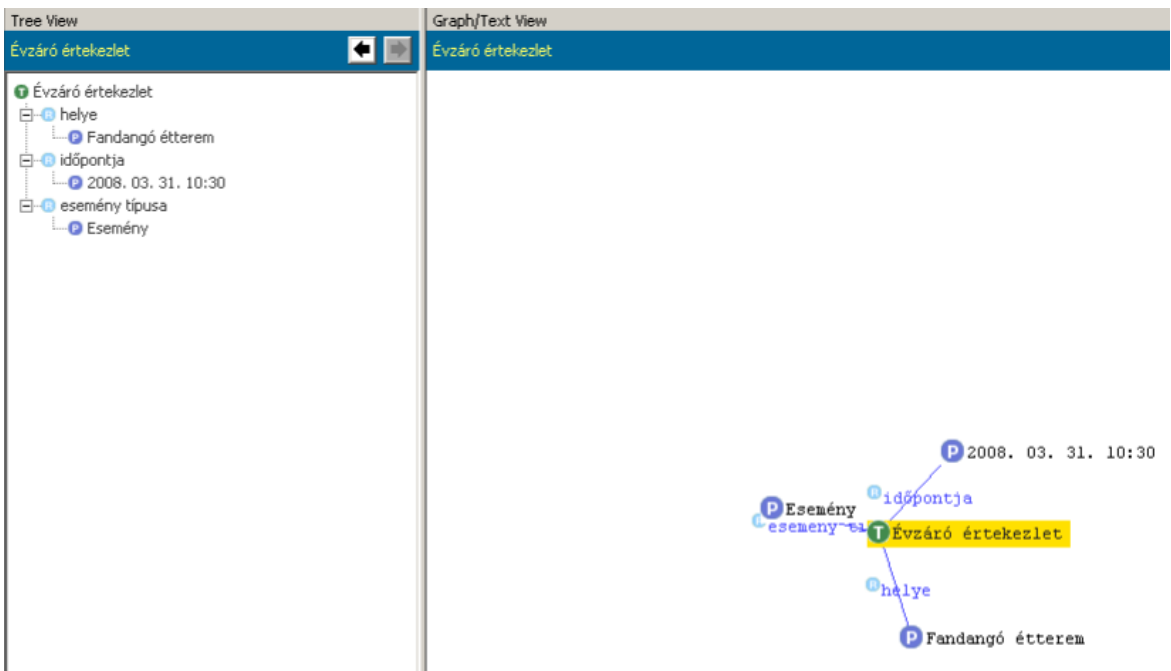
A témákhoz azon túl, hogy megadjuk azokat a témákat, melyekkel asszociációba tesszük, megadhatunk erőforrásokat. Például, a KRIPTO Kft témához, megadhatunk egy honlap címet, mint külső erőforrást. Belső erőforrásként megadhatjuk a székhelyét, az adószámát, a cégjegyzékszámát, és még sorolhatnánk a fontosabb adatait.

A tématerkép szerkesztése munkaigényes feladat, hiszen miközben az adatbevételt végezzük, egy ontológiát is kiépítünk az adott szakterületen. A szerkesztésbe és az ontológia kiépítésébe befektetett munkánk, a tématerkép használata során térül meg azáltal, hogy gyorsan megtalálhatjuk a keresett információt, akár a belső akár a külső erőforrások között.

Az alábbi ábra egy TM4L szerkesztő képernyő képe, és alatta a tématerkép egy vizualizációja.



3. ábra



4. ábra

## 5. BEFEJEZÉS

Tanulmányunkban a TM4L tématerkép szerkesztő és vizualizáló program fordításához kapcsolatosan mutattuk be a topic map technológia alapvető fogalmait, felhasználási lehetőségeit. A tématerképek használata feltételezésünk szerint nagy jövő előtt áll, elsősorban az oktatás területén. Gondolataink tématerképbe tétele nehéz feladat, hiszen háromtagú mondatokban kell leírni azt, ami a fejünkben van. A tématerképbe került tudásanyagot, az egyértelműség jellemzi leginkább, és ez adja a legnagyobb értékét. Beszélgetéseinkben, ahogy mondani szoktuk, félszavakból is értjük egymást. A tématerkép szerkesztő nem enged meg ilyen félszavakat, sőt még a teljes szavakat is bővíteni kell, ha más témátípusban, vagy témakörben beszélünk ugyanarról a fogalomról.

## IRODALOM

Blumberg, R. – Atré, S: (2003) „The Problem with Unstructured Data”  
[http://www.dmreview.com/article\\_sub.cfm?articleId=6287](http://www.dmreview.com/article_sub.cfm?articleId=6287)

Dichev, C. – Dicheva, D. – Fischer, J. (2007) „How to Name It, How to Find It”  
16<sup>th</sup> International World Wide Web Conference Banff, Alberta, Canada, 8-12. May, 2007. [http://www2007.org/workshops/paper\\_133.pdf](http://www2007.org/workshops/paper_133.pdf)

Garshol, L. M. (2004) „Metadata? Thesauri? Taxonomies? Topic Maps! Making sense of it all” <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>

Kovács – Lencse (2007) „The Management of Unstructured Information  
Typification of Managerial Notebooks Elektronik Notes  
” MicroCAD 2007 Nemzetközi Tudományos Konferencia, Miskolc, 2007.  
Miskolci Egyetem, Konferencia kiadvány (Pp. 71-79)

Passin, T. B. (2004) „Explorer’s Guide to the Semantic Web” Greenwich, Manning

Pepper, S. (2002) „The TAO of Topic Maps Finding the Way in the Age of Infoglut” <http://www.ontopia.net/topicmaps/materials/tao.html>