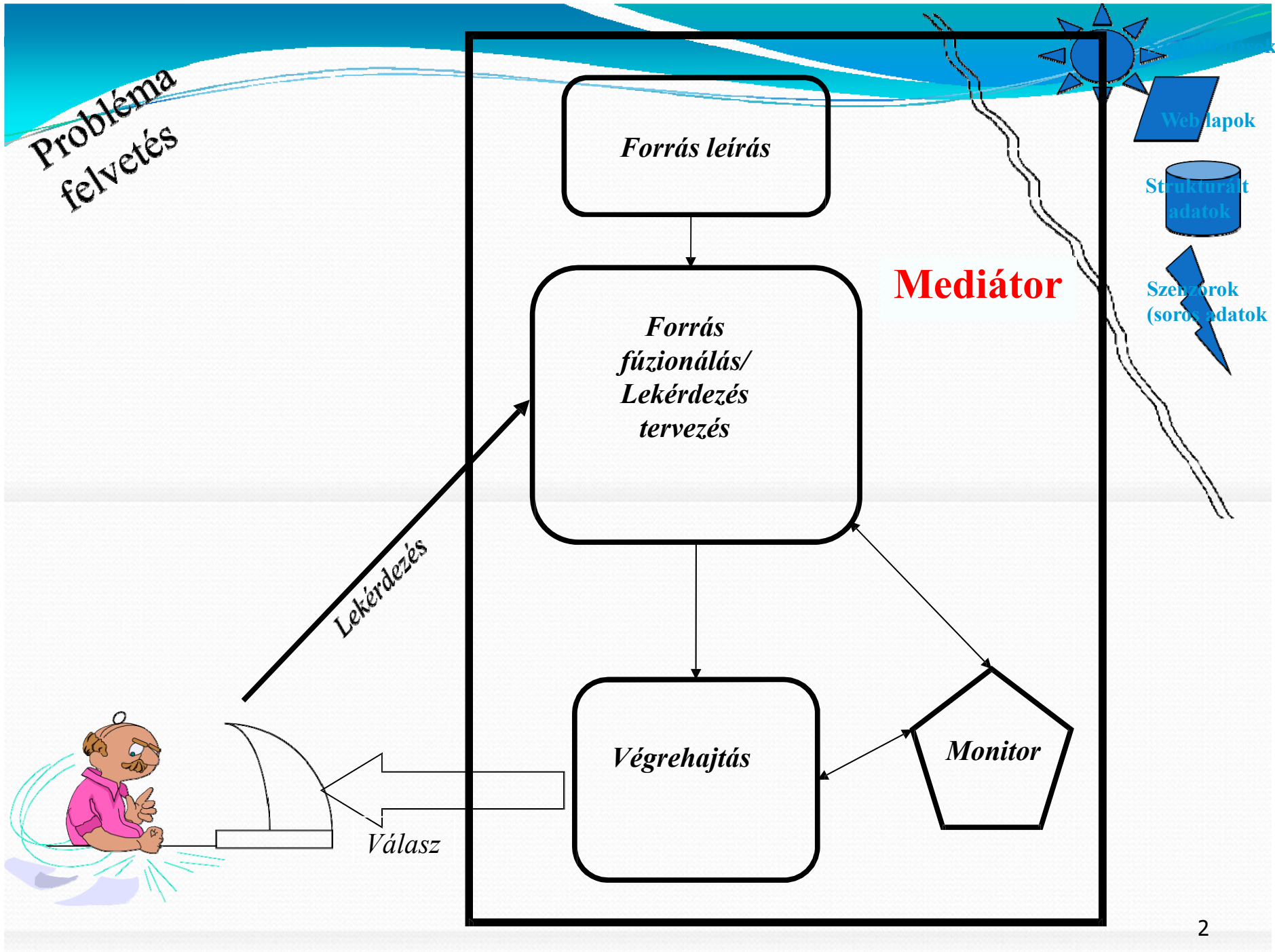


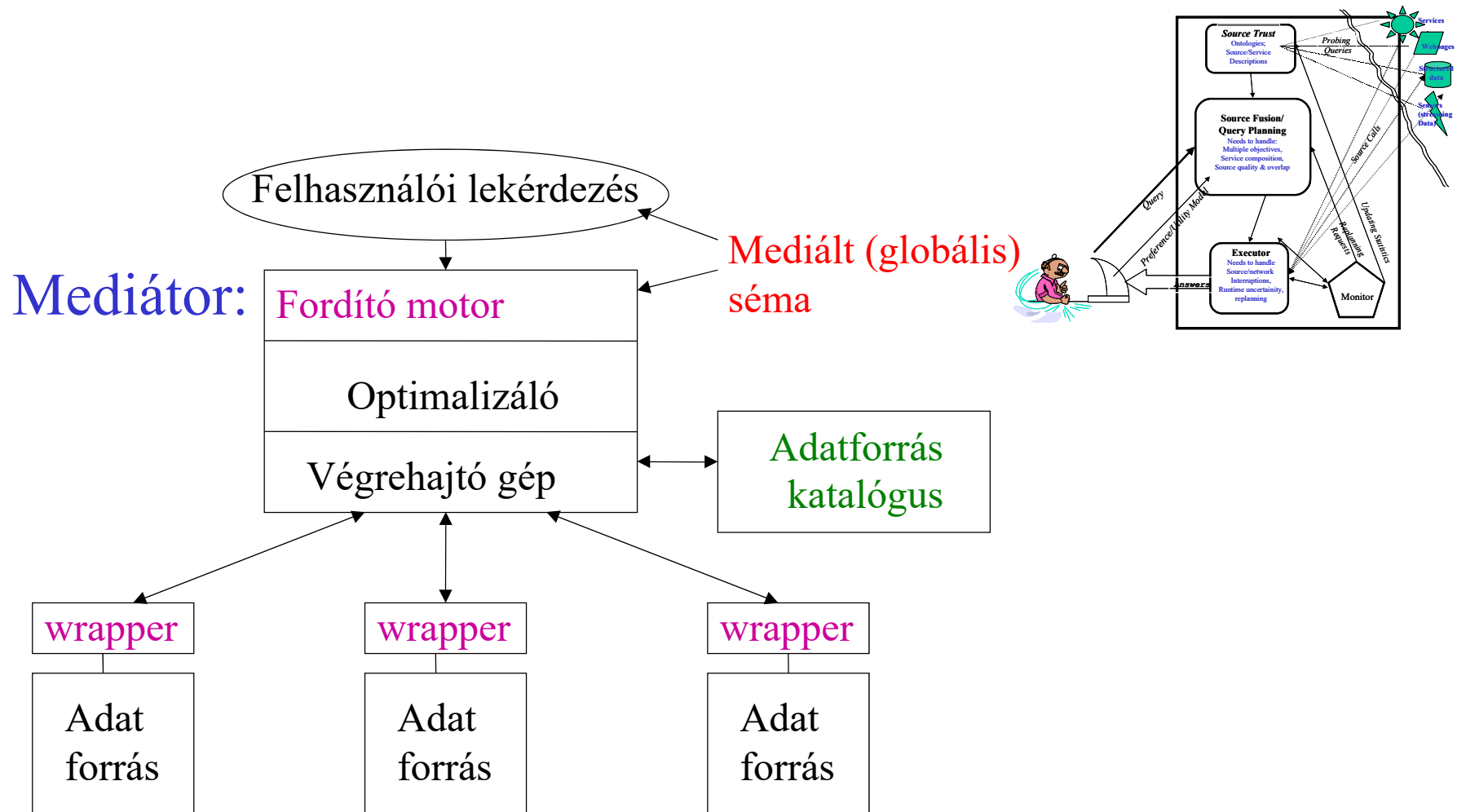
Integrációs és ellenőrzési technikák
VIMIAC04, 2019. tavasz

SZEMANTIKUS WEB

Méréstechnika és Információs Rendszerek Tanszék
<https://www.mit.bme.hu/oktatas/targyak/vimiaco4>



Virtuális integrátor architektúra



Források: relációs adatbázisok, weblapok, szövegek.

A web napjainkban

- Résztvevők dokumentumokat, adatokat publikálnak, URL címeket adva elérhetővé teszik az információkat
 - Kapcsolódás, hivatkozás linkekkel
- Integrációs próbálkozások: mashup oldalak (ad hoc)
 - Webszolgáltatások integrációja (eltérő API, logika, struktúra)
 - Adatgyűjtés kereső robotokkal (crawler programokkal)
 - Újra és újra felfedezzük a „kereket”
- Egészítsük ki a webet standard adatelérési módokkal: “Adatweb”

Az adatok webje

- Amire szükségünk van:
 - Publikáljuk adatainkat úgy, hogy felfedezhetőek legyenek a weben:
 - Standardizáljuk az adatok leírását, elérését
 - Dokumentumok eléréhez hasonló, de általános címezés: URI
 - Az URI-kal elérhető forrásokat kapcsoljuk
 - És engedjük a hálózat hatását érvényesülni, ahogy mi is böngészünk a weben...

- Példák: 2009 Semantic Technology Conference, San Jose, California, USA, June 15, 2009, Ivan Herman, W3C, ivan@w3.org

Böngészés a weben

Ivan Herman - Mozilla Firefox


File Edit View History Bookmarks Tools Help

http://www.w3.org/People/Ivan/

W3C

Ivan Herman

[My Work at W3C](#) | [Contact information](#) | [Short CV](#) | [Upcoming trips](#) | [Public presentations](#)



My Work at W3C

I am Semantic Web Activity Lead; that is my main work at W3C. I am member of [IW3C2](#) (International World Wide Web Conference Committee) (the committee coordinating the yearly WWW conference series), serving as a liaison for W3C, and of [SWSA](#) (Semantic Web Science Association), the committee responsible for the International Semantic Web Conferences series.

As part of my work, I also participate in lots of outreach activities, and I regularly make presentations, tutorials, etc. You can consult my [list of presentations](#) for further details.

Contact information

Email:
ivan@w3.org
(sha1sum: 5ac8022d5f0012aa1775ca2f001676befd5e80h)

Postal address:
[C/o Centre for Mathematics and Computer Sciences \(CWI\)](#)
[Kruislaan 413, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands.](#)

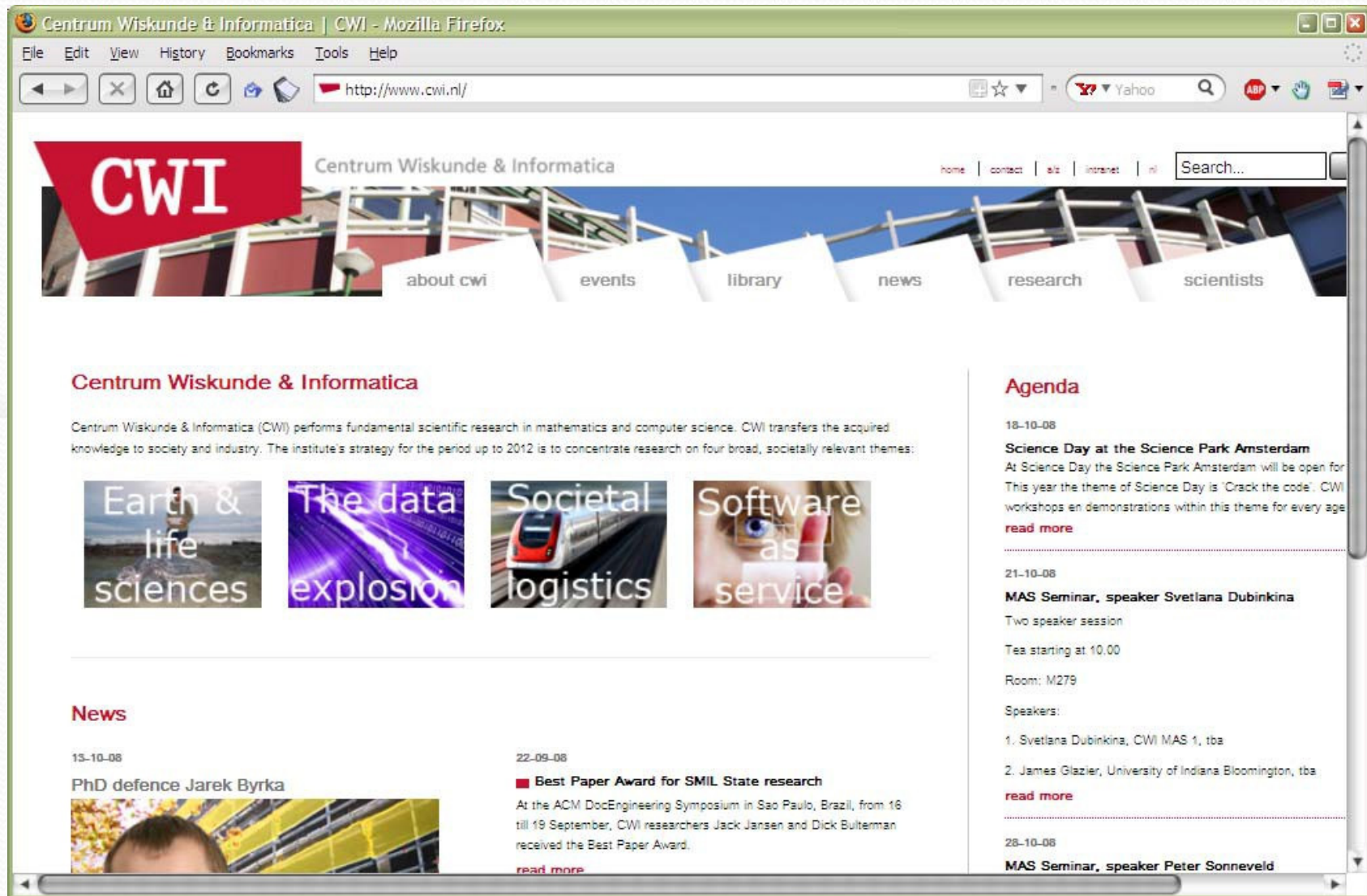
Phone numbers:
phone: +31-20-5924163
mobile phone: +31-641044153
fax: +31-20-5924312

PGP/GPG:
My GnuPGP key and signature is available on-line.

FOAF:
You can either extract a short FOAF information from this page, or consult my more complete, public FOAF file.

Misc:

... átugorva a munkahely linkre:



The screenshot shows the homepage of Centrum Wiskunde & Informatica (CWI) in a Mozilla Firefox browser window. The browser's address bar shows the URL <http://www.cwi.nl/>. The website features a red banner with the CWI logo and a navigation menu with links for 'about cwi', 'events', 'library', 'news', 'research', and 'scientists'. Below the banner, there is a section titled 'Centrum Wiskunde & Informatica' with a brief description of the institute's research focus. A grid of four images represents research areas: 'Earth & life sciences', 'The data explosion', 'Societal logistics', and 'Software as service'. A 'News' section on the left lists two recent articles: 'PhD defence Jarek Byrka' (dated 15-10-08) and 'Best Paper Award for SMIL State research' (dated 22-09-08). On the right, an 'Agenda' section lists upcoming events, including 'Science Day at the Science Park Amsterdam' (18-10-08) and 'MAS Seminar, speaker Svetlana Dubinkina' (21-10-08).

Centrum Wiskunde & Informatica

Centrum Wiskunde & Informatica (CWI) performs fundamental scientific research in mathematics and computer science. CWI transfers the acquired knowledge to society and industry. The institute's strategy for the period up to 2012 is to concentrate research on four broad, societally relevant themes:

- Earth & life sciences
- The data explosion
- Societal logistics
- Software as service

News

15-10-08
PhD defence Jarek Byrka

22-09-08
Best Paper Award for SMIL State research

At the ACM DocEngineering Symposium in Sao Paulo, Brazil, from 16 till 19 September, CWI researchers Jack Jansen and Dick Bulterman received the Best Paper Award.

Agenda

18-10-08
Science Day at the Science Park Amsterdam
At Science Day the Science Park Amsterdam will be open for This year the theme of Science Day is 'Crack the code'. CWI workshops en demonstrations within this theme for every age
[read more](#)

21-10-08
MAS Seminar, speaker Svetlana Dubinkina
Two speaker session
Tea starting at 10.00
Room: M279
Speakers:
1. Svetlana Dubinkina, CWI MAS 1, tba
2. James Glazier, University of Indiana Bloomington, tba
[read more](#)

28-10-08
MAS Seminar, speaker Peter Sonneveld

Weblapok értelmezése

- A felhasználó megértik, hogy a link egy munkahely honlapjára mutat
- Tudjuk értelmezni, hogy ez egy kutatóhely leírása
- *Ami hiányzik az „adatok webjének” építéséhez:*
 - *Jelentéssel kiegészített, értelmezhető linkek*

Tehát bővítsünk:

- Adjunk a linkekhez kiegészítő információt, címkézzük fel őket
- A címkék legyenek géppel értelmezhetőek
 - Kategorizálás
 - Esetleg következtetés

Adatok hálója

- Az adatok webje:
 - Használjunk URI-kat adatok és (nemcsak) dokumentumok publikálására
 - Kapcsoljuk össze az adatokat - linkek
 - Jellemezzük/osztályozzuk a linkeket információk hozzáadásával
 - Használjunk standard technológiákat
- *Ez a szemantikus web alapja*



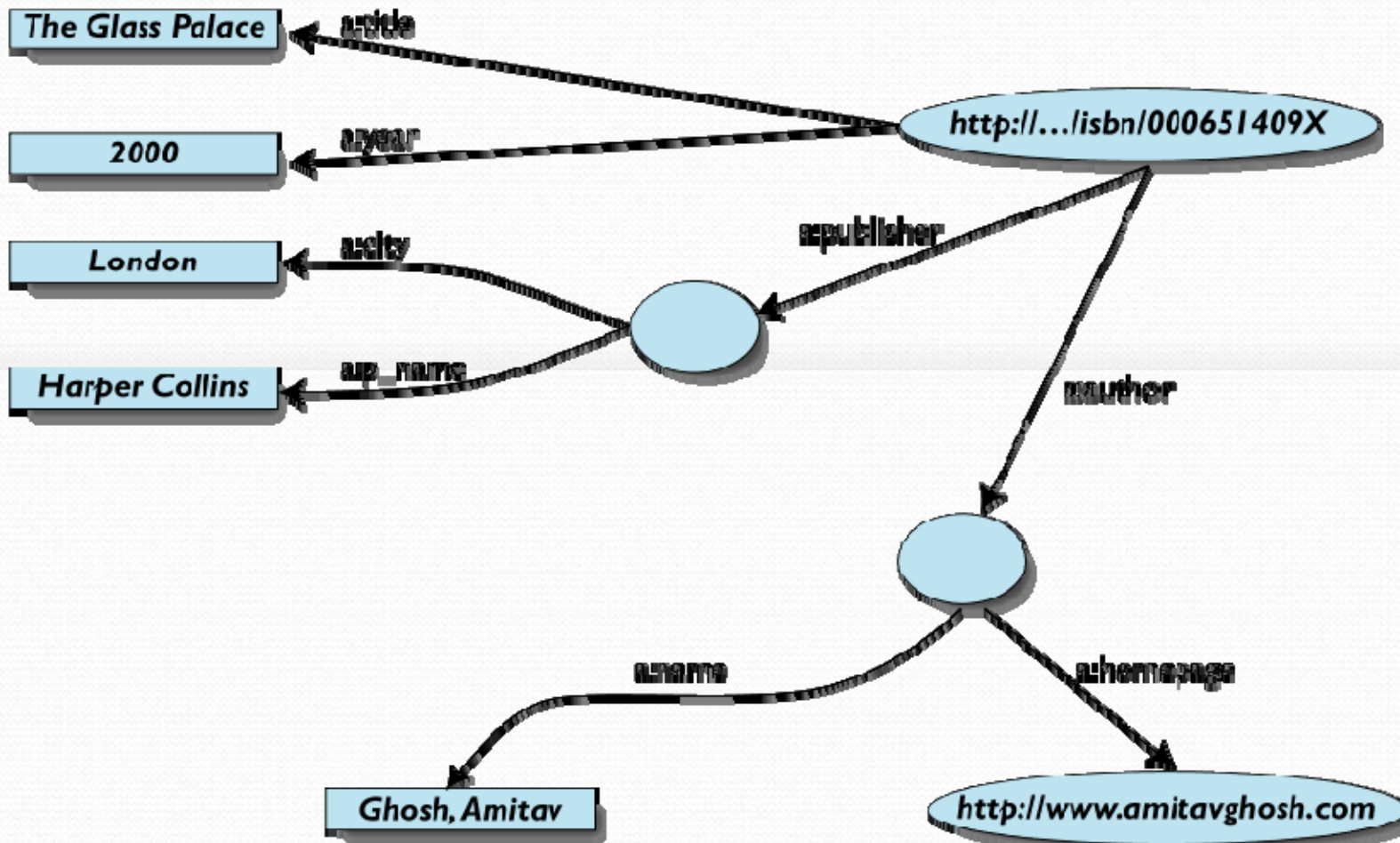
Példa: könyvesbolt adatai

ID	Author	Title	Publisher	Year
ISBN0-00-651409-X	id_xyz	The Glass Palace	id_qpr	2000

ID	Name	Home Page
id_xyz	Ghosh, Amitav	http://www.amitavghosh.com

ID	Publ. Name	City
id_qpr	Harper Collins	London

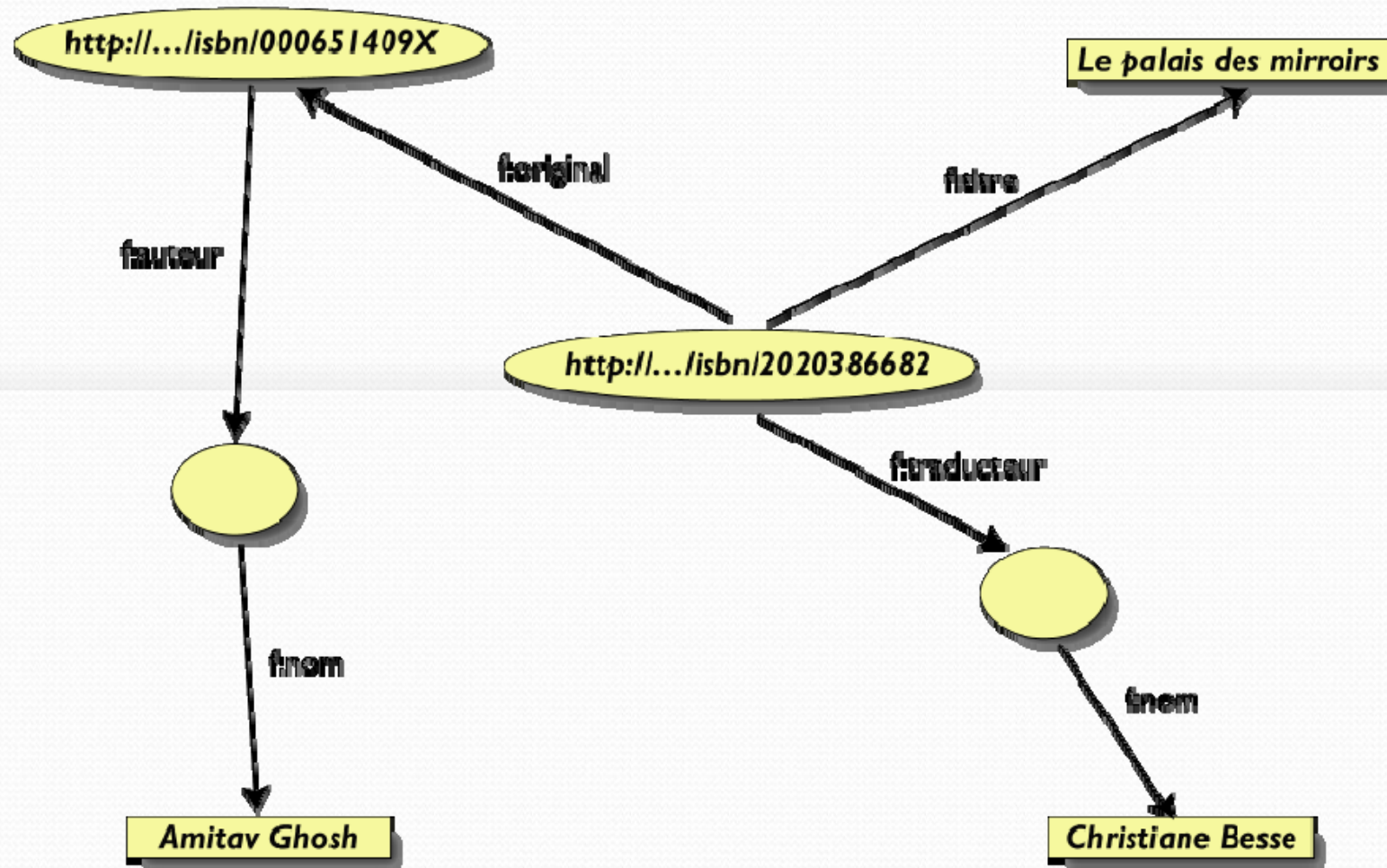
Adatok exportálása relációkként



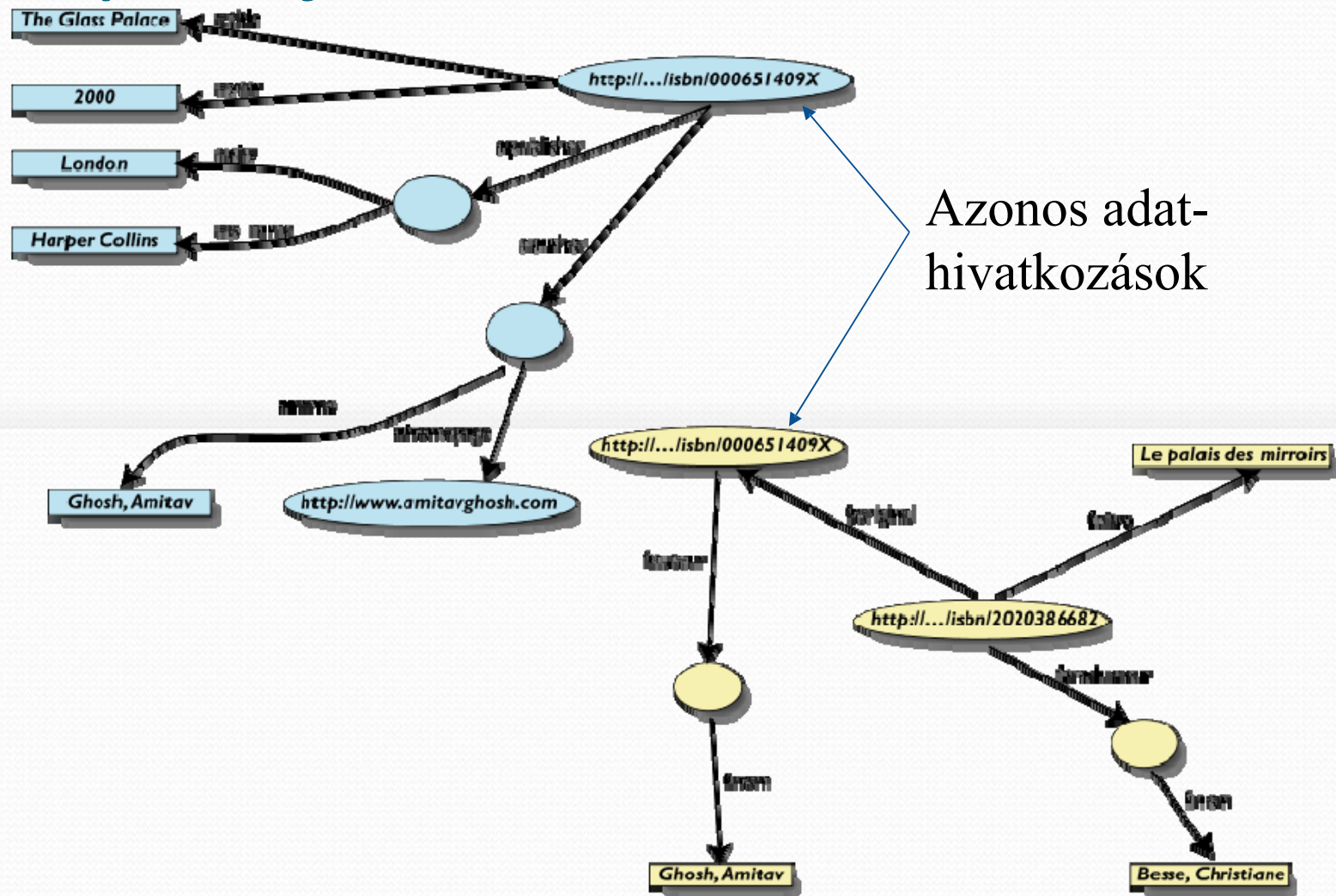
Egy másik könyvesbolt adatai

	A	B	D	E
1	ID	Titre	Traducteur	Original
2	ISBN0 2020386682	Le Palais des miroirs	A13	ISBN-0-00-651409-X
3				
6	ID	Auteur		
7	ISBN-0-00-651409-X	A12		
11	Nom			
12	Ghosh, Amitav			
13	Besse, Christianne			

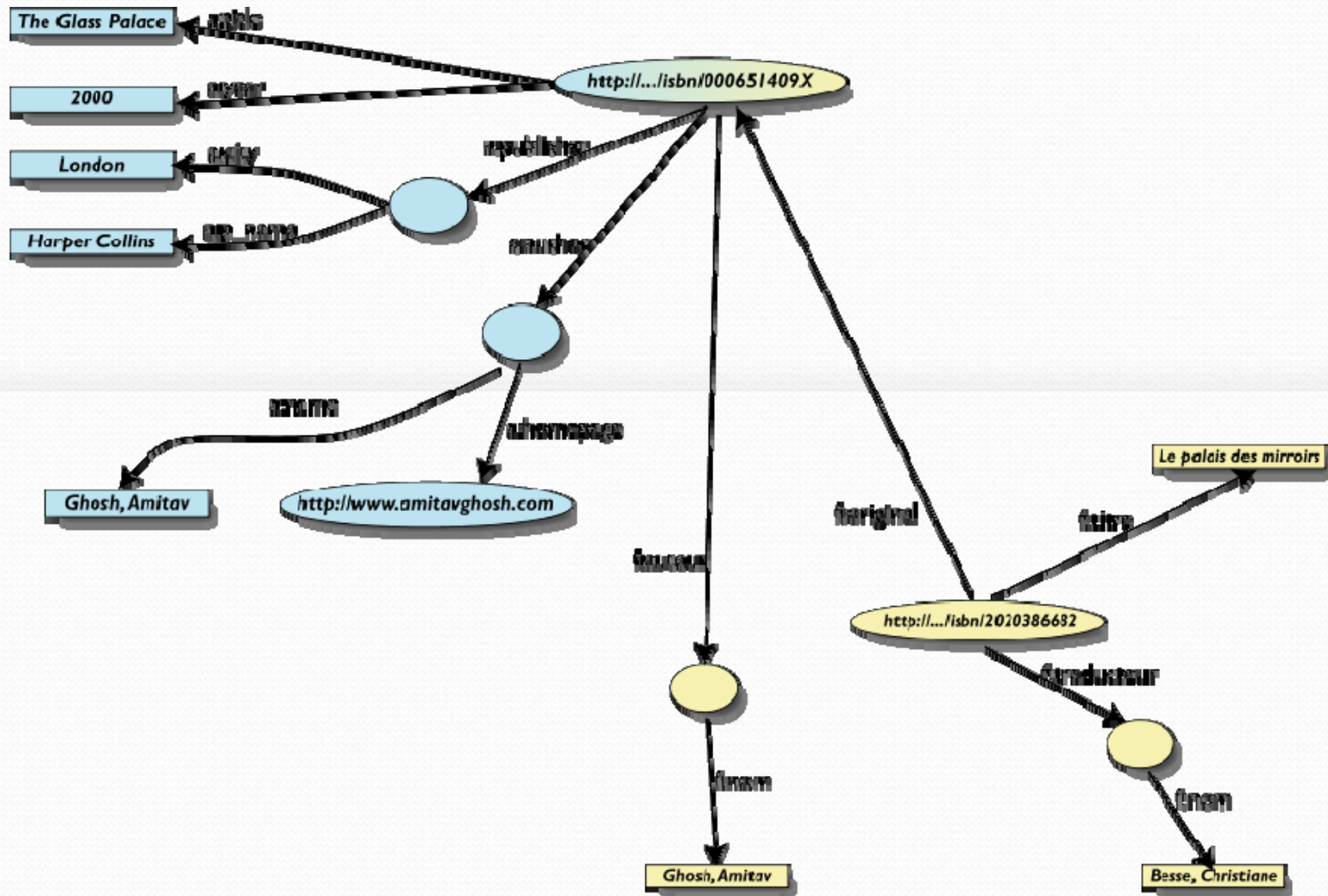
A második könyvesbolt adatainak exportja



Kapcsoljuk össze az adatokat

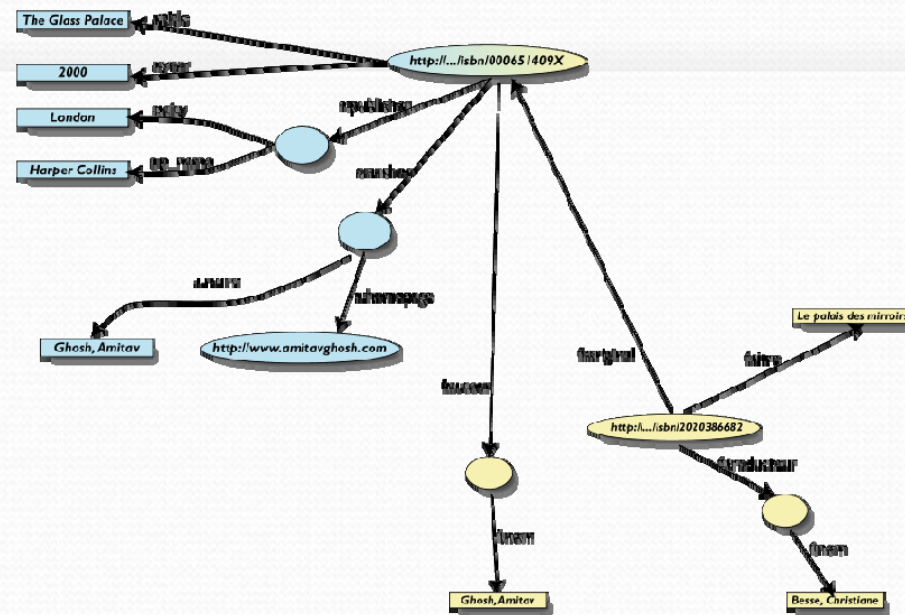


Amennyiben identikusak az elemek:



Írjunk lekérdezéseket a kapcsolatokon keresztül:

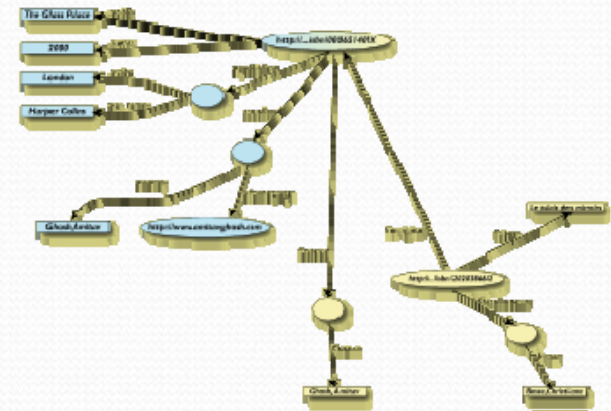
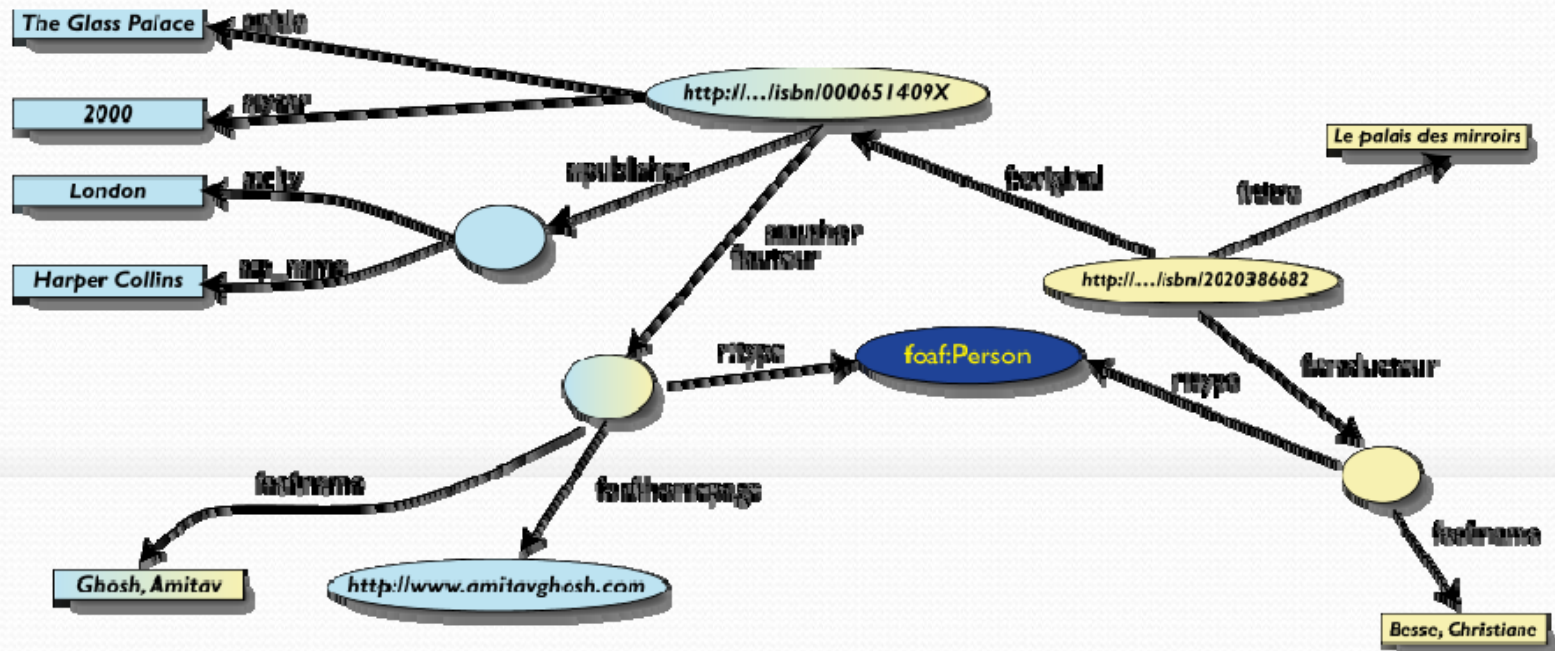
- Az első könyvesbolt adatai kiegészíthető például az eredeti könyvre vonatkozó információkkal



További kapcsolatok is felfedezhetőek...

- Vélhetően az **a : author** és az **f : auteur** azonos elemre mutat
- Automatikus összekapcsoláshoz: adjunk további információt a leíráshoz
 - **a : author** legyen azonos **f : auteur** erőforrással
 - Mindkettő személyt azonosít
 - Ilyen fogalmakat már a webes közösség definiált:
 - egy “Person” elem azonosítható a nevével és a honlapjával
 - Használjuk ezt kategóriaként

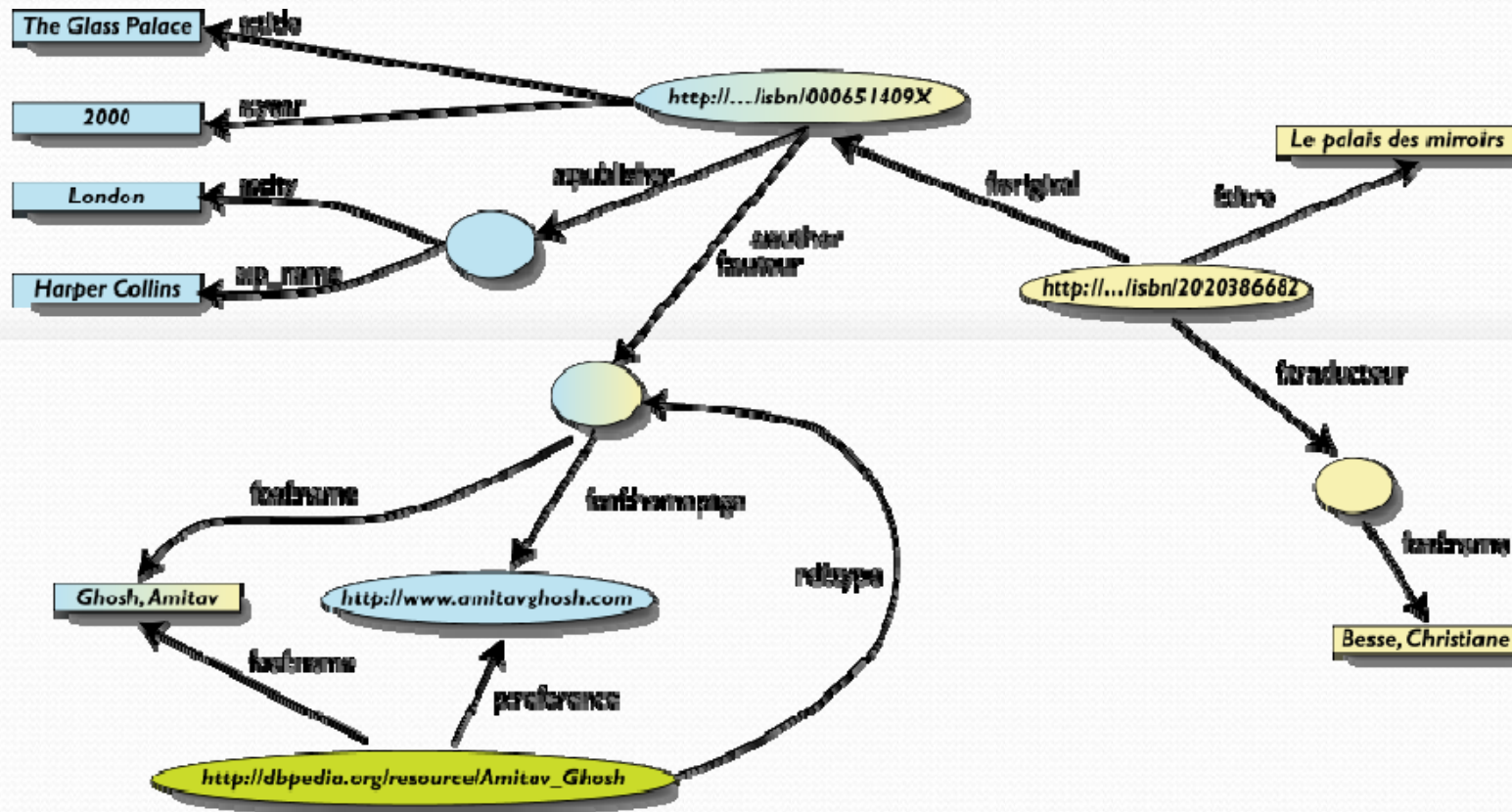
Adatháló kiegészíthető, lekérdezhető így:



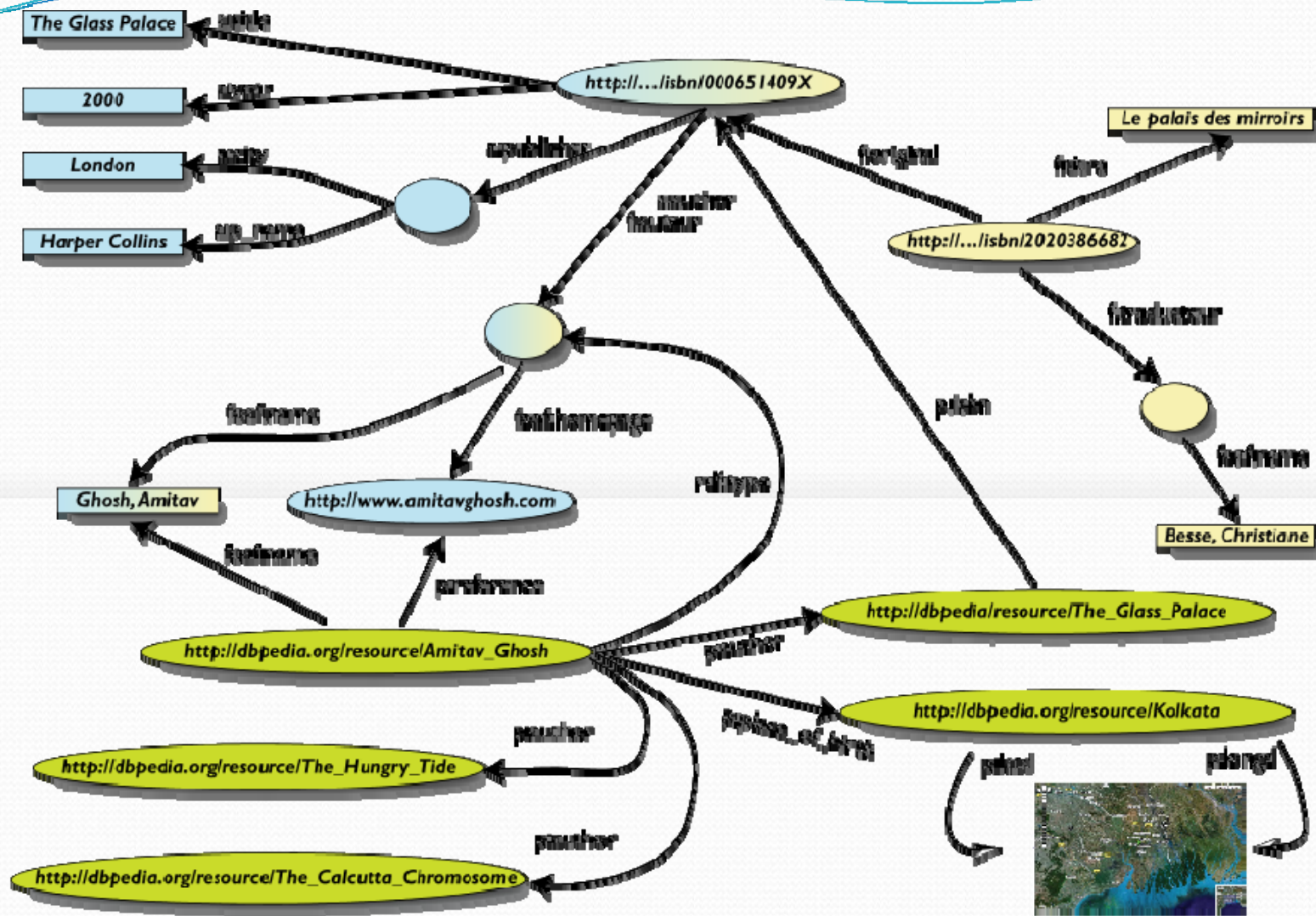
Bővítés újabb adathalmazokkal

- Például a “Person” típus esetén, használhatóak a Wikipédia adatai:
 - pl., a “[dbpedia](#)” projekt már feldolgozta a Wikipédia adatait

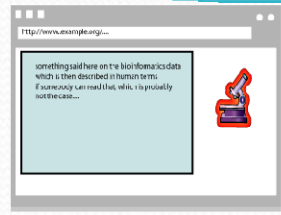
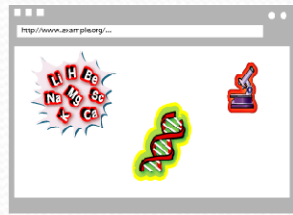
Összekapcsolás Wikipedia adatokkal



És így tovább...

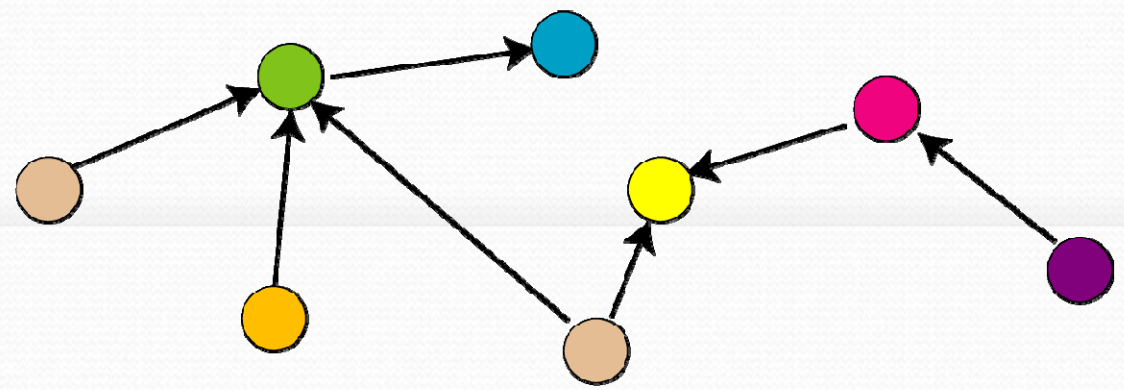


Az adatok webje kialakulófélben...



Alkalmazások

Lekérdezés,
adatmódosítás



Absztrakt adatstruktúra

Leképezés,



Forrás adatok különböző formátumokban

Szemantikus Web

Semantic Web

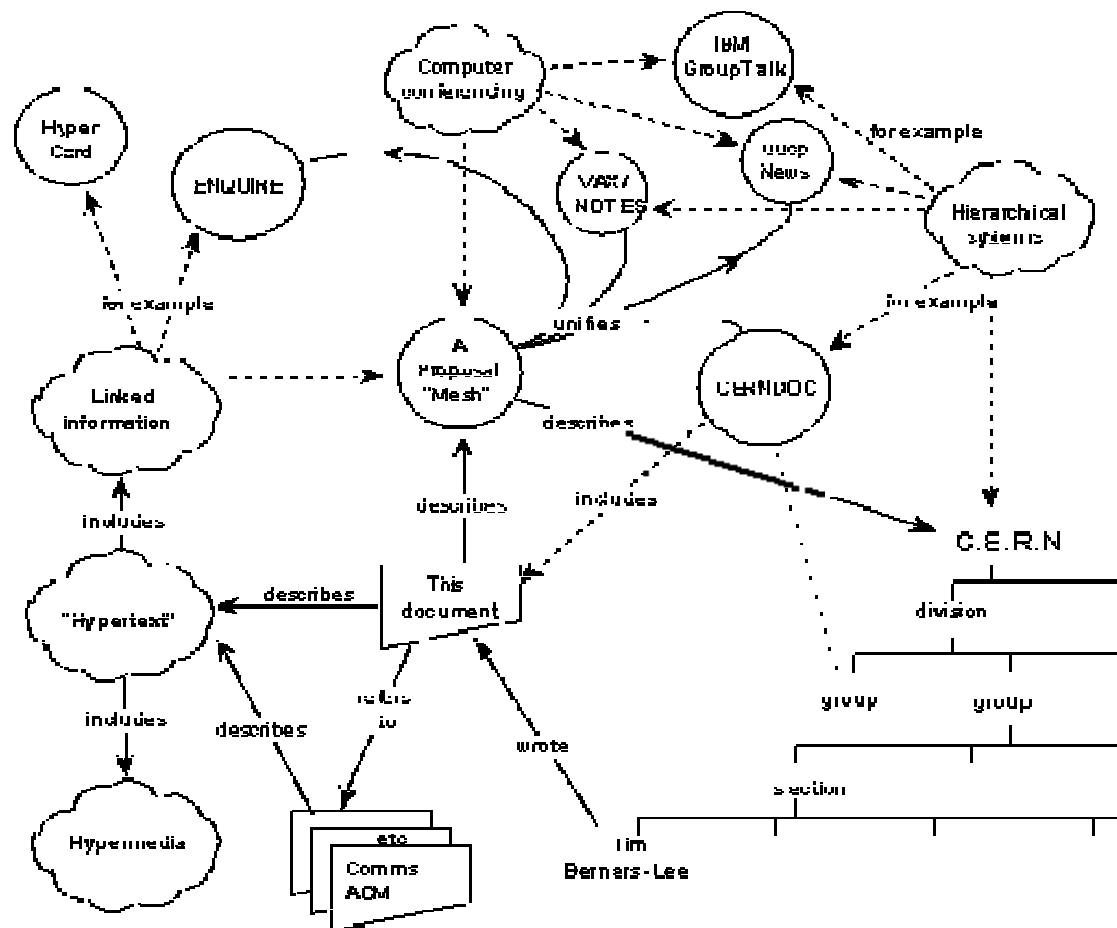
- A szemantikus web alkalmas megközelítés, illetve megfelelő nyelvekkel, eszközökkel támogatja az intelligens információs rendszerek fejlesztését az elosztott információs környezetben.
- A SzW alapja a hagyományos web hálózat, így egyáltalán nem nyilvánvaló, hogy alkalmas a feladatra.
- A SzW technológia lehetőséget teremt az ágens alapú intelligens megoldások felhasználására a web területen.



Egy IBM 360 számítógép (1971)

A Szemantikus Web eredete

- Tim Berners-Lee eredeti 1989-es WWW javaslata a Web-et információ menedzselő funkciókkal ellátott objektumok kapcsolataiként jellemzi.
- Guha designed MCF at Apple (~94)
- XML+MCF=>RDF (~96)
- RDF+OO=>RDFS (~99)
- RDFS+KR=>DAML+OIL (00)
- W3C's SW activity (01)
- W3C's OWL (02?)



<http://www.w3.org/History/1989/proposal.html>

W3C szervezet célkitűzései

- Megközelítés – számítógépek jobb kihasználtságának biztosítása:

„A szematikus web egy kiterjesztése a jelenlegi web-nek, amelyben az információknak jól definiált jelentést adhatunk, lehetővé téve a gépek és felhasználók jobb együttműködését..” -- Berners-Lee, Hender and Lassila, The Semantic Web, Scientific American, 2001

- A jelenlegi web *tárol* dolgokat, míg a szemantikus web képes *működtetni* dolgokat.

Mit működtet a szemantikus web?

- **Fogalom-alapú keresést**
≠ kulcsszó alapú keresés
- **Szemantikus navigáció**
≠ link alapú navigáció
- **Személyes megjelenés biztosítása**
≠ egységes megjelenés
- **Lekérdezés megválaszolása**
≠ dokumentum beszerzés
- **Szolgáltatások**
≠ CGI hívások, de nincs szolgáltatás leírási eszközkészlet, elemzés, kompozíció, stb.

Miért nehéz ez a feladat?

(példa az eredeti javaslat alapján)

Így látja a gép a web oldalt

林克昌 根留台灣 可能增高

在愛戴者熱心奔走之下，華裔指揮家林克昌根留台灣的可行性又提升了幾分。兩廳院主任李炎、國家音樂廳樂團副團長黃奕明日前親赴林克昌、石聖芳寓所拜會，並提出多場客席邀約。此外，台灣省立交響樂團團長陳澄雄也早早「下訂」，邀請林克昌赴台中霧峰，從八月十日起訓練省交，為期長達一個月。

在台灣諸多公家樂團中，陳澄雄是以實際行動表達對林克昌肯定的樂界人士之一，曾多次公開表示對林克昌指揮才華的欽佩，而且幾乎每個樂季都邀請林克昌客席演出。

此外，林克昌上個月赴俄羅斯與頂尖的「俄羅斯國家管絃樂團」灌錄了柴可夫斯基晚期三大交響曲以及「羅密歐與茱麗葉」、「斯拉夫進行曲」、「義大利隨想曲」，最後的DAT母帶也在前兩天寄回台灣。製作人楊忠衛與林克昌試聽之後，都對錄音效果—尤其音質表現感到相當滿意，楊忠衛估計呈現了七分林克昌指揮神韻。

俄羅斯國家管絃樂團首席布魯尼日前也讚譽林克昌的指揮藝術有三大特點：一是控制自如的彈性速度；二是強烈的動態對比；三是宛如呼吸歌唱的旋律處理。這些對錄音師而言都構成很大挑戰。俄國錄音師雖然採用多軌混音，但定位、場面都有可觀之處。

Természetes nyelv megértése, és különösen képi információk megértése igen nehéz.

HTML nem alkalmas a feladatra

Meg tudjuk fogalmazni a gép részére, hogy a különböző részek mit jelentenek?

林克昌 根留台灣 可能增高

name

education

在愛戴者熱心奔走之下，音樂指揮家林克昌根留台灣的可行性又提升了幾分。兩廳院主任李炎、國家音樂廳樂團副團長黃奕明日前親赴林克昌、石聖芳寓所拜會，並提出多場客席邀約。此外，台灣省立交響樂團團長陳澄雄也早早「下訂」，邀請林克昌赴台中響峰，從八月十日起訓練省交，為期長達一個月。

在台灣諸多公家樂團中，陳澄雄是以實際行動表達對林克昌肯定的樂界人士之一，曾多次公開表示對林克昌指揮才華的欽佩，而且幾乎每個樂季都邀請林克昌客席演出。

work

此外，林克昌上個月赴俄羅斯頂尖的「俄羅斯國家管絃樂團」灌錄了柴可夫斯基晚期三大交響曲以及「羅密歐與茱麗葉」、「斯拉夫進行曲」、「義大利隨想曲」，最後的DAT母帶也在前兩天寄回台灣。製作人楊忠衛與林克昌試聽之後，都對錄音效果—尤其音質表現感到相當滿意，楊忠衛估計呈現了七分林克昌指揮神韻。

CV

private

俄羅斯國家管絃樂團首席布魯尼日前也讚譽林克昌的指揮藝術有三大特點：一是控制自如的彈性速度；二是強烈的動態對比；三是宛如呼吸歌唱的旋律處理。這些對錄音師而言都構成很大挑戰。俄國錄音師雖然採用多軌混音，但定位、場面都有可觀之處。

XML használható?

Az XML hívei szerint jelentésekkel rendelkező tag-ek megoldják a problémát, de..

林克昌 根留台灣 可能增高

< name >

< education >

< work >

< private >

< CV >

在愛戴者熱心奔走之下，華裔指揮家林克昌根留台灣的可行性又提升了幾分。兩廳院主任李炎、國家音樂廳樂副團長黃奕明日前親赴林克昌、右聖芳寓所拜會，並提出多場客席邀約。此外，台灣省立交響樂團團長陳澄雄也早早「下訂」，邀請林克昌赴台中霧峰，從八月十日起訓練省交，為期長達一個月。

在台灣諸多公家樂團中，陳澄雄是以實際行動表達對林克昌肯定的樂界人士之一，曾多次公開表示對林克昌指揮水準的欽佩，而且幾乎每個樂季都邀請林克昌客席演出。

此外，林克昌上個月赴俄羅斯與頂尖的「俄羅斯國家管絃樂團」灌錄了柴可夫斯基晚期三大交響曲以及「羅密歐與茱麗葉」、「斯拉夫進行曲」、「義大利隨想曲」，最後的DAT母帶也在前兩天寄回台灣。製作人楊忠衛與林克昌試聽之後，都對錄音效果—尤其音質表現感到相當滿意，楊忠衛估計呈現了七分林克昌指揮神韻。

俄羅斯國家管絃樂團首席布魯尼日前也讚賞林克昌的指揮藝術有三大特點：一是控制自如的彈性速度；二是強烈的動態對比；三是宛如呼吸歌唱的旋律處理。這些對錄音師而言都構成很大挑戰。俄國錄音師雖然採用多軌錄音，但定位、場面都有可觀之處。

XML ≠ gépi megértés támogatás

De a tag-ek is ilyenek tünnek a gép számára....

林克昌 根留台灣 可能增高

< ναμε >

<εδουχαιον>

在愛戴者熱心奔走之下，華裔指揮家林克昌根留台灣的可行性又提升了幾分。兩廳院主任李炎、國家音樂廳樂團副團長黃奕明日前親赴林克昌、石聖芳寓所拜會，並提出多場客席邀約。此外，台灣省立交響樂團團長陳澄雄也早早「下訂」，邀請林克昌赴台中霧峰，從八月十日起訓練省交，為期長達一個月。

在台灣諸多公家樂團中，陳澄雄是以實際行動表達對林克昌肯定的樂界人士之一，曾多次公開表示對林克昌指揮才華的欽佩，而且幾乎每個樂季都邀請林克昌客席演出。

<ωορκ>

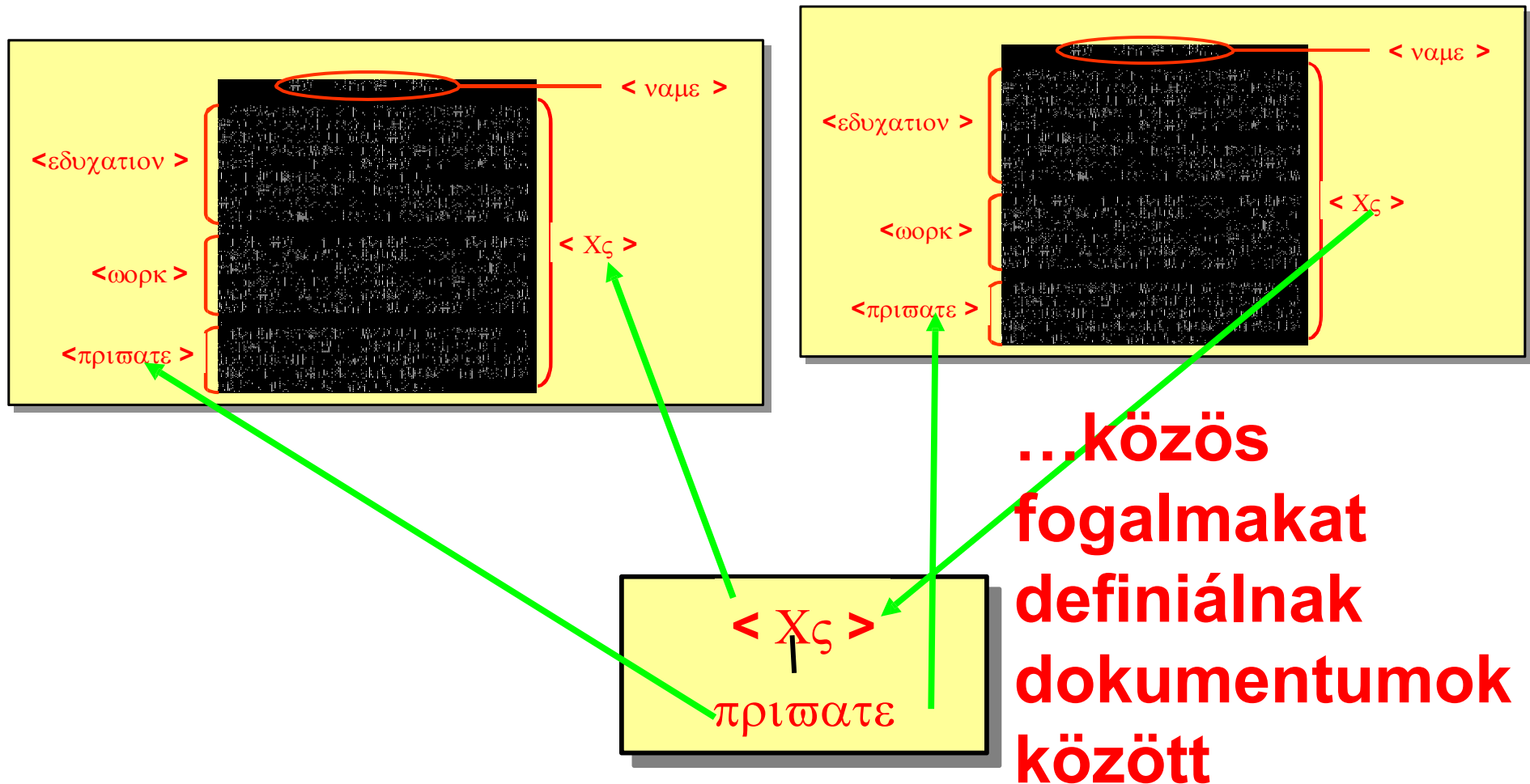
此外，林克昌上個月赴俄羅斯與頂尖的「俄羅斯國家管絃樂團」灌錄了柴可夫斯基晚期三大交響曲以及「羅密歐與茱麗葉」、「斯拉夫進行曲」、「義大利隨想曲」，最後的DAT母帶也在前兩天寄回台灣。製作人楊忠衡與林克昌試聽之後，都對錄音效果—尤其音質表現感到相當滿意，楊忠衡估計呈現了七分林克昌指揮神韻。

< Xς >

<πριωατε>

俄羅斯國家管絃樂團首席布魯尼日前也讚譽林克昌的指揮藝術有三大特點：一是控制自如的彈性速度；二是強烈的動態對比；三是宛如呼吸歌唱的旋律處理。這些對錄音師而言都構成很大挑戰。俄國錄音師雖然採用多軌混音，但定位、場面都有可觀之處。

A sémák alkalmazása előrelépés:



Eltérő sémák problémája

林克昌 根留台灣 可能增高

在愛戴者熱心奔走之下，華裔名指揮家林克昌根留台灣的可行性又提升了幾分。兩廳院主任李炎、國家音樂廳樂團副團長黃奕明日前親赴林克昌、石聖芳寓所拜會，並提出多場客席邀約。此外，台灣省立交響樂團團長陳澄雄也早早「下訂」，邀請林克昌赴台中霧峰，從八月十日起訓練省交，為期長達一個月。

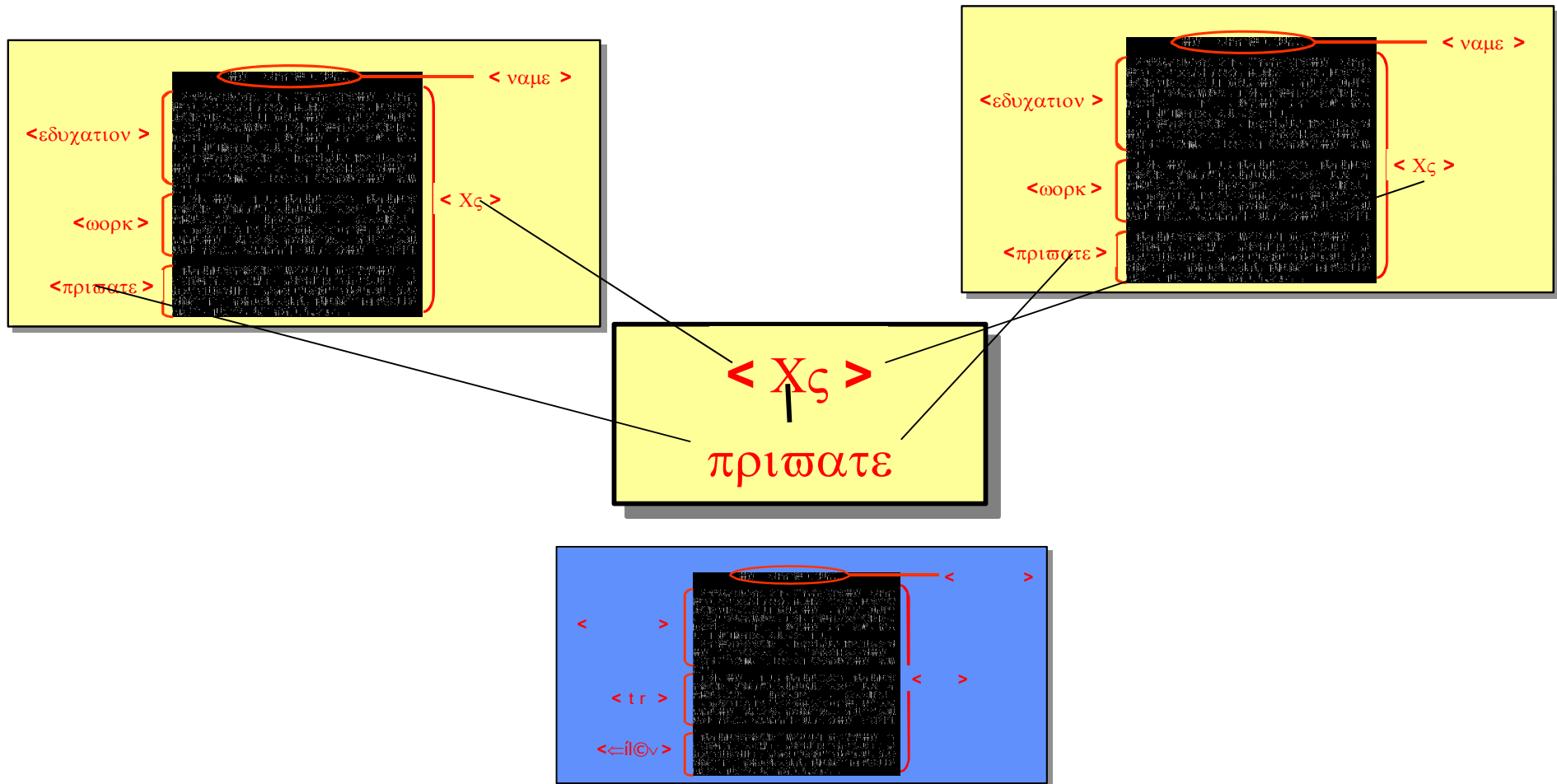
在台灣諸多公家樂團中，陳澄雄是以實際行動表達對林克昌肯定的樂界人士之一，曾多次公開表示對林克昌指揮才華的欽佩，而且幾乎每個樂季都邀請林克昌客席演出。

此外，林克昌上個月赴俄羅斯與頂尖的「俄羅斯國家管絃樂團」灌錄了柴可夫斯基晚期三大交響曲以及「羅密歐與茱麗葉」、「斯拉夫進行曲」、「義大利隨想曲」，最後的DAT母帶也在前兩天寄回台灣。製作人楊忠衛與林克昌試聽之後，都對錄音效果—尤其音質表現感到相當滿意，楊忠衛估計呈現了七分林克昌指揮神韻。

俄羅斯國家管絃樂團首席布魯尼日前也讚譽林克昌的指揮藝術有三大特點：一是控制自如的彈性速度；二是強烈的動態對比；三是宛如呼吸歌唱的旋律處理。這些對錄音師而言都構成很大挑戰。俄國錄音師雖然採用多軌混音，但定位、場面都有可觀之處。

Más számára mást jelenthet

A jelentés még mindig hiányzik

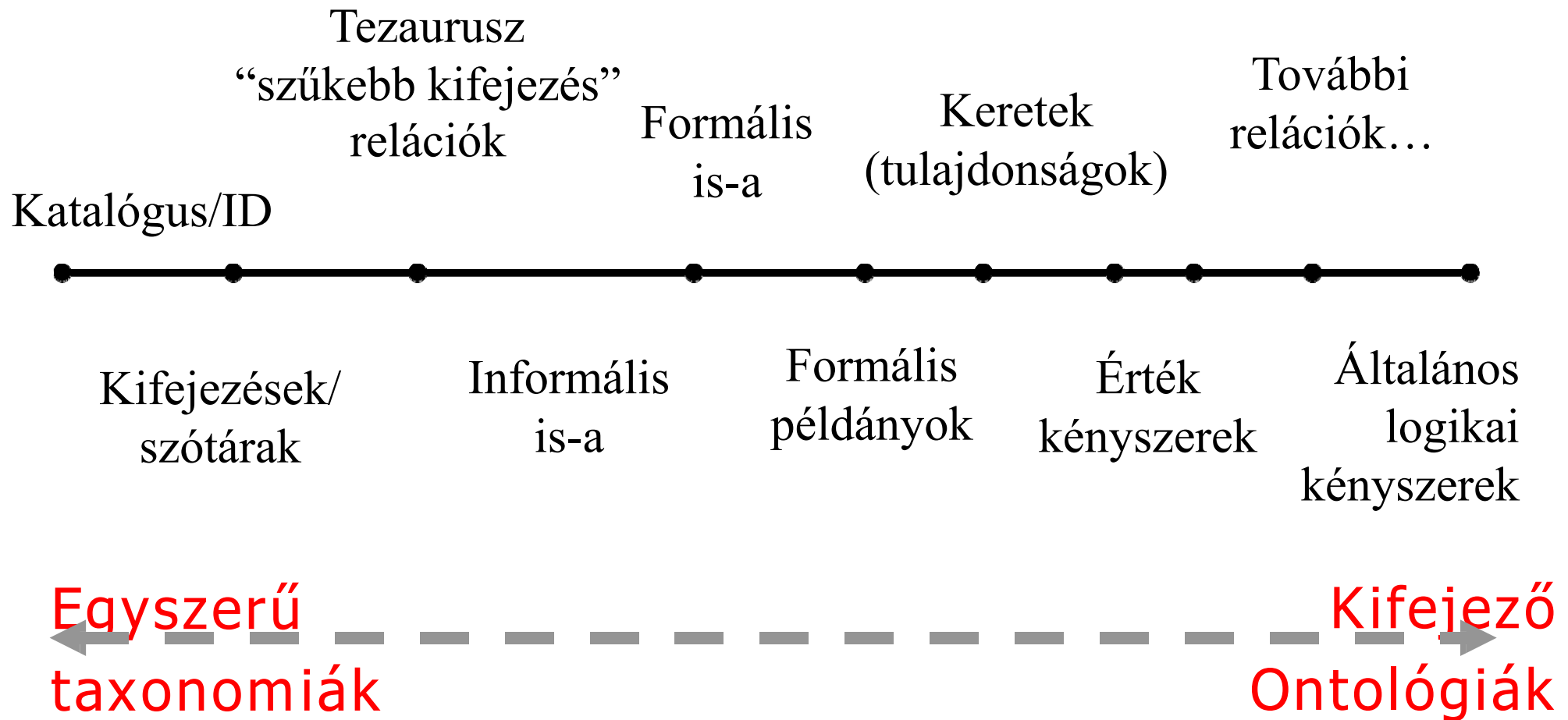


Ontológiák segíthetnek

- **Az ontológia fogalmakat definiál, amely egy adott terület leírásához szükséges.**
- **Valójában tudást tárolnak az adott területre vonatkozó, vagy a területeken felüli információkra.**

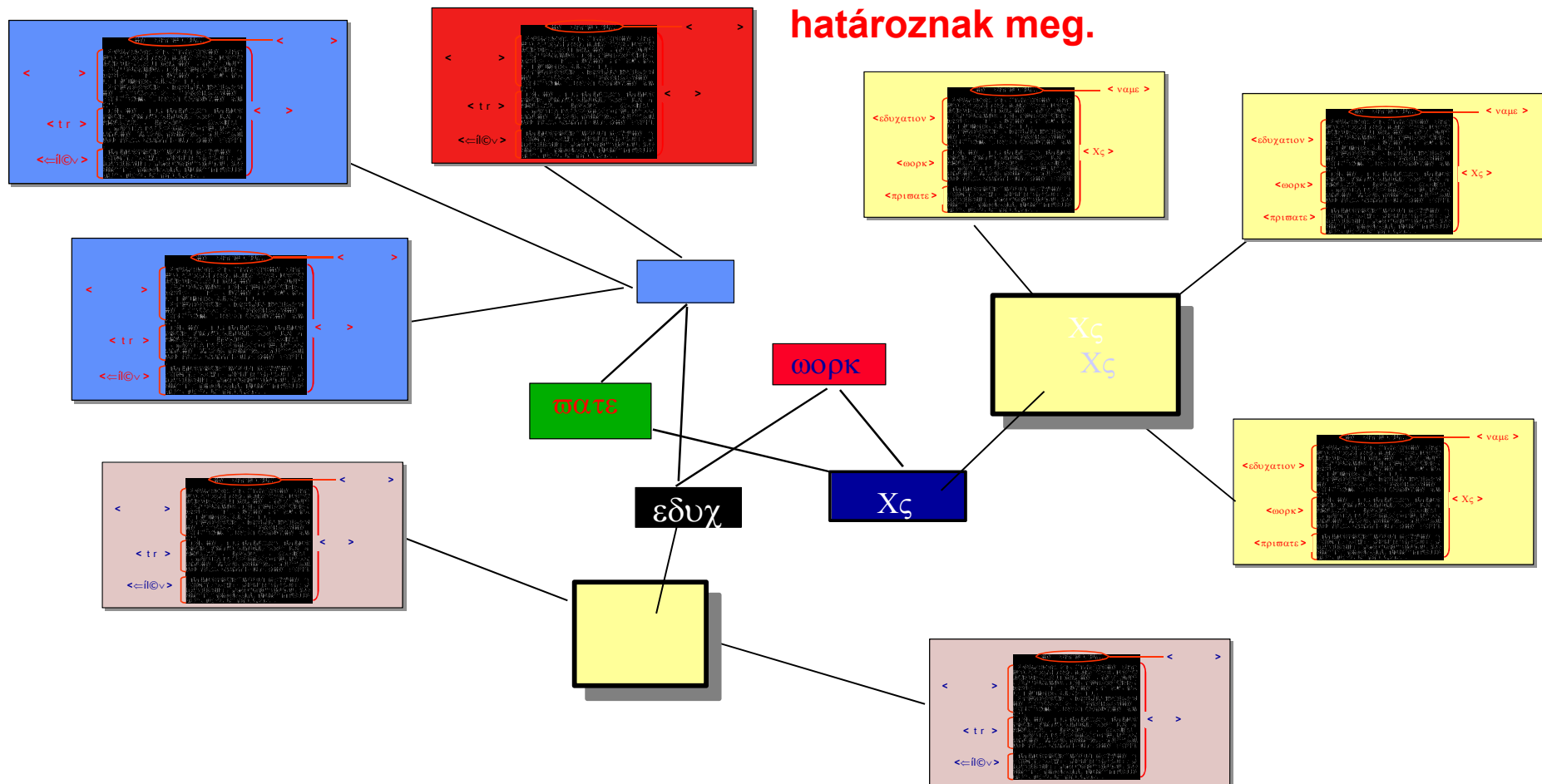
Working Draft, Web Ontology Working Group.

Ontológiák segíthetnek ...



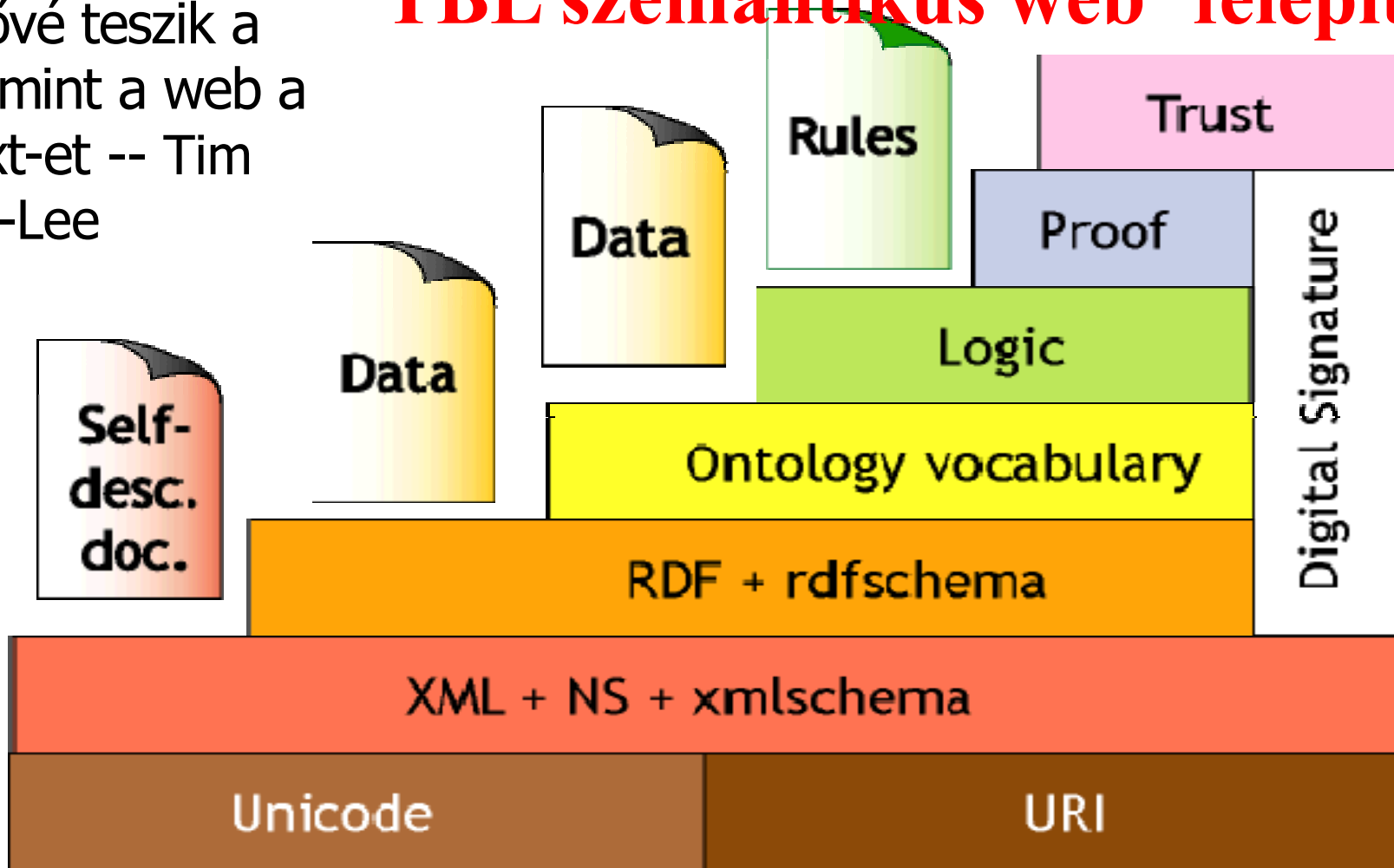
Külső hivatkozások beillesztése

SW nyelvek leképezéseket és struktúrát határoznak meg.



"A szemantikus web elérhetővé teszi a tudást, mint a web a hipertext-et -- Tim Berners-Lee

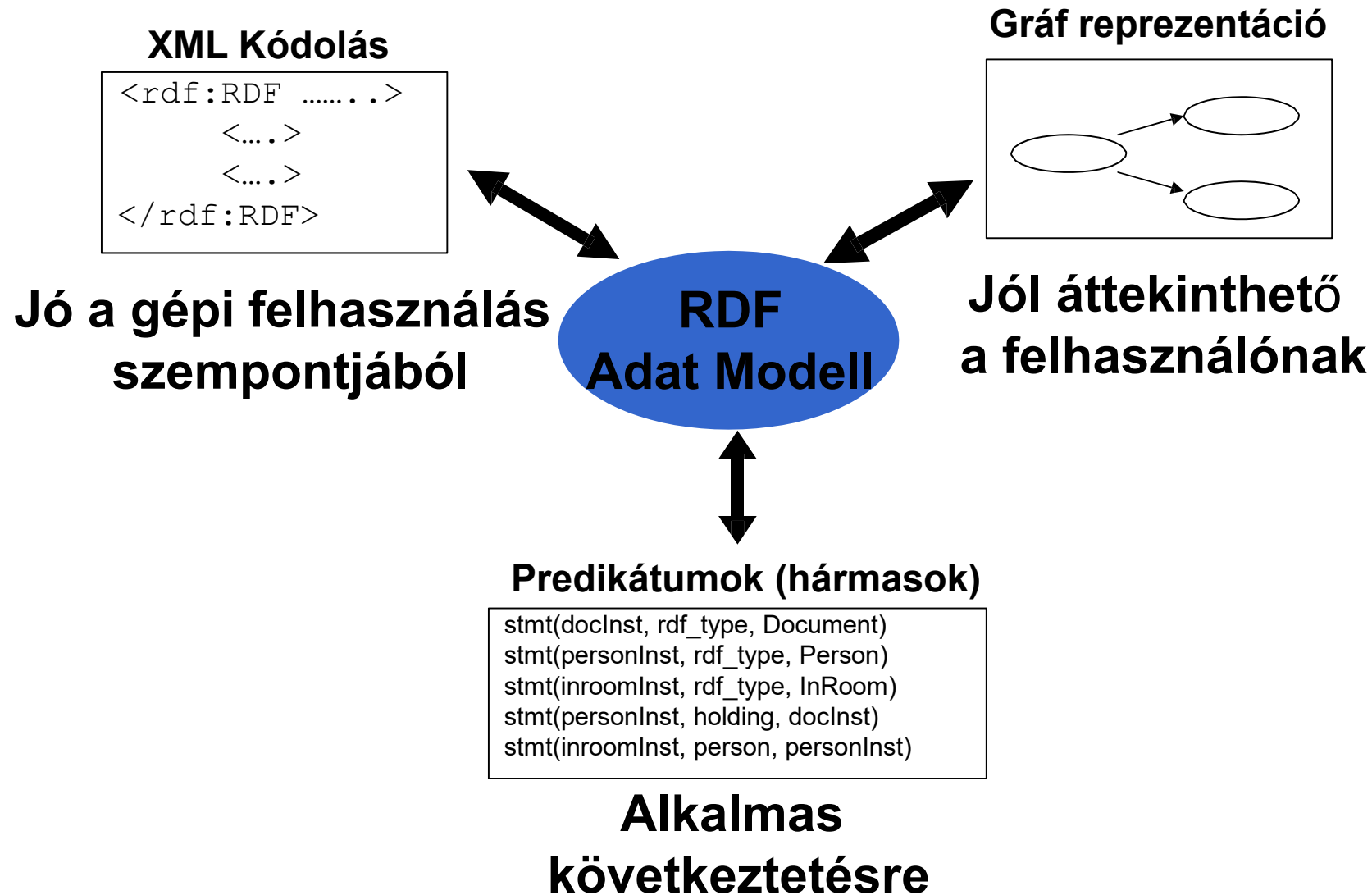
TBL szemantikus web felépítése



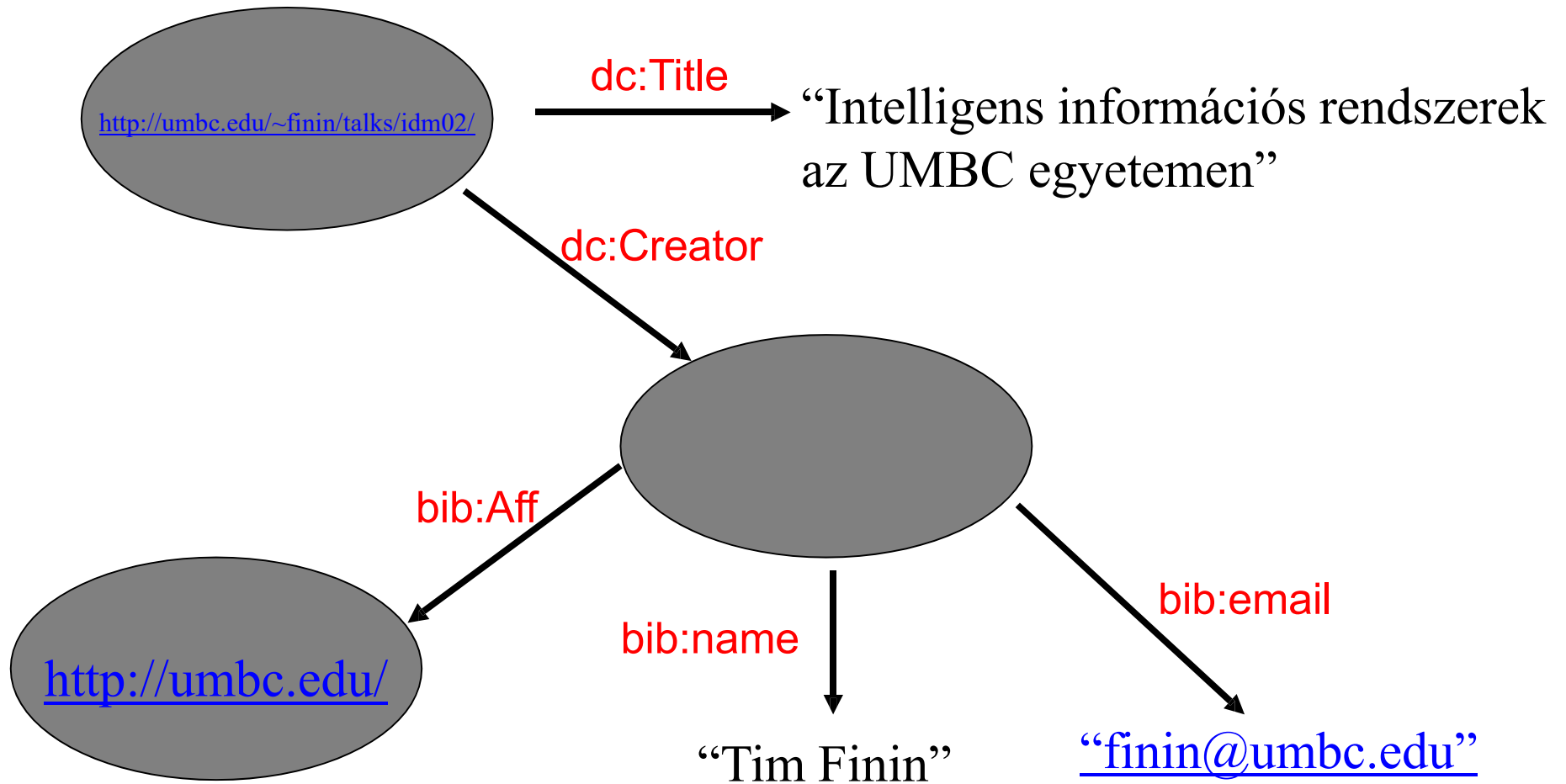
Szemantikus web leíró nyelvek ma

- Két fontos nyelv:
 - DAML – Darpa Agent Markup Language
<http://www.daml.org/>
 - RDF – Resource Description Framework
<http://www.w3.org/RDF/>
- Illetve új fejlesztés (W3C)
 - OWL – Ontology Web Language
<http://www.w3.org/2001/sw/>

RDF az első SzW nyelv



Egyszerű RDF példa



A példa XML szintaxissal

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:bib="http://daml.umbc.edu/ontologies/bib/">
<description about="http://umbc.edu/~finin/talks/idm02/">
  <dc:title>Intelligent Information Systems on the Web and in the
Aether</dc>Title>
  <dc:creator>
    <description>
      <bib:Name>Tim Finin</bib:Name>
      <bib:Email>finin@umbc.edu</bib:Email>
      <bib:Aff resource="http://umbc.edu/" />
    </description>
  </dc:Creator>
</description>
</rdf:RDF>
```

Hármasokat alkalmazó reprezentáció

- RDF kifejezések leírhatóak hármasokkal:
- *<alany> <állítmány> <tárgy>*

Megengedett szintaxis:

<URI><URI><URI>

<URI><URI><string>