

Connecting Web and User aneb

Sémantické vakuum a jeho důsledky (in Czech)

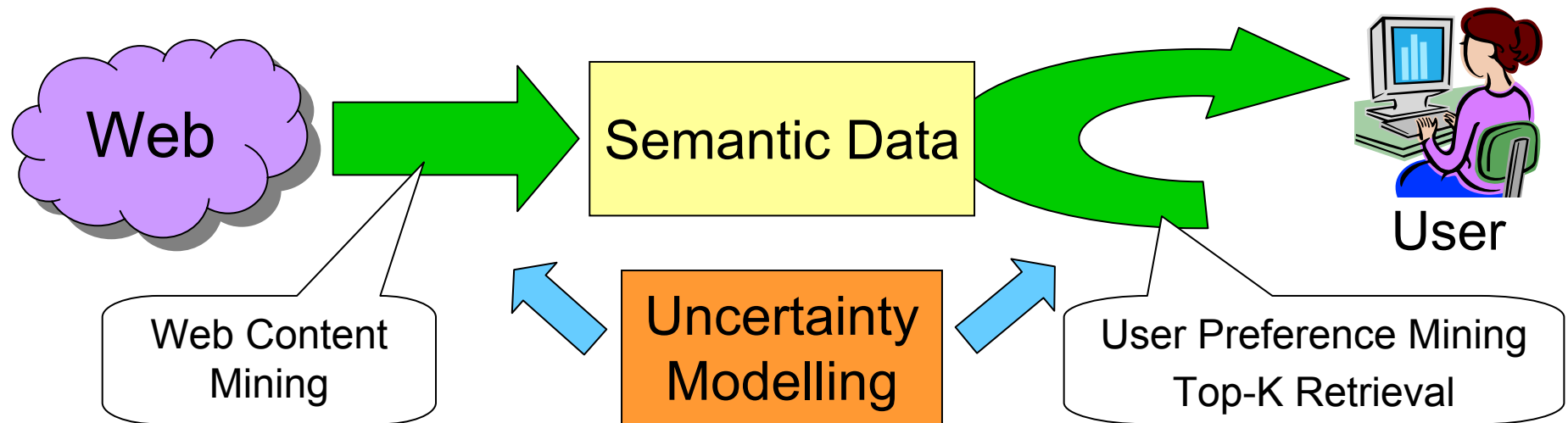
J. Dědek, A. Eckhardt, P. Vojtáš

Karlova Univerzita, AVČR, Praha

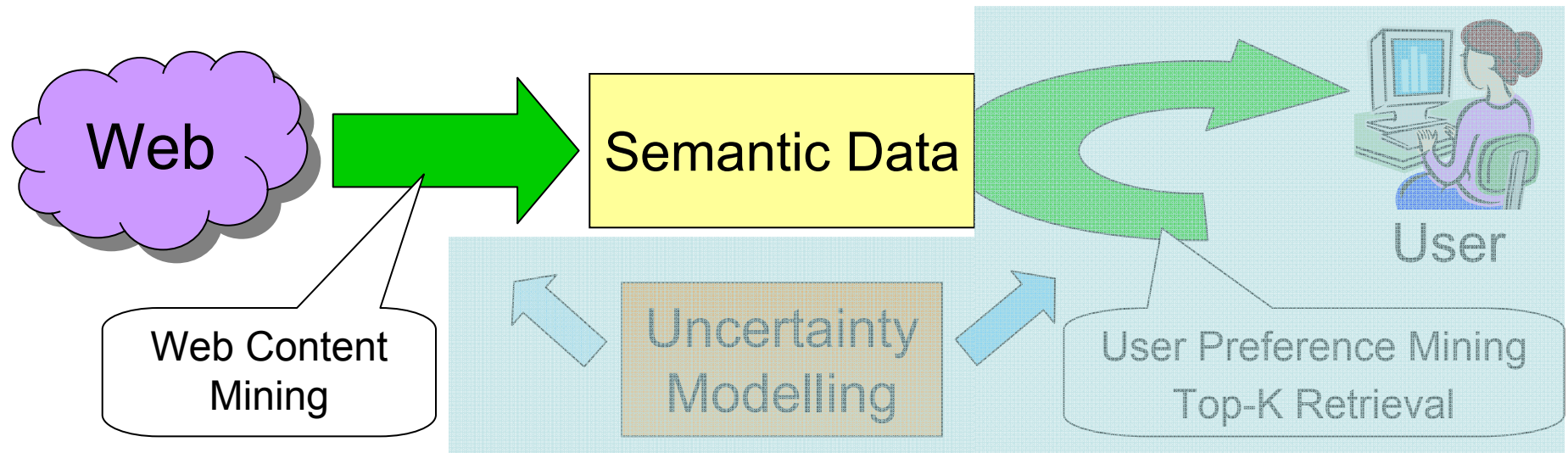
W2U – (Connecting) Web to User

- Introduction
- Web content mining
 - Dominantly Tabular Pages
 - Dominantly Textual Pages
- Modelling user preferences
- Modelling of Uncertainty
- Web Semantization
- Conclusions, future work, ...

W2U



WCM - Web content mining



- Dominantly Tabular Pages
- Dominantly Textual Pages
- Annotation of web resources

WCM - dominantly tabular pages

Nejprodávanejší zboží této kategorie:

- HP PAVILION dv6319ea/ Core Duo T2080/ N7400/ 15.4" WXGA BV/ 1GB/ SATA 120GB 5.4k/ DVD±RW +DL-RAM/ WIFI/ CAM/ VIS HP **25 597,00**
- HP COMPAQ nx7400/ C-M 430/ 15.4" WXGA BV/ 512MB/ 120GB 5.4k/ DVD±RW +DL-RAM/ WIFI/ Bez OS **14 568,00**
- FUJITSU-SIEMENS Amilo Pa1538/ Turion 64 X2 TL52/ 15.4" WXGA CV/ N7400/ 1GB/ SATA 160GB/ DVD±RW +DL/ WIFI/ VIS HP **22 248,00**
- ACER Aspire 3693WLM/ C-M 430/ 15.4" WXGA CB/ 1GB/ 120GB 5.4k/ DVD±RW +DL-RAM/ WIFI/ CAM/ XP Ho **19 042,00**
- ASUS F3TC-AP075/ Turion 64 M26/ 15.4" WXGA CS/ N7300/ 1GB/ SATA 120GB 5.4k/ DVD±RW +DL-RAM/ WIFI/ BT/ CAM/ Bez OS **21 279,00**

TOP FUJITSU-SIEMENS Amilo PRO V3205/ Core Duo T2350/ 12.1" WXGA CV/ 1GB/ SATA 120GB 5.4k/ DVD±RW +DL/ WIFI/ BT.0/ XP Pr



Miniaturní elegantní notebook v jehož útroběch nalezneme výkonný dvoujádrový procesor firmy Intel ...

30 118,-
s DPH 35 841,-

1 Koupit
kód:ND032e

TOP ASUS F3TC-AP008/ Turion64 X2 TL52/ 15.4" WXGA CS/ N7300/ 1GB/ SATA 120GB 5.4k/ DVD±RW +DL-RAM/ WIFI/ BT/ CAM/ Bez OS



Moderní výkonný notebook osazený dvoujádrovým procesorem AMD Turion64 X2 TL52 s frekvencí 1,6 GHz ...

23 444,-
s DPH 27 898,-

1 Koupit
kód:NA424g

TOP MSI MegaBook L745-023CZ/ Core 2 Duo T5600/ 17" WXGA/ N7600/ 1GB/ 120GB 5.4k/ DVD±RW+DL/ WIFI/ BT/ CAM/ Bez OS + ZDARMA Promo MSI StarReader - externí USB2.0 čtečka 52v1 pro paměťové karty CF/ MD/ MS/ MSPRO/ MSDUO/ MMC/ RS-MMC/ SD/ SIM



Spolehlivý výkonný notebook v elegantním provedení přináší perfektní zážitek při sledování videa ...

33 013,-
s DPH 39 286,-

1 Koupit
kód:NB026

TOP ACER Aspire 1652ZWLM/ P-M 735/ 15.4" WXGA CB/ AX1300/ 512MB/ 80GB/ DVD±RW +DL/ WIFI/ XP Ho



Výkonný notebook nabízející mobilní technologii profesionálním i domácím uživatelům, kteří požadují ...

TOP IBM/LENOVO THINKPAD T43/ P-M 760/ 14.1" SXGA+/ AX300/ 512MB/ 80GB 7.2k/ DVD±RW +DL/ WIFI/ BT/ FPR/ XP Pr

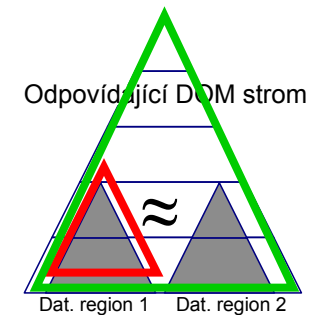
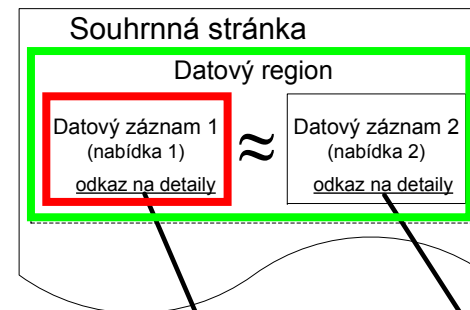
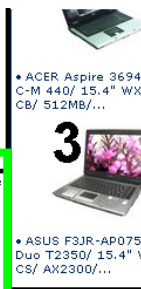


Velmi odolný a spolehlivý notebook určený pro náročné uživatele požadující vysoký výkon při nízké ...

TOP NEW ASUS F3JR-AP120C/ Core Duo T2080/ 15.4" WXGA CS/ AX2300/ 2GB/ SATA 120GB 5.4k/ DVD±RW +DL-RAM/ WIFI/ BT.0/ CAM/ VIS HP + ZDARMA KWORLD 323U - analogový TV / digitální DVB-T tuner, externí USB2.0 dongle, software, dálkové ovládání, anténa



Moderní spolehlivý notebook, který Vás jistě zaujme svou velmi příznivou cenou.



Amount of human work - ontology; training (patterns, detailed pages, ..); tolerance to changes

WCM - Tabular pages – our tool

Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://localhost:8080/Notebook/EXTRACT/http://www.alza.cz/18842920/notebooky.htm

VidomeXPI

Extractor Ontology Server

Mine instance from >>













Show instance Clear instance

About

VIDOMEXPI
User interface for ontology mining

Notebooky | alza.cz

http://localhost.../notebooky.htm

| | | |
|--|---|--|
| <p>TOP ASUS F3KE-AP012/ A64 X2 TK53/ 15.4" WXGA CS/ 1GB/ X2300 128MB/ 160GB 5.4k/ DVD±RW/ WF/ BT/ CAM/ Bez OS</p> <p>Inovativní zajímavý notebook nabízející dokonalou mobilitu, maximální výkon, uspokojivou výdrž na ...</p>  <p>-2,8% 17 599,- s DPH 20 367,-</p> <p>17 115,-</p> <p>Koupit</p> <p>kód:NA1261</p> | <p>TOP MSI MegaBook L730X-074CZ/ T64 X2 TL52/ 17" WXGA/ 1GB/ 120GB 5.4k/ DVD±RW+DL/ WF/ BT/ Bez OS</p> <p>Univerzální notebook v elegantním designu, který každého zaujme na první pohled. Nabízí špičkový ...</p>  <p>-14,4% 20 367,- s DPH 21 190,-</p> <p>17 807,-</p> <p>Koupit</p> <p>kód:NB024e</p> | <p>TOP ACER Aspire 3115.4" WXGA CE DVD±RW/ WiFi/ VLS HB</p> <p>Zajímá zvlášť Sempr Power</p>  <p>-14,9% 14 599,- s DPH 14 744,-</p> <p>12 390,-</p> <p>Koupit</p> |
| <p>TOP ACER Aspire 5613ZWLMI/ CD T2080/ 15.4" WXGA CB/ 1GB/ 80GB 5.4k/ DVD±RW/ WF/ VLS HP v zvláštní ceně</p> <p>Inovativní multimediální notebook, který je vybaven průkopnickou technologií poskytující poslušný ...</p>  <p>-9,9% 15 999,- s DPH 17 016,-</p> <p>14 299,-</p> <p>Koupit</p> <p>kód:NC2716</p> | <p>TOP ACER Aspire 5633ZWLMI/ CD T5500/ 15.4" WXGA CB/ 2GB/ 160GB 5.4k/ DVD±RW/ WF/ VLS HP</p> <p>Zajímavý moderní notebook v jehož útrobcích nalezneme výkonný procesor firmy Intel Core 2 Duo T5500 ...</p>  <p>-9,8% 20 367,- s DPH 21 669,-</p> <p>18 226,-</p> <p>Koupit</p> <p>kód:NC2726</p> | <p>TOP LENOVO 3000 V WXGA VV/ 1GB WF/ BT/ FPR/ CAM/ VLS BL</p> <p>Minta (stě) design</p>  <p>-17,1% 30 333,- s DPH 30 333,-</p> <p>25 490,-</p> <p>Koupit</p> |
| <p>TOP NEW ACER Aspire 5315-050508M/ CH 530/ 15.4" WXGA CB/ 512MB/ 80GB 5.4k/ DVD±RW/ WF/ Bez OS</p> <p>Inovativní notebook v neopakovatelném designu Gemstone spojuje výkon procesoru Intel Celeron s ...</p>  <p>-10,1% 10 499,- s DPH 11 226,-</p> <p>9 434,-</p> <p>Koupit</p> <p>kód:NC274g</p> | <p>TOP NEW ASUS X51H-AP020C/ CD T2450/ 15.4" WXGA CS/ 1GB/ 120GB 5.4k/ DVD±RW/ WF/ BT/ VLS HP</p> <p>Přesná přeměna počítače určený pro domácí nebo kancelářské využití s velmi příznivou cenou. Je ...</p>  <p>-19% 16 999,- s DPH 16 216,-</p> <p>13 627,-</p> <p>Koupit</p> <p>kód:NA172e</p> | <p>TOP ASUS FSRL-AP0 WXGA CS/ 1GB WF/ BT/ CAM/ Bez OS</p> <p>Bleag přízní nabíd</p>  <p>-18,9% 16 599,- s DPH 16 111,-</p> <p>13 539,-</p> <p>Koupit</p> |
| <p>TOP ASUS F2KA-AP012C/ T64 X2 TL60/ 15.4" WXGA CS/ 1GB/ X2600 512MB/ 160GB 5.4k/ DVD±RW/ WF/ BT/ CAM/ VLS HP</p> <p>Zajímavý spolehlivý notebook, se kterým se přinejmenším do světa vizuálních a zvukových záběrů. ...</p>  <p>-13,1% 26 999,- s DPH 27 739,-</p> <p>23 310,-</p> <p>Koupit</p> <p>kód:NA1265</p> | <p>TOP ACER Extensa 5220-050508M/ CH 520/ 15.4" WXGA/ 512MB/ 80GB 5.4k/ DVD±RW/ WF/ Bez OS</p> <p>Atraktivní moderní notebook, který oceníte především díky nízké ceně a dokonalé mobilitě. Základní ...</p>  <p>-9% 10 399,- s DPH 11 261,-</p> <p>9 463,-</p> <p>Koupit</p> <p>kód:NC052i</p> | <p>TOP ACER Extensa 5 T2310/ 15.4" V 5.4k/ DVD±RW/ WF/ Bez OS</p> <p>Modele štude a síře</p>  <p>-3,6% 13 499,- s DPH 15 462,-</p> <p>12 993,-</p> <p>Koupit</p> |

Recognized record 1

REGION_ID: 0
PRICE: 20 367,- or 17 115,- or 17 609,-
PRODUCER: ASUS
DATE: 2007-11-21 14:38:19.289
SIZE_OF_HDD: 160GB
TOKENIZED_TEXT: Inovativní zajímavý notebook nabízející dokonalou mobilitu,maximální výkonuspokojivou výdrž na ... [ASUS F3KE-AP012]A64 X2 TK53[15.4" WXGA CS]1GB[X2300 128MB]160GB 5.4k[DVDRW]WF[BT]CAM]Bez OS[Skladem 3-5 ks]s DPH 20 367,-[17 115,-[17 609,-[2,8%]NA426]kód:]
LINKS_TO_DETAIL: /notebook-asus-f3ke-ap012-d79797.htm
DISPLAY_SIZE: 15.4"

Recognized record 2

REGION_ID: 0
PRICE: 21 190,- or 17 807,- or 20 796,-
PRODUCER: MSI
DATE: 2007-11-21 14:38:19.289
SIZE_OF_HDD: 120GB
TOKENIZED_TEXT: Univerzální notebook v elegantním designukterý každého zauj na první pohled. Nabízí špičkový ... [MSI MegaBook L730X-074CZ][T64 X2 TL52]1 WXGA[1GB]120GB 5.4k[DVDRW+DL]WF[BT]Bez OS[Skladem 1-2 ks]s DPH 21 190,-[17 807,-[20 796,-[14,4%]NB024e]kód:]
LINKS_TO_DETAIL: /notebook-msi-megabook-1730x-074cz-d74966.htm
DISPLAY_SIZE: 17"

Recognized record 3

REGION_ID: 0
WiFi: WiFi
PRICE: 14 744,- or 12 390,- or 14 559,-
PRODUCER: ACER
DATE: 2007-11-21 14:38:19.289
SIZE_OF_HDD: 80GB
TOKENIZED_TEXT: Zajímavý inovativní notebook založený na procesoru Mobile AMD Sempron s technologií AMD PowerNow! ... [ACER Aspire 3105WLMi]S64 3600+[15.4" WXGA CB]1GB[80GB 5.4k[DVDRW]WiFi[VLS HB]Skladem 1-2 ks]s DP 14 744,-[12 390,-[14 559,-[14,9%]NC205n]kód:]
LINKS_TO_DETAIL: /notebook-acer-aspire-3105wlm-d80383.htm
DISPLAY_SIZE: 15.4"

Recognized record 4

REGION_ID: 0
PRICE: 17 016,- or 14 299,- or 15 869,-
PRODUCER: ACER
DATE: 2007-11-21 14:38:19.289

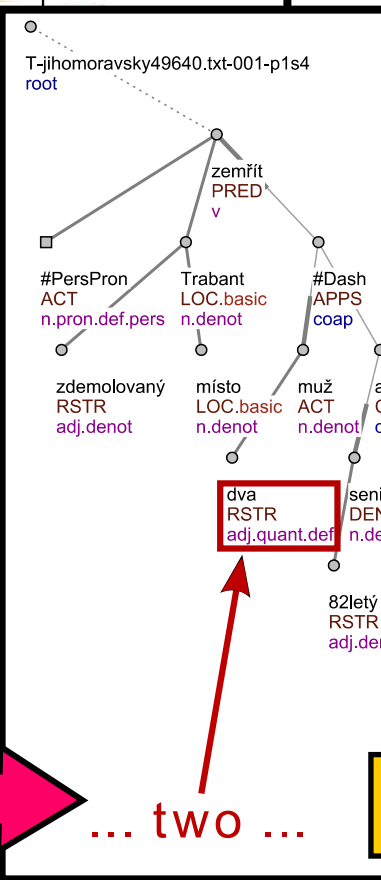
WCM - dominantly textual pages



Nehoda byla operačním střediskem HZS ohlášena ve 13.13 hodin na místě zasahovala jednotka profesionálních hasičů ze stanice Tříšnově. Jednalo se o čelní srážku autobusu Karosa s vozidlem Trabant 601. Podle dostupných informací trabant jedoucí ve z Brna do Kuřimi zřejmě vyjel do protisměru, kde narazil do linkového autobusu dopravní společnosti ze Žďáru nad Sázavou. Ve zdemolovaném trabantu na místě zemřeli dva muži – 82letý senior a další muž, jehož totožnost zjišťují policisté

Hasiči udělali na vozidle protipožární opatření a po vyšetření zadokumentování nehody dopravní policii vrak trabantu zaklesnutý pod autobusem pomoci lana odtrhli. Po odstranění střechy trabantu pak kabiny vyprostili těla obou mužů. Obě vozidla – trabant i autobus, postupně odstranili na kraj vozovky a uvolnili tak jeden jízdní pruh. Úřadníci vrak trabantu uklidili vozovku a ukončili zásah.

Source web page



```
tree_root(node0_0). node(node0_0).
id(node0_0, t_jihomoravsky49640_txt_001_p1s4).
%%%%%%%%% node0_1 %%%%%%%%%%
node(node0_1).
functor(node0_1, pred).
gram_sempos(node0_1, v).
t_lemma(node0_1, zemrit).
%%%%%%%%% node0_2 %%%%%%%%%%
node(node0_2).
functor(node0_2, act).
gram_sempos(node0_2, n_pron_def_pers).
t_lemma(node0_2, x_perspron).
%%%%%%%%% node0_3 %%%%%%%%%%
node(node0_3). id(node0_3,
functor(node0_3, loc).
gram_sempos(node0_3, n_denot).
t_lemma(node0_3, trabant).
...
edge(node0_3, node0_7).
edge(node0_3, node0_8).
...

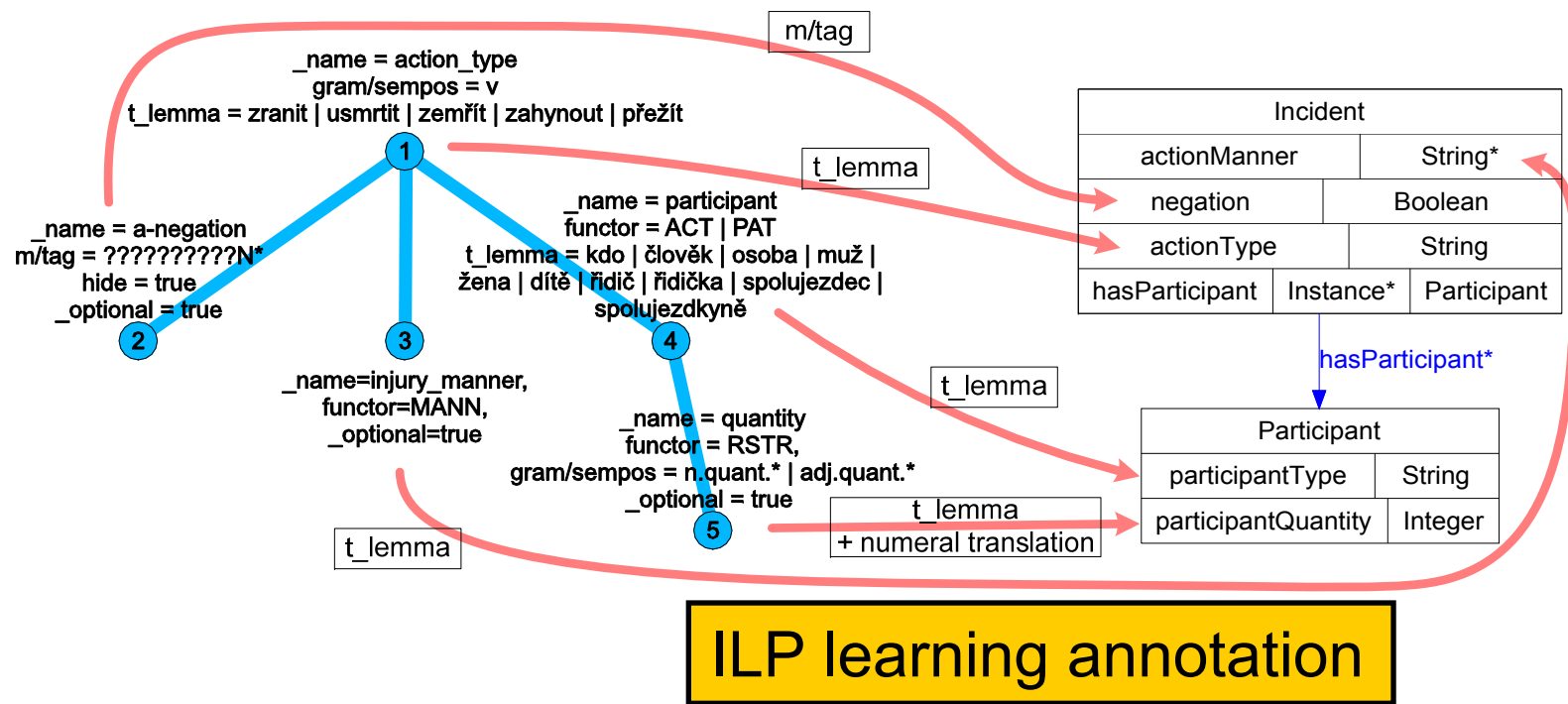
```

Logic representation

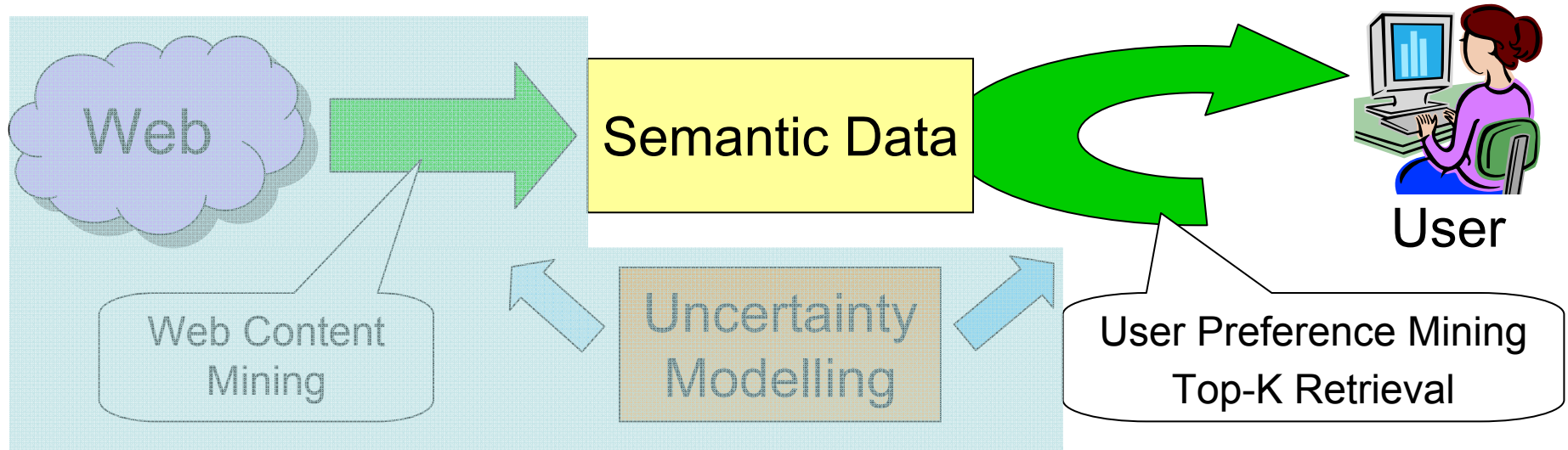
Linguistic trees - UFALware

... two ...

WCM – web ressource annotation



Modelling user preferences



- unexperienced user
- inconsistencies
- preferences developing in time
- connections to economy, OR, HCI

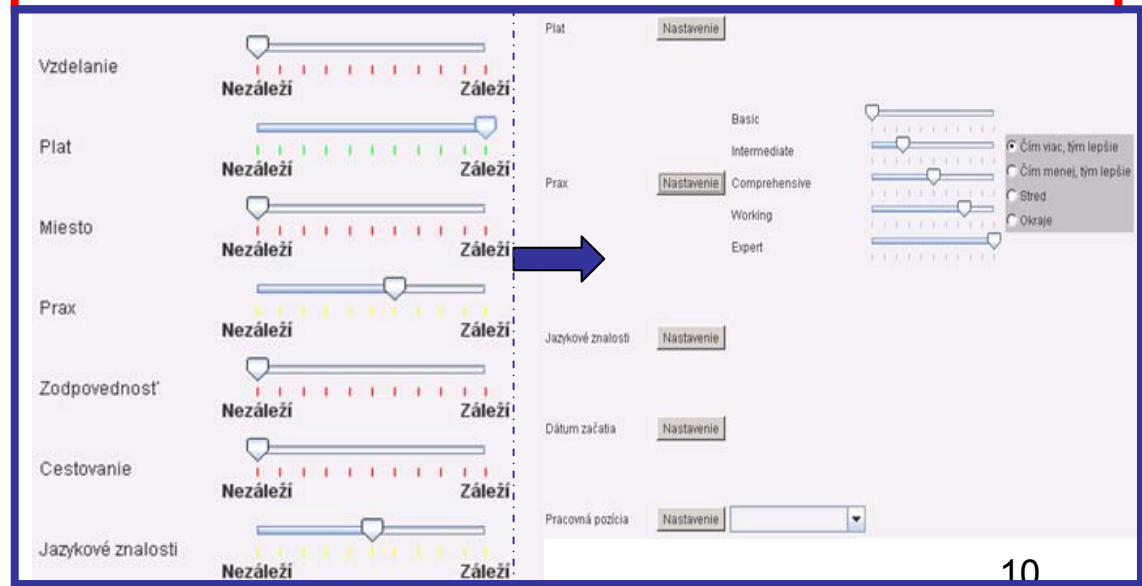
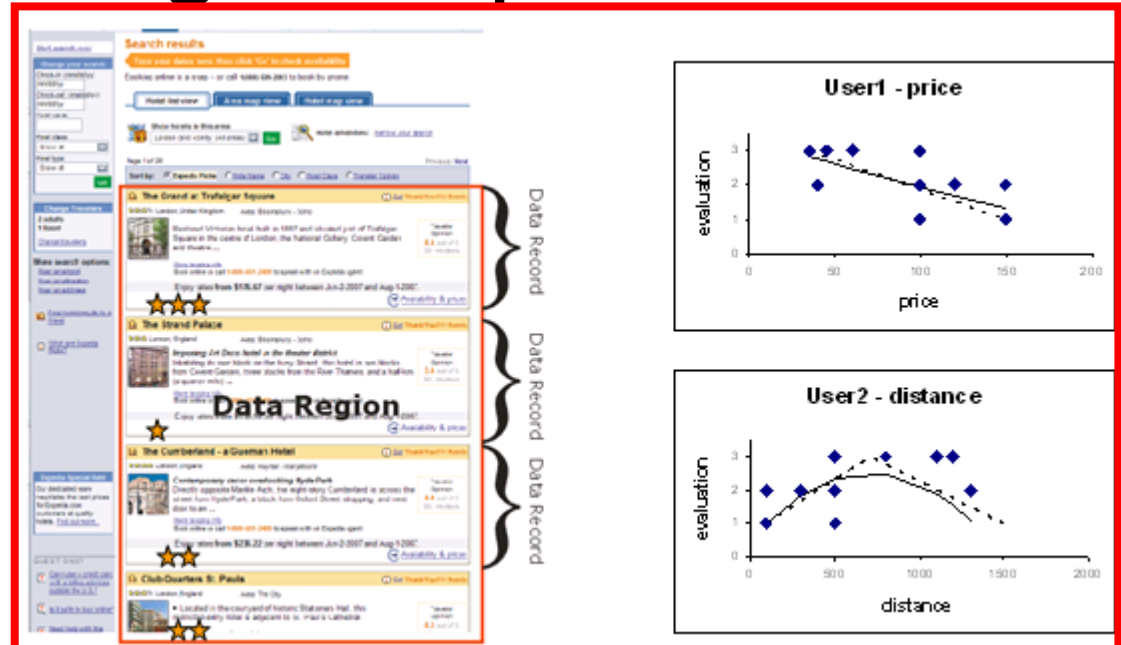
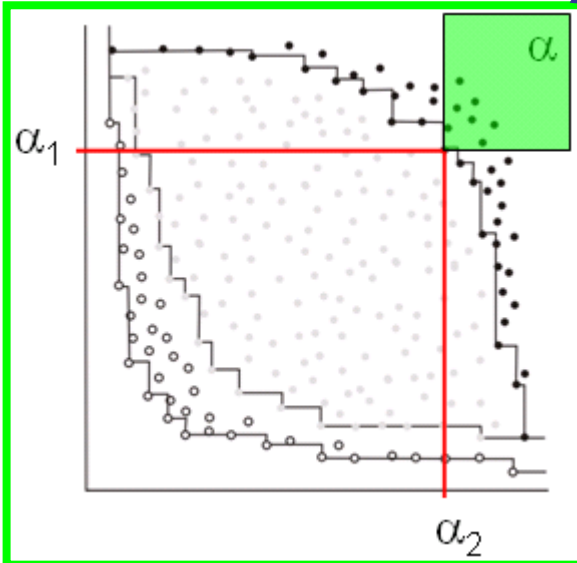
Modelling-learning user preferences

User preferences

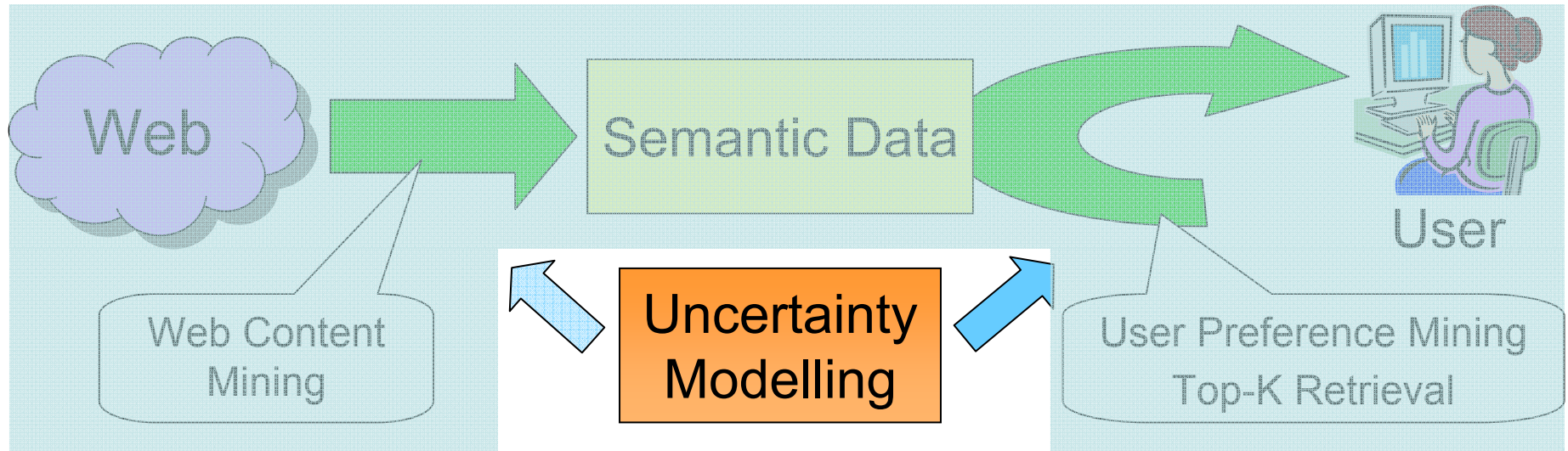
- **explicit** →
- implicit
- **direct** ↘

Fuzzy preference

ILP



Uncertainty modelling



Various sources of uncertainty

Various models of uncertainty

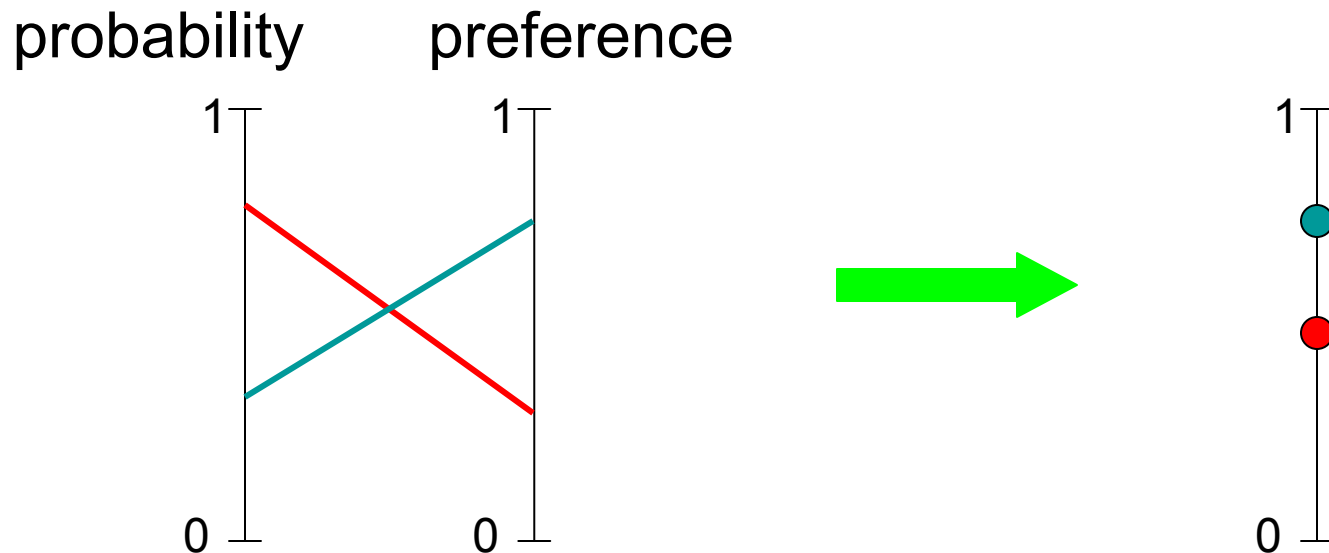
probabilistic

preferential

No formal models for combination - heuristics

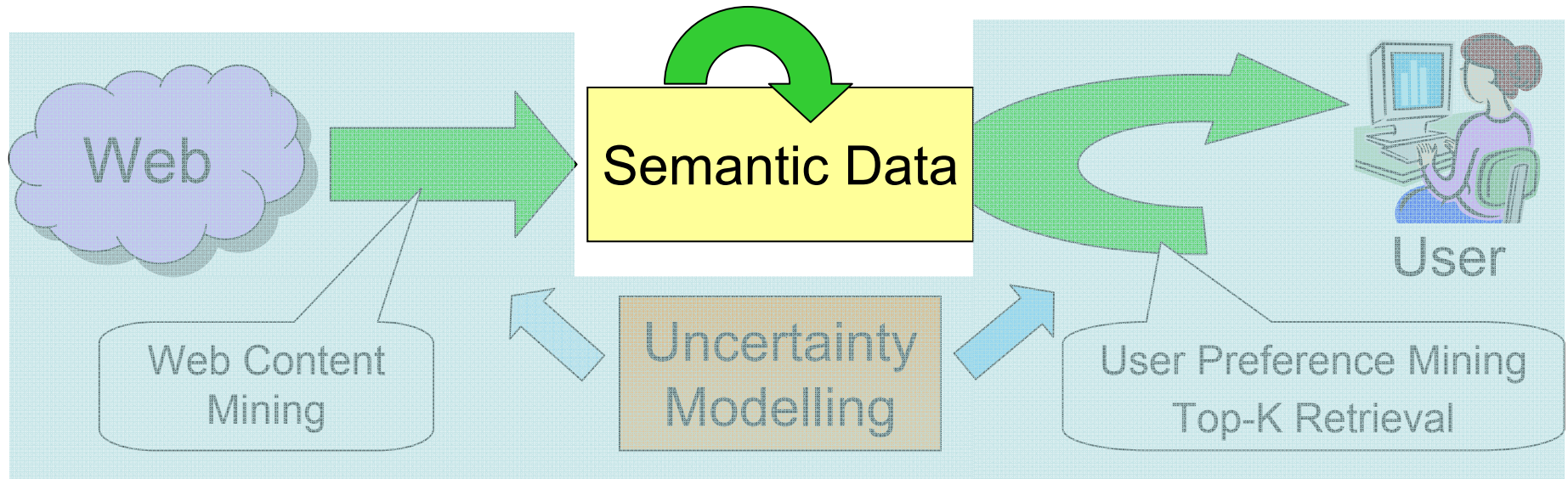
Probabilities versus preference

- Probability of our WIE algorithm correctly recognized attribute values and-or understood sentences
- User preferences



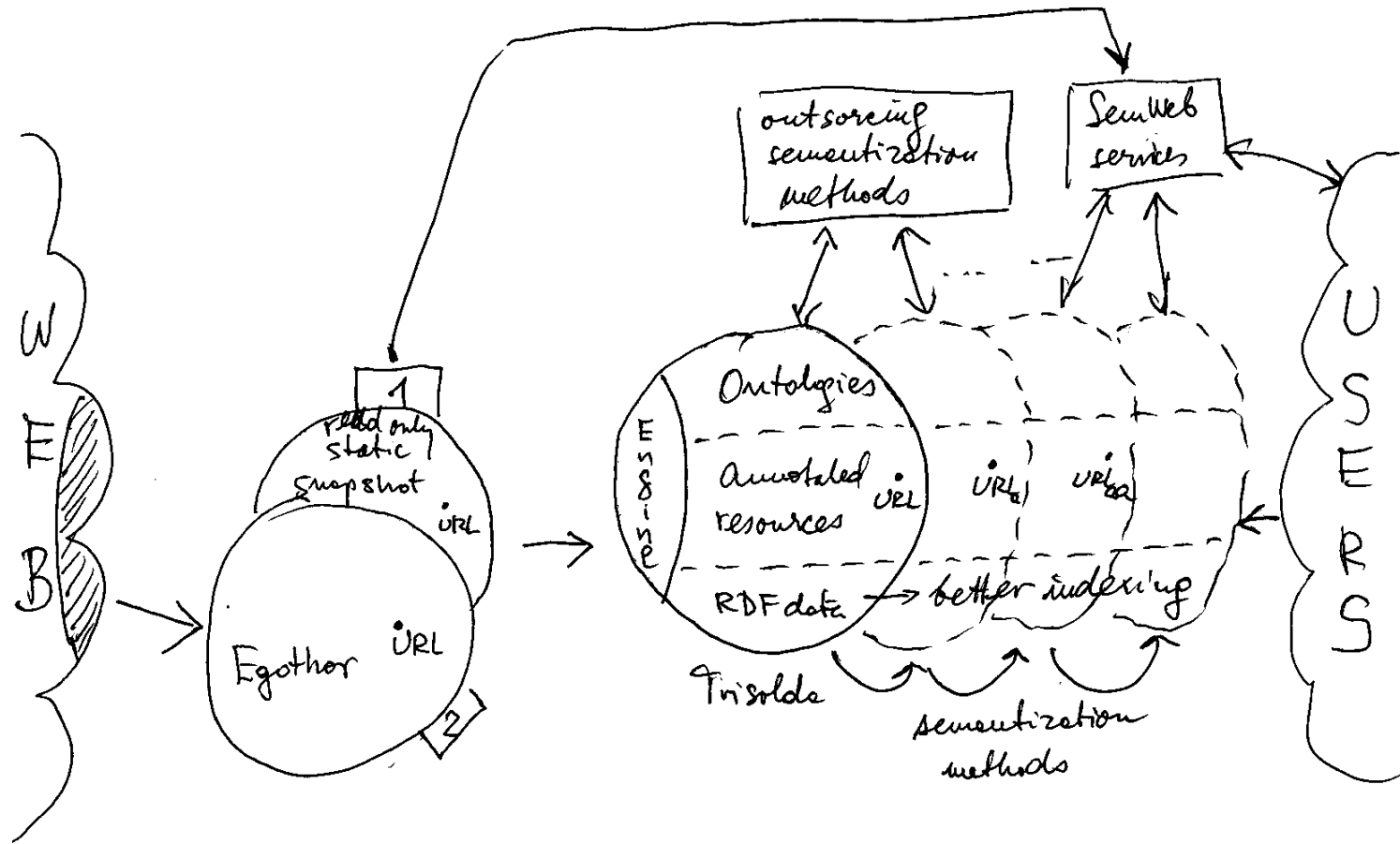
- Agregation – learning from user scoring/ranking

Web semantization



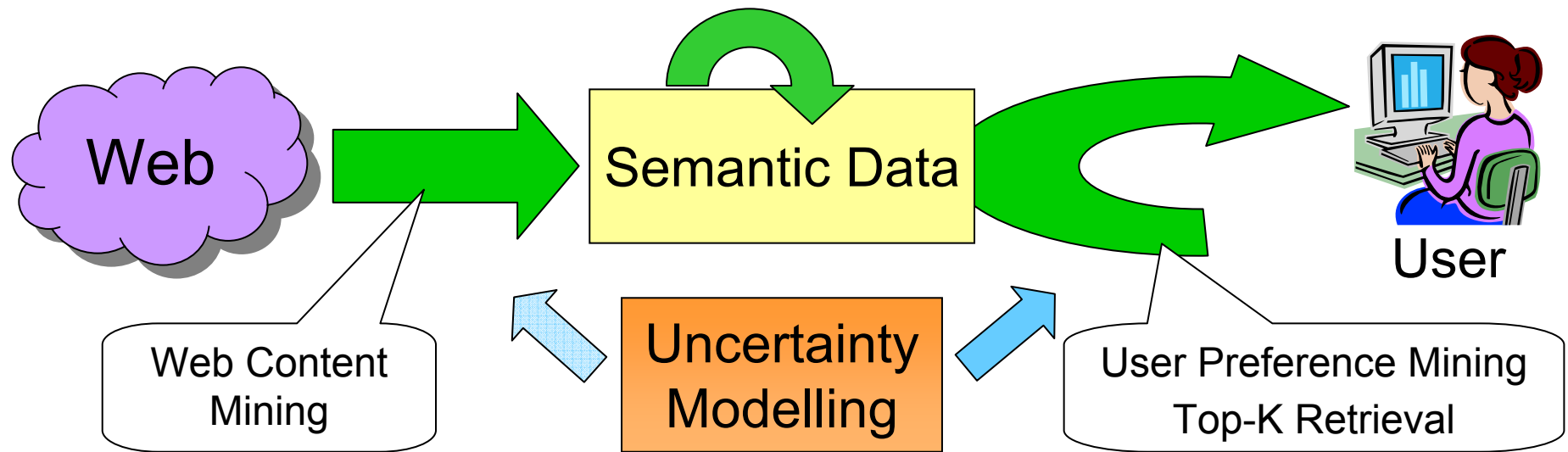
- „Zesemantičňování“ webu (in Czech)
- durable web resources
- search better than key word search
- first experiments promising

Web semantization – a process



obr. 2

Conclusions, future work, ...



- W2U – models, methods, prototypes, experiments
- web services
- experiments with real users
- scaling problem