

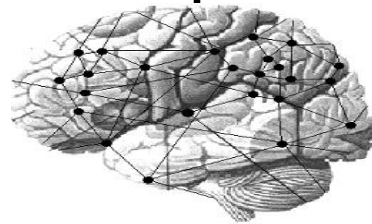
MINISTRSTVO ZA NOTRANJE ZADEVE

ODKRIVANJE NOVIH ZNANJ IZ POIZVEDB NA SPLETU

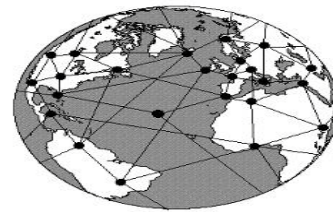
AVTORJI: Dr. Karl Petrič, prof. dr. Teodor
Petrič, prof. dr. Marjan Krisper in prof. dr.
Vladislav Rajkovič



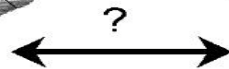
Zapletena omrežja



Človeški možgani



Globalno omrežje



DSI Portorož, 2011



1 Uvodna spodbuda

- Poizvedbe na spletu spadajo v okvir kolektivno zavednega in nezavednega, ki ga lahko merimo z analitikom spletnega prometa.
- V bistvu na ta način merimo kolektivne informacijske potrebe in / ali želje povsem globalno ali pa znotraj določene organizirane združbe.
- Skratka, večja ali manjša množica lahko določenemu raziskovalcu nevede in nehote nekaj pomembnega sporoča.
- Kdor zna brati kolektivna hotenja je zmožen ustvarjati kolektivno in poslovno inteligenco.

2 Raziskovanje dejavnosti uporabnikov

- **Raziskovanje dejavnosti uporabnikov** na osnovi metod in vizualizacijskih tehnik s področja odkrivanja zakonitosti v podatkih / besedilih na spletu lahko pomeni poslovno inteligenco, ki je v bistvu proces spreminjanja podatkov v informacije in le-te kasneje v znanje.
- **Cilji predstavitve**
 - a. analiza poizvedb s pomočjo statističnih metod in vizualizacijskih tehnik,
 - b. odkrivanje novih znanj iz poizvedb na podlagi neeksaktnih oziroma intuitivnih metod.

3 Priprava podatkov

- Iz Google analytics smo v Excel izvozili 1421 poizvedb vključno s podatki o številu poizvedb (podatki so bili pridobljeni v času od 27.11.2008 do 01.10.2010).
- Poizvedbe smo najprej razvrstili s pomočjo lastno izdelane klasifikacije od 1 do 7, nakar smo v nadaljevanju omenjene poizvedbe razvrstili še po univerzalni decimalni klasifikaciji (v nadaljevanju: UDK). Pri sleherni poizvedbi smo določili tudi predpostavljeni namen uporabnika.

3.1 Del poizvedb znotraj preglednice

LEGENDA

Preglednica prikazuje pripravljene podatke za poizvedbe spletnega obiskovalca (v nadaljevanju: PSO), obiski na stran SKMNZ, (v nadaljevanju: OB), klasificirane poizvedbe po lastnem scenariju (v nadaljevanju: KE), klasificirane poizvedbe po UDK in nenazadnje predpostavljeni namen določene poizvedbe (v nadaljevanju: ND). OSX je določena oseba (OSX1, OSX2, OSX3 itd).

PSO	OB	KE	UDK	ND
mnz	1431	1	35	Splošne informacije
ministrstvo za notranje zadeve	1015	1	35	Splošne informacije
specialna knjižnica mnz	272	1	35	Splošne informacije
revija varnost	242	3	0	Viri
e publikacije	136	3	0	Faktografsko znanje
kult žrtve	136	3	0	Faktografsko znanje
OSX1, mnz	98	2	9	Splošne informacije
knjižnica mnz	73	1	35	Splošne informacije
preprečevanje gospodarske kriminalitete	67	5	343.85	Faktografsko znanje
mnz knjižnica	64	1	35	Splošne informacije
ministerstvo za notranje zadeve	47	1	35	Splošne informacije
specialna knjižnica	40	1	35	Splošne informacije
e pregled novosti	34	3	0	Viri
OSX2	31	2	9	Splošne informacije
novosti knjige	30	3	0	Viri
nacionalni preiskovalni urad	29	1	351.74	Splošne informacije

3.2 Programska orodja, algoritmi in procesiranje podatkov

- ORA Casos → prog. orodje za izdelavo in analizo omrežij
 - Algoritem korelacijske podobnosti
 - Algoritem ključne informacije
 - Algoritem Eigenvektorjeve centralnosti
1. Določili smo štiri kazalce, ki so: PSO, OB, KE in ND.
 2. Ustrezno pripravljeno .TXT datoteko s 1421 poizvedbami in omenjenimi kazalci smo uvozili ORA Casos.
 3. Podatke smo procesirali tako, da smo najprej opredelili omrežje s PSO = vir (angl.: resource), OB = akcija (angl. action), KE = znanje (angl: knowledge) in ND = mnenje (angl.: belief). V nadaljevanju smo opredelili vir, cilj in lastnosti omrežja.
 4. V naslednji stopnji smo sprožili ukaz "Vizualiziraj (angl.: visualize)" in s tem vključili modul za vizualizacijo vzpostavljenega omrežja.
 5. Potrebno je bilo še določiti moč in usmerjenost povezave, barvo ter velikost vozlišč. Z algoritmom korelacijske podobnosti (angl.: correlation similarity) smo določili sorodnost vozlišč, medtem ko smo z algoritmom ključne informacije (angl.: centrality information) in eigenvektorjeve centralnosti (angl.: eigenvector centrality) določili velikost vozlišč.

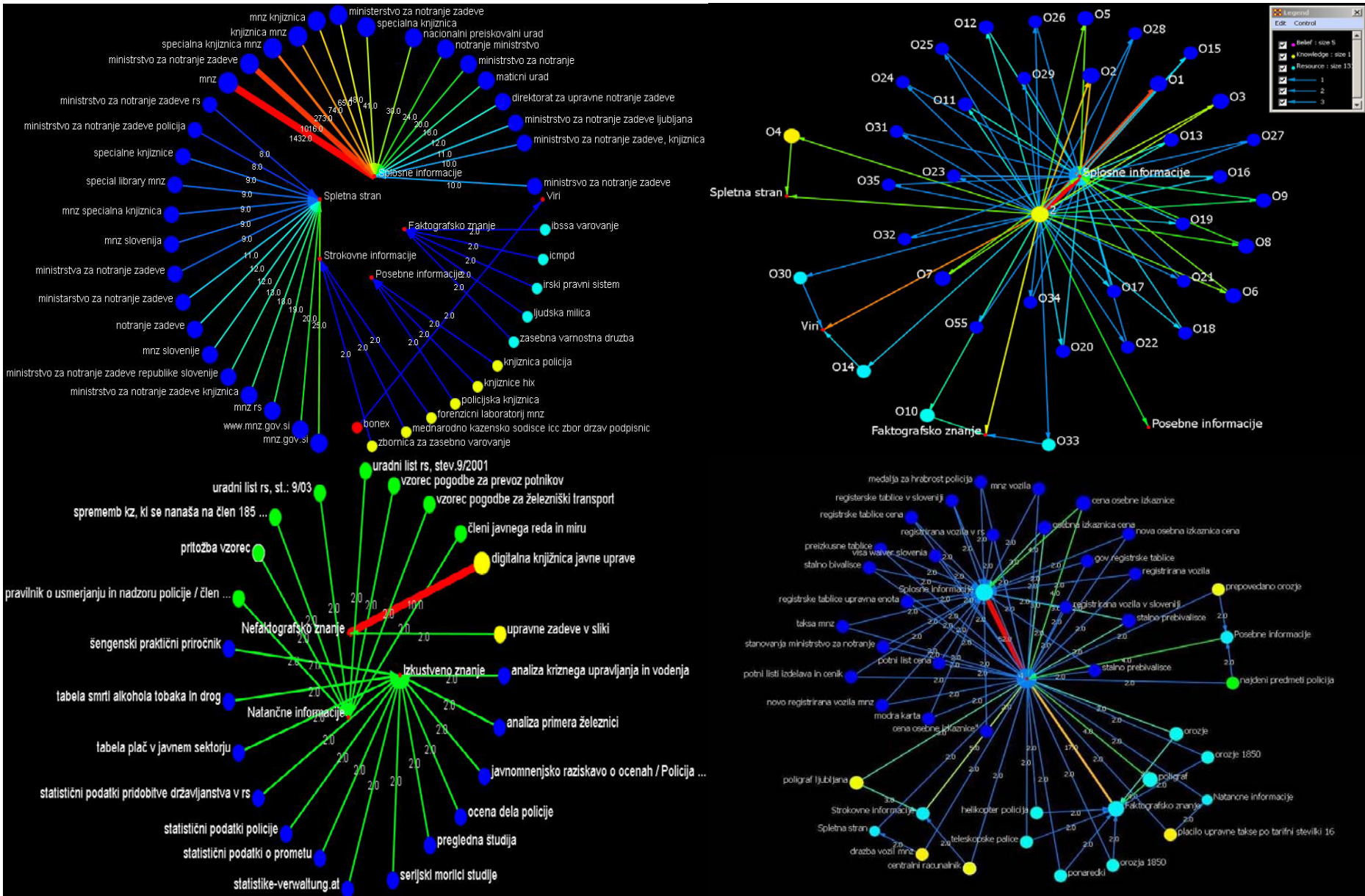
3.3 Lastna klasifikacija poizvedb

- - KE 1: v to skupino se razvrščajo poizvedbe, ki so označile sociološke sisteme in njihove dele (npr. MNZ, Specialna knjižnica MNZ, matični urad),
- - KE 2: v to skupino se razvrščajo poizvedbe o osebah (npr. OSX1, OSX2, ženske, moški, tujci, sekretar),
- - KE 3: se razvrščajo poizvedbe o intelektualnih kulturnih stvaritvah (npr. knjige, revije, dokumenti, tehnični sistemi, obrazci, sezname, pregledi, inovacije),
- - KE 4: se razvrščajo poizvedbe o predmetih in / ali materialih (npr. stanovanja, cena osebne izkaznice, modra karta, vozila, registrske tablice),
- - KE 5: se razvrščajo poizvedbe o področjih kulturnega pomena v najširšem smislu (npr. znanost, šport, umetnost, stroka, kriminalistika, poslovna inteligenca),
- - KE 6: se razvrščajo poizvedbe o dejavnostih, stanjih in dogodkih (npr. komuniciranje, preprečevanje nasilja, proslave),
- KE 7: in nazadnje se v to skupino razvrščajo poizvedbe v obliki vprašanj (npr. kaj potrebuje tujec za bivanje?, kako postati policist?)

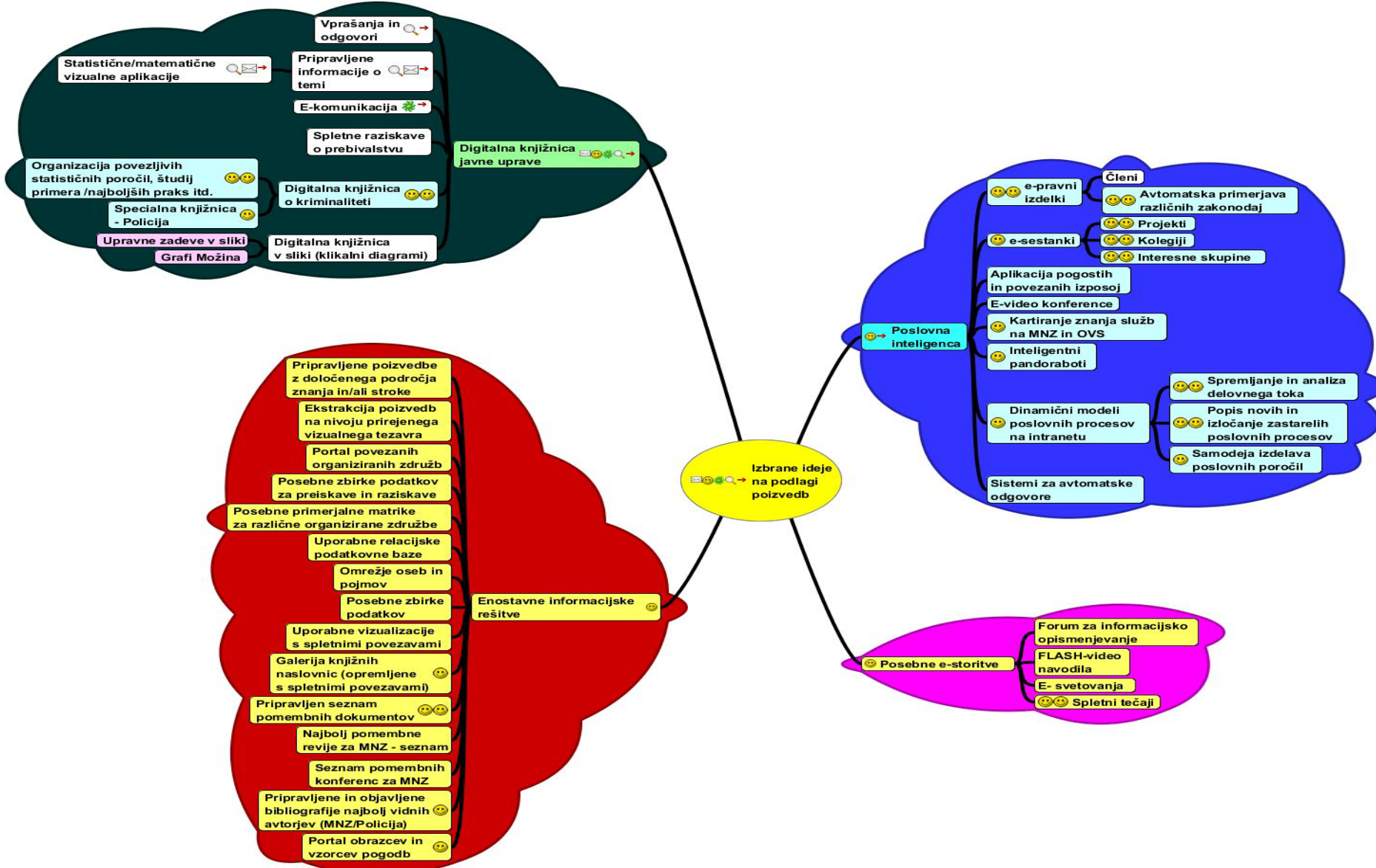
3.4 Hipotetični namen uporabnika

- ND pomeni z vidika spletnega obiskovalca motiv za poizvedovanje in posredno tudi za obisk določene spletne strani. Nameni uporabnikov:
 - želja po splošnih informacijah (npr. splošni podatki o določenih osebah, oddelkih, podjetjih, storitvah),
- - želja po dostopu do določene spletne strani,
 - želja po virih (npr. bibliografija, zbirke podatkov, digitalne knjižnice),
 - želja po faktografskem znanju (npr. kriminaliteta, sodne vede),
 - želja po natančnih informacijah (npr. uporabnik želi na vpogled točno določen člen iz zakonodaje, izvedba natančno določenega postopka),
 - želja po strokovnih informacijah (npr. merila in standardi za organizacijo, poročila, pritožbe, postopki),
 - želja po posebnih informacijah (npr. javnomnenjske raziskave, tiralice),
 - želja po izkustvenih znanjih (npr. razne analize, statistike, meritve) in
 - želja po nefaktografskem znanju (gre za znanja, ki se jih ljudje niso v šolah ali na fakultetah naučili npr. nove metode, zvijače, intuitivna znanja).

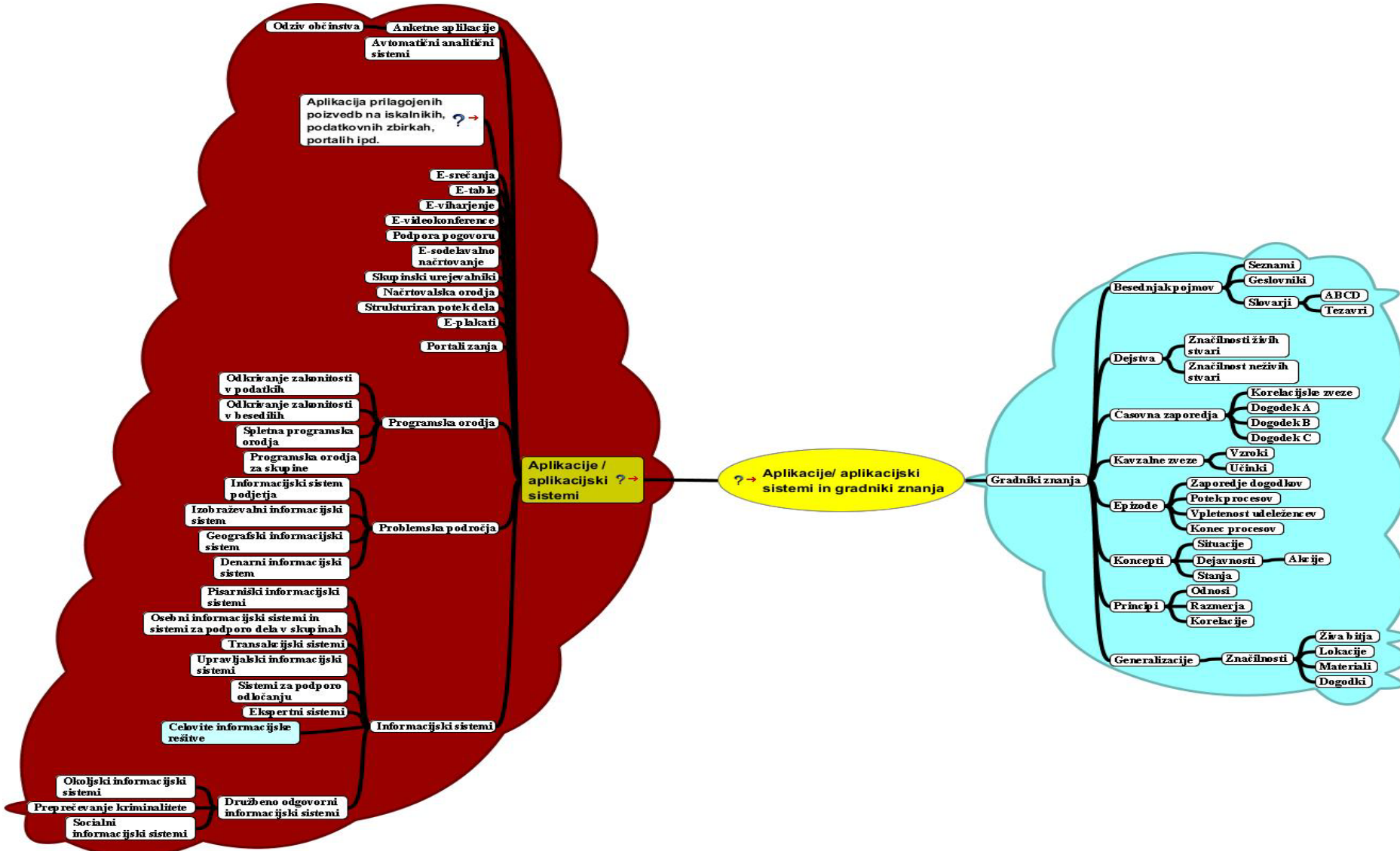
4 Odkrivanje zakonitosti v poizvedbah od KE 1 do KE 7 in po UDK



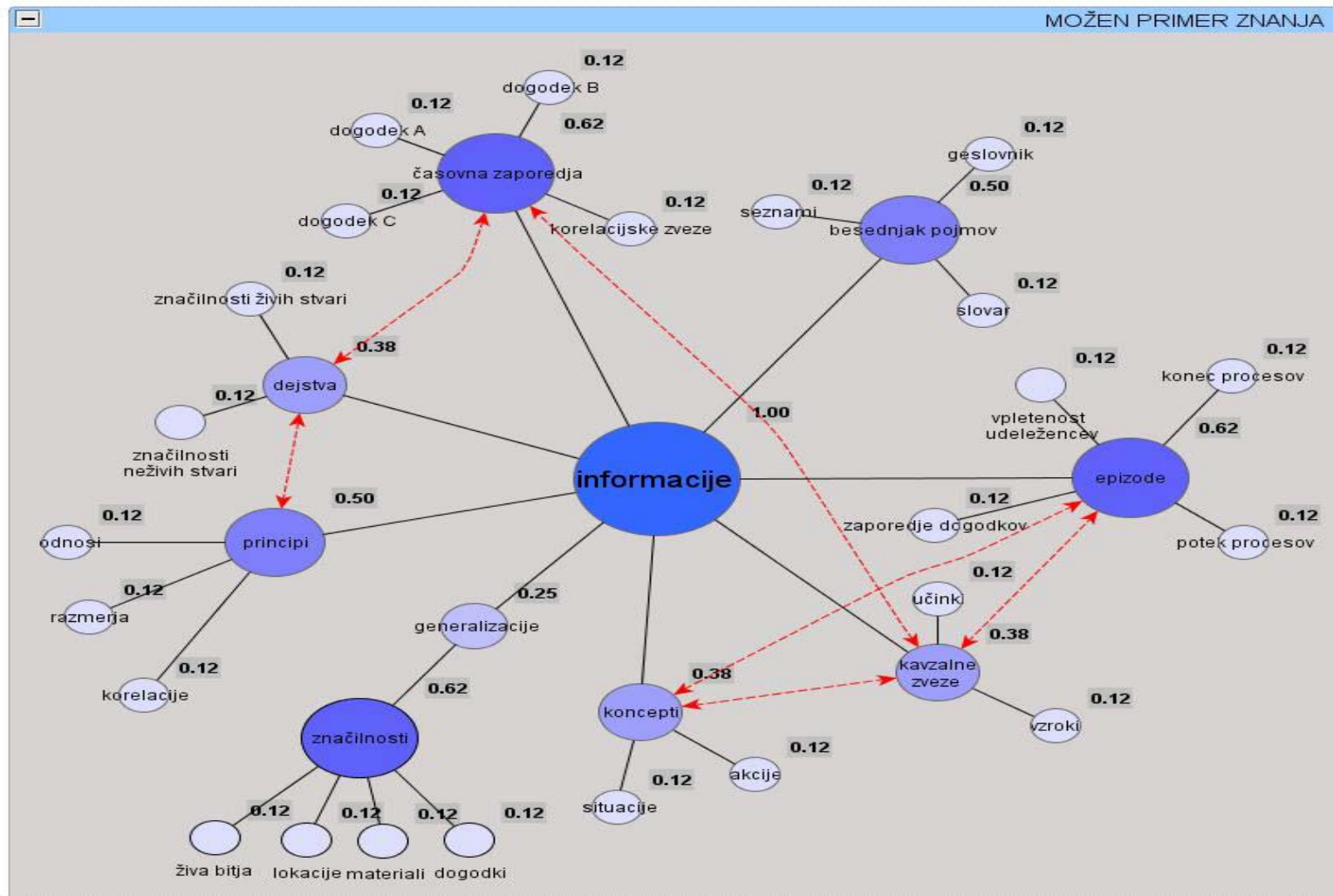
5 Odkrivanje novih idej in znanj iz poizvedb



5.1 Miselni vzorec aplikacij/aplikacijskih sistemov in gradnikov znanja



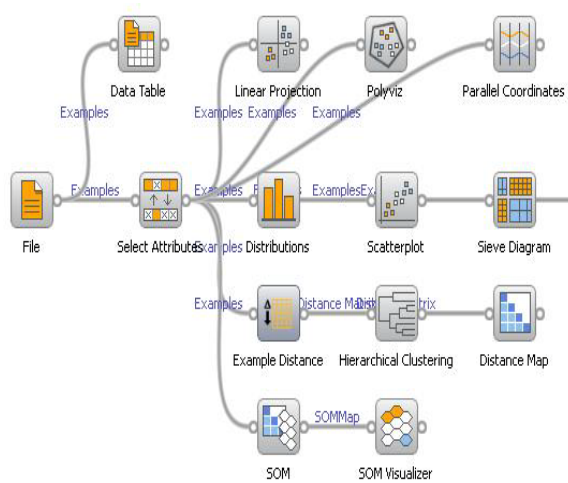
5.2 Gradniki znanja in možen primer



5.3 Del primerjalne matrike med izpeljanimi kazalci

OI	UP	KEZ	NI
e-pravni izdelki	3	8	pravni informacijski sistemi
e-sestanki	1	3	skupinsko delo
aplikacija pogostih in povezanih izposoj	1	2	sistemi za podporo odločanju
e-video konference	1	3	skupinsko delo
kartiranje znanja služb na MNZ in OVS	3	8	avtomatični analitični sistemi
inteligentni pandoraboti	1	2	ekspertni sistemi
dinamični modeli poslovnih procesov na intranetu	3	5	upravljavski informacijski sistemi
sistemi za avtomatske odgovore	3	1	ekspertni sistemi
vprašanja in odgovori	3	1	e-table
pripravljene informacije o temi	2	2	portali znanja
e-komunikacija	1	6	skupinsko delo

5.4 OZVP s pomočjo Orange Canvas



Variable: UP

Displayed outcomes: 1, 2, 3, 4, 5, 6, 7

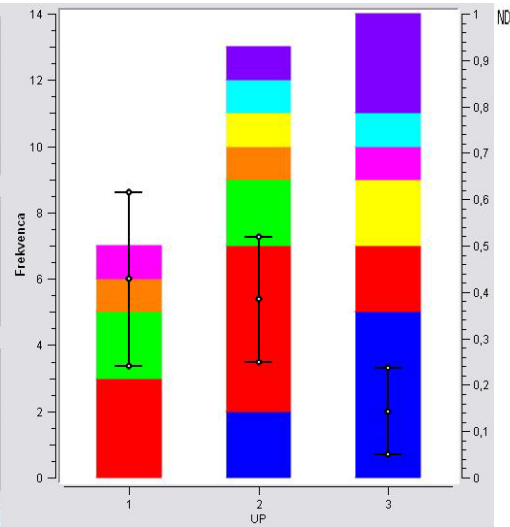
General graph settings:

- Main title
- X axis title: UP
- Y axis title: Frekvenca
- Show continuous class graph
- Number of bars: 5

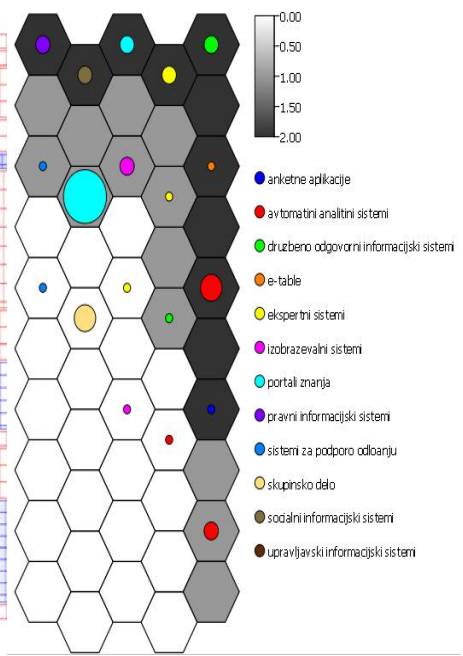
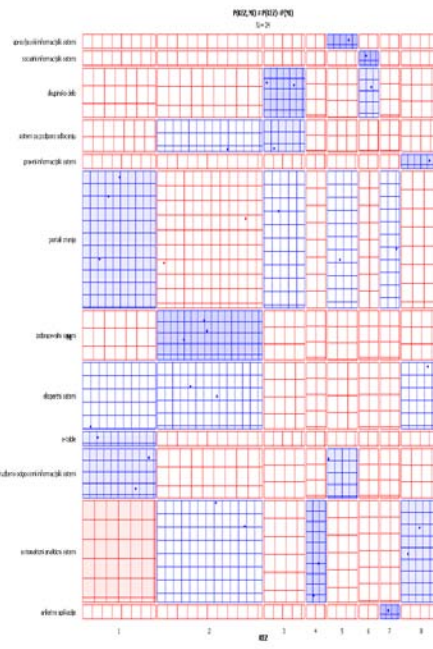
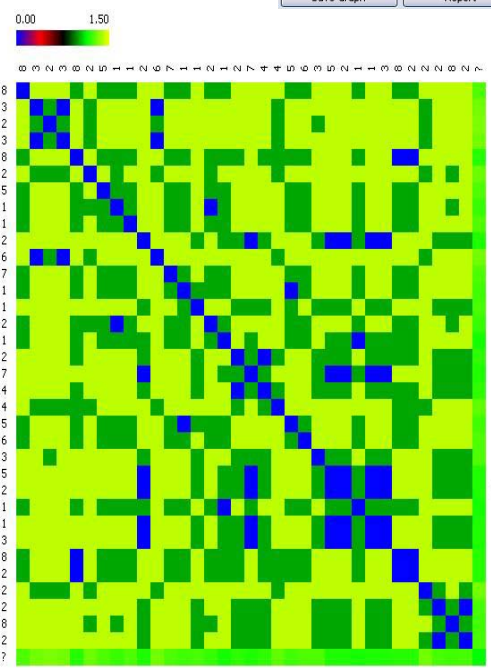
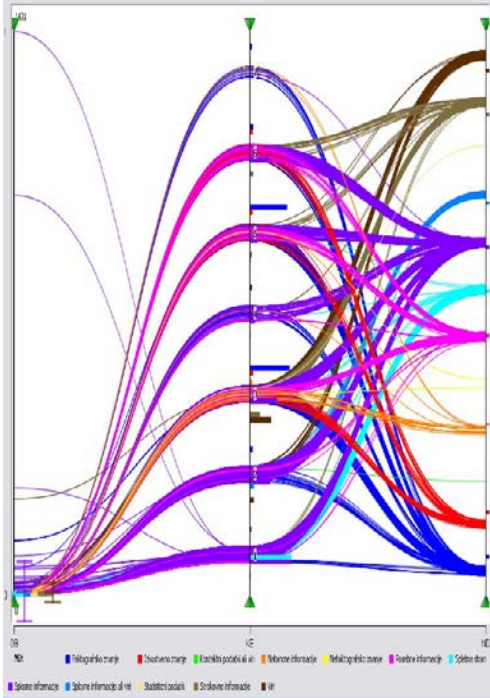
Probability plot:

- Show probabilities
- Target value: 2
- Axis title
- Show confidence intervals
- Smooth probability lines

Buttons: Save Graph, Report



Category	Total Values
Faktografsko znanje	386 (27.2%)
Izkustveno znanje	33 (2.3%)
Kontaktirni podatki ali viri	1 (0.1%)
Natančne informacije	13 (0.9%)
Nefaktografsko znanje	2 (0.1%)
Posebne informacije	57 (4.0%)
Spletna stran	196 (13.8%)
Splosne informacije	533 (37.6%)
Splosne informacije ali viri	16 (1.1%)
Statistični podatki	1 (0.1%)
Strokovne informacije	70 (4.9%)
Viri	110 (7.8%)



6 Zaključek

- Skratka na podlagi številnih predhodnih preučevanj, lahko odgovorimo na že zastavljeno vprašanje:
- **»Kakšno/-e aplikacijo/-e in/ali aplikacijski/-e sistem/-e bi lahko na podlagi ekstrahiranih novih znanj razvili, da čim bolj celovito zadovoljimo informacijske potrebe javnih uslužbencev MNZ in OVS ter drugih državljanov?«**
- Aplikacijski sistem za MNZ in OVS naj bi bil celovit in podprt s koristnimi aplikacijami ter naj bi zadovoljil naslednja pričakovanja:

6 Zaključek – nadaljevanje

- 1. Informacijski sistem naj bi bil usmerjen v družbeno problematiko, ki bi se ukvarjal z reševanjem kompleksnih družbenih problemov (npr. kriminaliteta, migracije, onesnaževanje okolja, družinsko nasilje, pravna vprašanja)
- 2. Zbrana znanja znotraj portala znanja (npr. primerjava različnih zakonodaj znotraj Evropske unije)
- 3. Zbrana izvedenska znanja in izkušnje znotraj dinamičnega ekspertnega sistema, ki omogoča komunikacijo med eksperti in drugimi (npr. omrežje ekspertov, pandoraboti – avtomatični in hitri odgovori)
- 4. Znotraj celovite informacijske rešitve naj bi bilo tudi možno izvesti izobraževanja in usposabljanja (npr. javni uslužbenci)
- 5. V celovito informacijsko rešitev naj bi bila tudi vgrajen sistem za avtomatično analizo podatkov in procesov (npr. avtomatična analiza uporabljenih in potrebnih znanj znotraj različnih služb na MNZ in OVS, avtomatična analiza poslovnih procesov).
- Kako bi takšno celovito informacijsko rešitev poimenovali z eno besedo je stvar ideje in dogovora med razvijalci ter drugimi udeleženci (npr. DOMIIS – družbeno odgovorni mrežni inteligenčni informacijski sistem?)

8 Viri

- [1] CAPPELLIN, R. & WINK, R. International knowledge and innovation networks : knowledge creation and innovation in medium-technology clusters. Cheltenham ; Northampton : Edward Elgar, 2009. VI, 275 str. (New horizons in regional science)

- [2] DIGIACOMO, J. Implementing knowledge management as a strategic initiative : thesis. Monterey: [J. DiGiacomo], 2003. 99 str.
- [3] Hafner, T. (2008). Poslovna uporaba orodja Google analytics : diplomsko delo. Kranj : [T. Hafner], 88 f., [3] f. pril.
- [4] MIAO, Q., Li, Q. & Zeng, D. (2010). Fine-grained opinion mining by integrating multiple review sources. V: Journal of the American Society for Information Science and Technology. - ISSN 1532-2882. - Vol. 61, no 11, str. 2288 – 2299
- [5] PETRIČ, K. (2008). Model adaptivne digitalne knjižnice na podlagi izdelave profila uporabnika : doktorska disertacija. Ljubljana: [K. Petrič], 187 str.
Dostopno na spletu:
<http://eprints.fri.uni-lj.si/701/1/petric-1.pdf> (2011-01-25)
- [6] PETRIČ, K. (2011). Odkrivanje novih znanj iz poizvedb na spletnih straneh Specialne knjižnice MNZ : obdobje od 27.11.2008 do 01.10.2010. Ljubljana: Ministrstvo za notranje zadeve RS, 58 str.
Dostopno na spletu:
http://www.mnz.gov.si/fileadmin/mnz.gov.si/pageuploads/SK/slike/2011/E_publikacije_2011/OZVP_poizvedbe_2010.pdf (2011-01-27)
- [7] TEXT mining : classification, clustering, and applications. Boca Raton ; London ; New York : CRC Press, 2009. XXX, 290 str., [8] str. pril. (Chapman & Hall/CRC data mining and knowledge discovery series)
- [8] THELWALL, M. A., Wilkinson, D. & Uppal, S. (2010). Data mining emotion in social network communication : gender differences in MySpace. V: Journal of the American Society for Information Science and Technology. - ISSN 1532-2882. - Vol. 61, no 1, str. 190 - 199.

Vprašanja, predlogi ipd.

