

SVEUČILIŠTE U ZAGREBU
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 366

**RAČUNALNO OTKRIVANJE MREŽE
POVEZANIH POJMOVA**

Eva Štos

Zagreb, lipanj 2012.

Sadržaj

Uvod	1
1. Pregled postojećih programskih alata.....	2
1.1. Web-tražilice	4
1.1.1. Cluuz.....	4
1.1.2. Hakia.....	6
1.1.3. Microsoft Academic Search	8
1.1.4. iSeek	9
1.1.5. MnemoMap	9
1.1.6. TouchGraph SEO Browser.....	11
1.1.7. WikiMindMap	11
1.1.8. VisionLink Search	12
1.1.9. Thinkpedia	13
1.1.10. Thinkbase	14
1.1.11. SILObreaker Network	15
1.1.12. MarketVisual	16
1.1.13. Bibex Social Network.....	17
1.2. Primijenjena programska podrška	18
1.2.1. Maltego.....	18
1.2.2. NetGlub	19
1.2.3. Sentinel Visualizer.....	20
1.2.4. Bibex – Bibliographic Exploration Tool	21
1.2.5. HistCite.....	22
1.2.6. Cogito Intelligence Platform	23
1.2.7. XANALYS Link Explorer	24
1.2.8. Centrifuge Visual Analytics Network	26

1.2.9.	IQeXplore	26
1.2.10.	Vizster.....	28
1.2.11.	OpenCalais	28
1.3.	Sažetak programskih alata.....	30
2.	Maltego.....	36
2.1.	Entiteti i transformacije	37
2.2.	Transform Distribution Server (TDS)	39
2.3.	Dostupnost i podržane platforme.....	40
3.	Pisanje transformacija za Maltego.....	42
3.1.	Entiteti	43
3.1.1.	Pregled entiteta	44
3.2.	Transformacije.....	45
3.2.1.	Amazon.....	48
3.2.2.	Library of Congress, knjižnični sustavi FER-a i FFZG-a	48
3.2.3.	Google Scholar	49
3.2.4.	Općenito	49
3.2.5.	Pregled transformacija.....	50
4.	Upute za pisanje transformacija	55
	Zaključak	57
	Literatura	58
	Sažetak.....	60
	Summary.....	61
	Skraćenice.....	62
	Privitak	63

Uvod

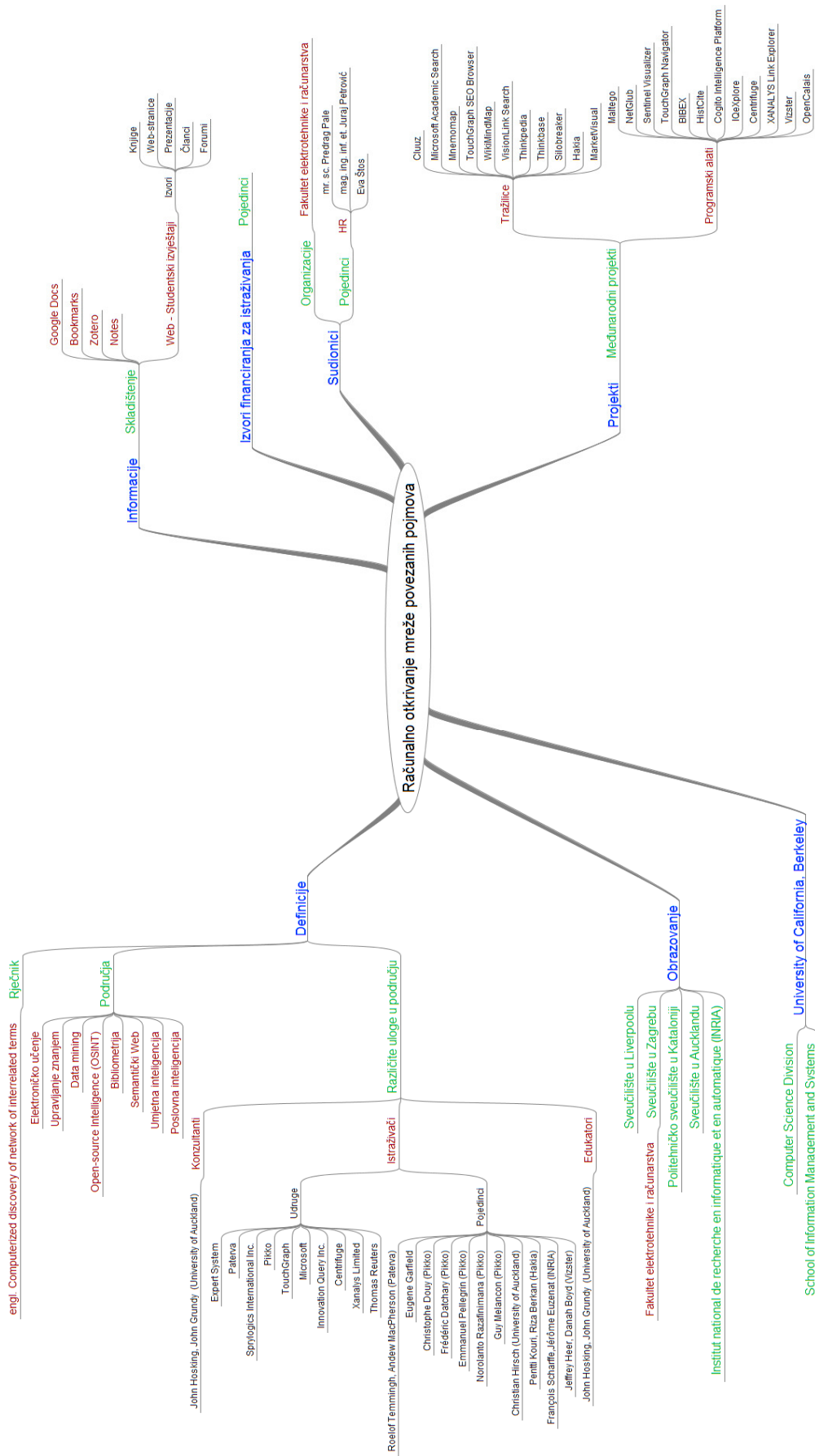
Na početku učenja ili istraživanja nekog područja ključno je obuhvatiti krupnu, grubu sliku područja kako bi se daljnji naponi na učenju i istraživanju pravilno usmjeravali. Internet pruža veliku količinu informacija o predmetnom području, međutim potrebno je uložiti veliku količinu rada i vremena u pronalaženje povezanih pojmova i definiranje mreže koja ih povezuje. Većina raspoloživih web-alata za pretraživanje pruža linearan ispis rezultata pri čemu je nemoguće uvidjeti međuovisnosti rezultata te važnost pojedinačnih rezultata u području pretraživanja. Stoga bi prilikom jednog takvog pretraživanja veliku pomoć pružila mogućnost automatskog, računalnog pronalaženja povezanih pojmova i definiranje njihove mreže.

Zadatak je ovog rada istražiti stanje znanosti i tehnike u području automatiziranog prikazivanja znanja pomoću generirane mreže povezanih pojmova važnih za neko područje te evaluirati raspoložive programske alate. Naglasak je potrebno staviti na jednostavnost upotrebe za korisnike s minimalnim informatičkim vještinama te na dobavljalivost i troškove korištenja alata. Konačno, potrebno je prilagoditi jedan takav alat i istražiti njegovu upotrebljalivost za učenje i istraživanje u nekom tehničkom području.

U prvom poglavlju ovog rada opisani su postojeći alati i njihove mogućnosti. U narednim poglavljima predstavljen je alat Maltego i rad vezan uz taj alat. Također, uz upute za korištenje dosadašnjih rezultata rada, navedene su i upute za budući rad u Maltegu.

1. Pregled postojećih programskih alata

U sklopu prvog dijela rada vezanog uz istraživanje pronađeni su i klasificirani programski alati koji pružaju neke od mogućnosti kvalitetnijeg pretraživanja, analize podataka te vizualizacije veza među podacima. Također, identificirana su područja znanosti koja se bave dotičnom tematikom, a mapa dotičnog područja vidi se na slici 1.1:



Slika 1.1 Mapa područja

Pošto je tema rada poprilično sveobuhvatna tako se i područja koja se dotiču te teme poprilično razlikuju, a pronađene programske alate (engl. *Tools*) moguće je ugrubo svrstati pod sljedeće vrste alata:

1. Web-tražilice (engl. *Search engines*)
2. Primijenjena programska podrška/aplikacije (engl. *Applications*)

1.1. Web-tražilice

Vizualno pretraživanje relativno je novi trend vezan uz pretraživanje na Webu, pogotovo od pojave semantičkih tehnologija¹. Veliki broj trenutno dostupnih vizualnih tražilica podržava različite načine vizualizacije pomoću grupiranja rezultata, interaktivnih grafova, korištenja boja na grafovima, mogućnosti proširivanja pretrage itd. Neke od takvih tražilica koje mogu biti korisne pri kvalitetnom pretraživanju vezanom uz učenje i akademsko istraživanje opisane su u nastavku rada.

1.1.1. Cluuz

Cluuz² je web-tražilica koja omogućava pretraživanje pojma, ispis top rezultata, grafički prikaz mreže vezanih pojmova, ispis najvažnijih web-stranica te entiteta vezanih uz svaku od web-stranica. Pretraživanje se može odvijati po sljedećim vrstama pojmova:

- Osobe
- Tvrtke/organizacije
- Telefonski brojevi
- E-mail adrese
- Adrese

¹ Semantika se kao grana lingvistike bavi proučavanjem značenja jezičnih znakova, dok s druge strane informacijske znanosti nisu usredotočene na značenje informacija. Međutim, područja umjetne inteligencije (engl. *Artificial intelligence*) kao što su obradba prirodnog jezika (engl. *Natural language processing*), strojno učenje (engl. *Machine learning*) i strojno prevođenje (engl. *Machine translation*) zanima obradba prirodnog jezika na svim jezičnim razinama, pa i na semantičkoj. Semantičke tehnologije ne zanima struktura poveznica među različitim mrežnim sjedištima, odnosno dokumentima, nego odnosi među elementima i njihovim svojstvima. Utvrđivanjem odnosa među elementima i njihovim svojstvima s pomoću metapodataka omogućava se strukturiranje nestrukturiranih ili polustrukturiranih podataka na mreži. Semantičke tehnologije uključuju alate za automatsko prepoznavanje tema i koncepata, ekstrakciju značenja i informacija te kategorizaciju podataka.

² <http://www.cluuz.com>

- Domene
- Datumi
- Geografija
- Ostalo

Domene pretraživanja su Web ili News, a alat je besplatan i koristi se na engleskom jeziku.

“Cluuz radi neki posao za vas. Standardna tražilica kao rezultat prikazuje popis linkova. Cluuz pretražuje Web stranice, izdvaja bitne pojmove i slike, oblikuje ih u skupine i prikazuje u obliku grafa povezanih pojmova (semantičkog grafa) gdje možete kliknuti na bilo koji entitet te dodatno usredotočiti pretraživanje.” [2]

Primjer izvođenja pretrage za pojam „firewall“ moguće je vidjeti na slici 1.2:

The screenshot shows the Cluuz search interface. At the top, there are navigation links: Home, Search Basics, How to use Cluuz?, Terms of Use, Missing Link?, and About. The search bar contains the word "firewall" and a "Search" button. Below the search bar, there are links for "Advanced" and "Preferences", and radio buttons for "Web" (selected) and "News".

The search results section shows: "Search for firewall found: 10 out of 70100000 results in 0.41 secs;". Under "Top Cluuz", there are three results:

1. Internet Firewalls: Frequently Asked Questions, compiled by Matt Curtin, Marcus Ranum and Paul Robertson (25)
2. Firewall (film), a 2006 thriller film written by Joe Forte, starring Harrison Ford (24)
3. SearchNetworking.com provides links to more about firewalls in general (20)

On the right side, there is a "Top Linked Entities" section with two items:

1. Creative Commons Att.. +
2. en.wikipedia.org +

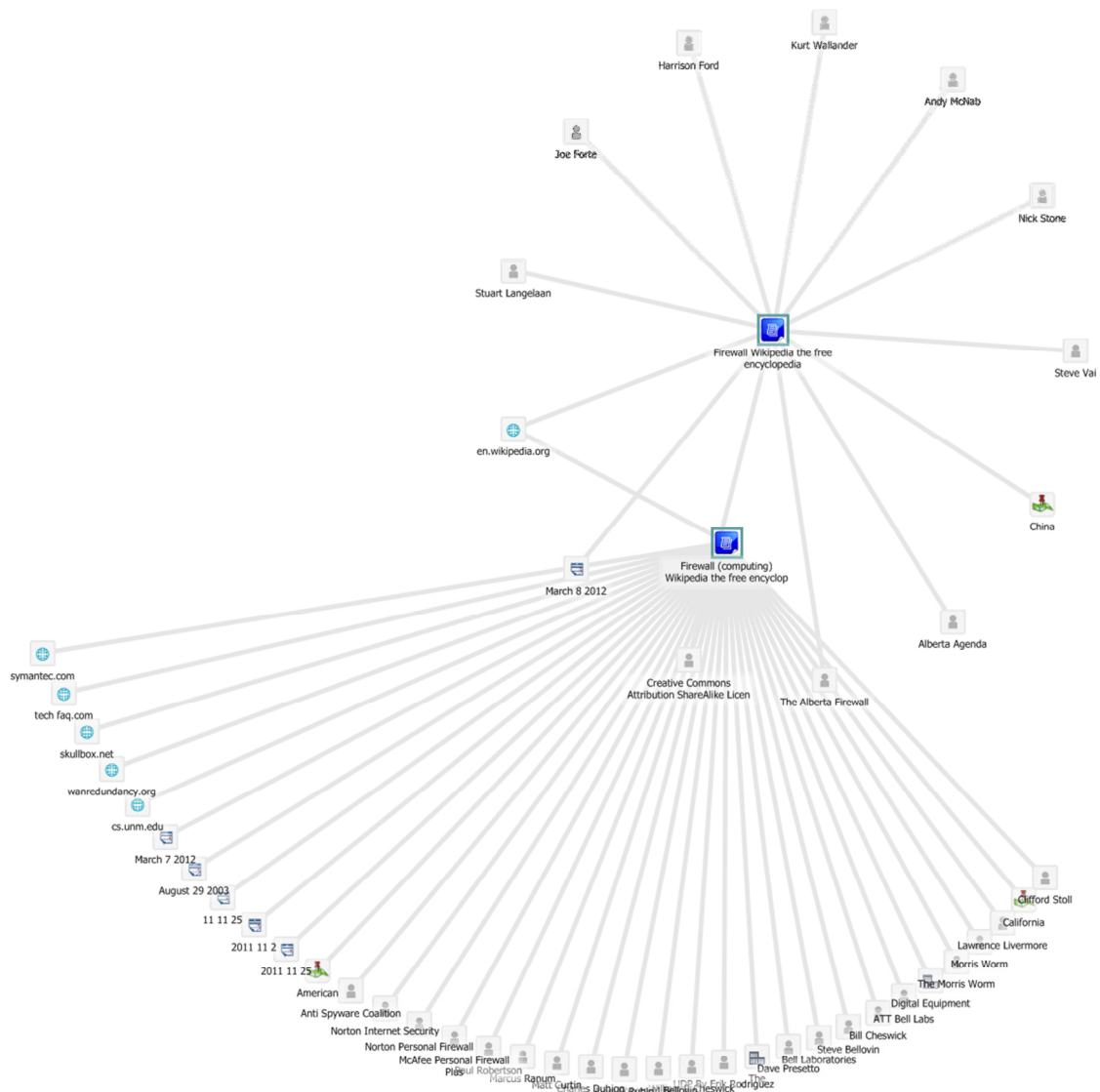
Below the search results, there are two semantic graphs. The first graph is titled "Stuart Langelaan (50..)" and shows a central node "Stuart Langelaan" connected to various other nodes including "Andy McNab", "Joe Forte", "Kurt Wallander", "Harrison Ford", "Nick Stone", "Creative Commons Attribution ShareAlike License", and "en.wikipedia.org". The second graph is titled "Lisa Phifer (23)" and shows a central node "Lisa Phifer" connected to "Den Kaminsky", "RSA Conference", and "Tim Rains".

At the bottom, there are two Wikipedia snippets for "Firewall". The first snippet is for "Firewall - Wikipedia, the free encyclopedia" and includes a small graphic of puzzle pieces. The second snippet is for "Firewall (computing) - Wikipedia, the free encyclopedia" and also includes a small graphic of puzzle pieces.

Slika 1.2 Cluuz search – pretraga za pojam „firewall“

Na desnoj strani ispisa rezultata (slika 1.2) mogu se vidjeti semantički interaktivni grafovi.

Primjer jednog od grafova vidi se na slici 1.3:



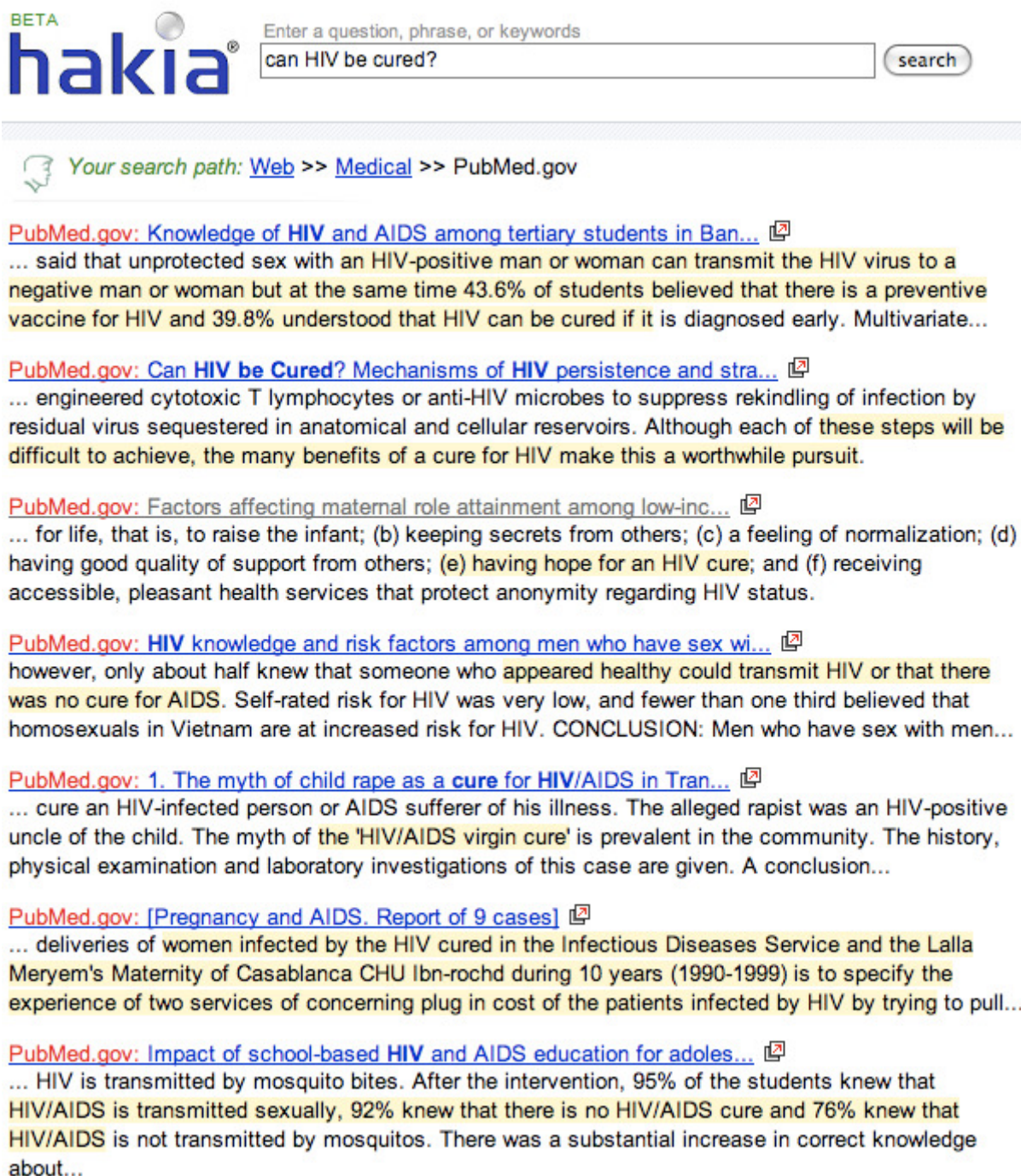
Slika 1.3 Cluuz search – semantički graf za pojam „firewall“

1.1.2. Hakia

Hakia³ je semantička tražilica koja ne ispisuje rezultate samo na temelju podudarnosti traženih riječi na web-stranicama već koristi mehanizme kojima pokušava oponašati ljudske kognitivne sposobnosti prepoznavanja onoga što je bitno u web-stranicama. Također, alat svrstava rezultate u grupe kao što su: službene stranice, biografije, slike, vijesti i intervjui, fan stranice, nagrade, govori, mitovi, polemike, resursi, inovacije,

³ <http://hakia.com/>

statistike, literatura i filmografija. Upiti se pišu na engleskom jeziku, a domena pretraživanja je Web. Iako nema mogućnost vizualizacije, Hakia je navedena kao jedna od korisnih web-tražilica zbog semantičkih metoda pretrage i odabira rezultata.



The screenshot shows the Hakia search engine interface. At the top left is the Hakia logo with 'BETA' above it. To the right is a search bar with the text 'Enter a question, phrase, or keywords' and a search button. The search query 'can HIV be cured?' is entered in the search bar. Below the search bar, the search path is shown: 'Your search path: Web >> Medical >> PubMed.gov'. The search results are listed below, each with a title and a snippet of text. The results are:

- PubMed.gov: Knowledge of HIV and AIDS among tertiary students in Ban...**
... said that unprotected sex with an HIV-positive man or woman can transmit the HIV virus to a negative man or woman but at the same time 43.6% of students believed that there is a preventive vaccine for HIV and 39.8% understood that HIV can be cured if it is diagnosed early. Multivariate...
- PubMed.gov: Can HIV be Cured? Mechanisms of HIV persistence and stra...**
... engineered cytotoxic T lymphocytes or anti-HIV microbes to suppress rekindling of infection by residual virus sequestered in anatomical and cellular reservoirs. Although each of these steps will be difficult to achieve, the many benefits of a cure for HIV make this a worthwhile pursuit.
- PubMed.gov: Factors affecting maternal role attainment among low-inc...**
... for life, that is, to raise the infant; (b) keeping secrets from others; (c) a feeling of normalization; (d) having good quality of support from others; (e) having hope for an HIV cure; and (f) receiving accessible, pleasant health services that protect anonymity regarding HIV status.
- PubMed.gov: HIV knowledge and risk factors among men who have sex wi...**
however, only about half knew that someone who appeared healthy could transmit HIV or that there was no cure for AIDS. Self-rated risk for HIV was very low, and fewer than one third believed that homosexuals in Vietnam are at increased risk for HIV. CONCLUSION: Men who have sex with men...
- PubMed.gov: 1. The myth of child rape as a cure for HIV/AIDS in Tran...**
... cure an HIV-infected person or AIDS sufferer of his illness. The alleged rapist was an HIV-positive uncle of the child. The myth of the 'HIV/AIDS virgin cure' is prevalent in the community. The history, physical examination and laboratory investigations of this case are given. A conclusion...
- PubMed.gov: [Pregnancy and AIDS. Report of 9 cases]**
... deliveries of women infected by the HIV cured in the Infectious Diseases Service and the Lalla Meryem's Maternity of Casablanca CHU Ibn-rochd during 10 years (1990-1999) is to specify the experience of two services of concerning plug in cost of the patients infected by HIV by trying to pull...
- PubMed.gov: Impact of school-based HIV and AIDS education for adoles...**
... HIV is transmitted by mosquito bites. After the intervention, 95% of the students knew that HIV/AIDS is transmitted sexually, 92% knew that there is no HIV/AIDS cure and 76% knew that HIV/AIDS is not transmitted by mosquitos. There was a substantial increase in correct knowledge about...

Slika 1.4 Hakia – primjer pretrage za „can HIV be cured?“⁴

⁴ <http://www.zdnet.com/blog/semantic-web/hakia-adds-pubmed-enhances-medical-search/138>

1.1.3. Microsoft Academic Search

Microsoft Academic Search⁵ osim pretraživanja akademskih publikacija i autora te ispisa rezultata po temama nudi i neke mogućnosti vizualizacije. Pretragu je moguće ograničiti po određenim područjima te pretraživati sljedeće vrste pojmova: Author, Conference, Journal, Organization, Year, DOI. Također moguće je koristiti sljedeće vizualizacije (slika 1.5):

- Academic Map – prikazuje geografski organizacije i autore vezane uz određenu domenu
- CFP Calendar – pretražuje konferencije po domenu, vremenu i geografskoj lokaciji
- Domain Trend – vizualizacija istraživačkih trendova u području računalne znanosti
- Organization Comparison – uspoređivanje dvaju organizacija
- Co-author Graph – prikaz relacija između traženog autora i povezanih istraživača
- Co-author Path – prikaz relacija između dva istraživača preko zajedničkog koautora
- Genealogy Graph – nije trenutno u funkciji
- Paper Citation Graph - nije jasno kako koristiti opciju ili nije u funkciji

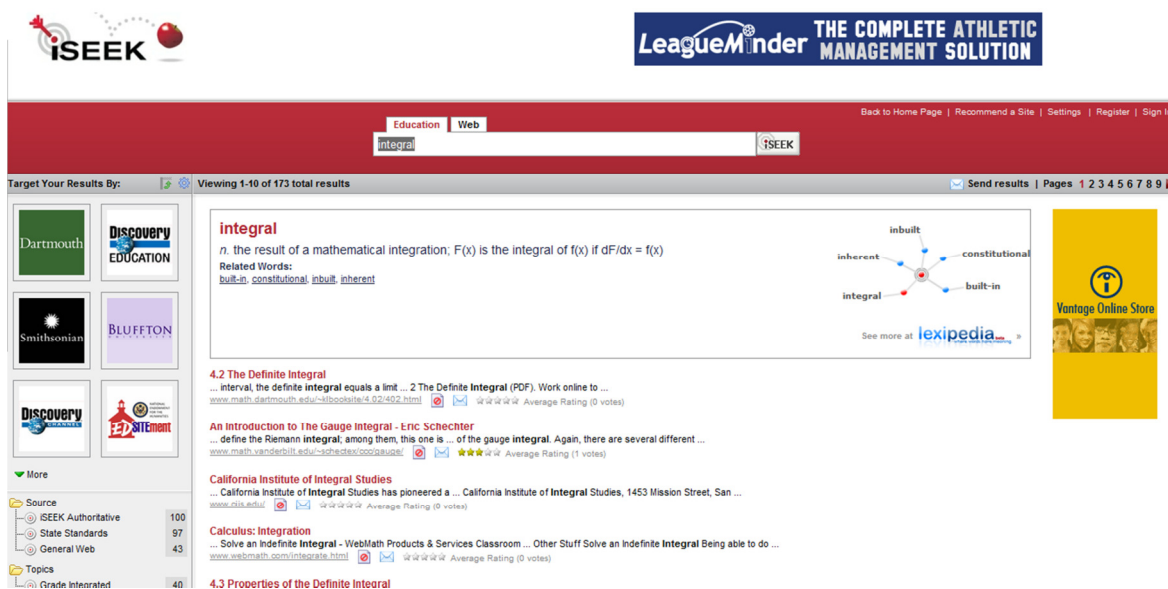


Slika 1.5 Microsoft Academic Search – mogućnosti vizualizacije⁵

⁵ <http://academic.research.microsoft.com/>

1.1.4. iSeek

iSeek⁶ je tražilica nove generacije i omogućava pretraživanje u dva područja pretraživanja: Web i Education. Nema mogućnosti vizualizacije, ali omogućava ispis web-stranica, povezane pojmove, poveznica na poznate udruge i sveučilišta koja se bave traženom temom te prikaz rezultata po sljedećim skupinama: Source, Topics, Subject, Resource Type, Grade Level, Places, People, Organizations, Standards by State. Također, svaka skupina rezultata sadrži pojmove sortirane od najvećeg broja podudarnosti u web-stranicama do najmanjeg. Upite je moguće postavljati na engleskom jeziku.



Slika 1.6 iSeek – ispis rezultata za „integral“

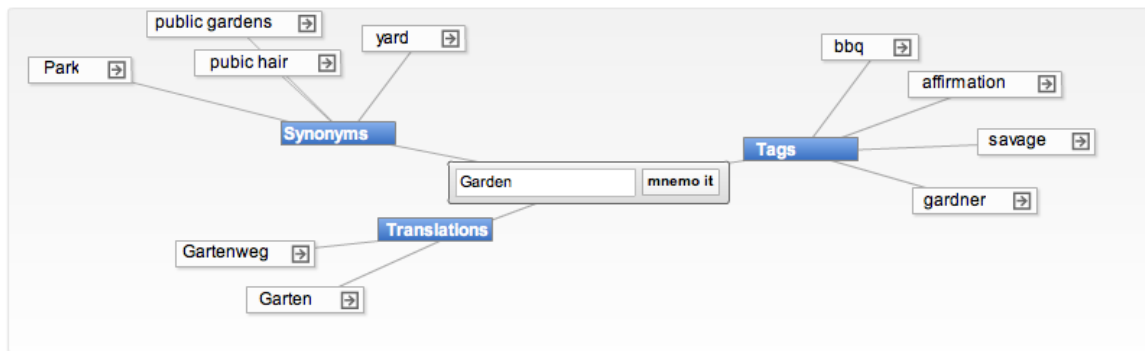
1.1.5. MnemoMap

MnemoMap⁷ je tražilica koja je još u početnom stadiju razvoja. Funkcionalnost alata se dijeli na 3 dijela:

- Atomic-Tree - Ispis sinonima, tagova i prijevoda. Svaku od pronađenih riječi moguće je postaviti kao centar atomskog stabla ili dodati kao ključnu riječ u pretraživanje.

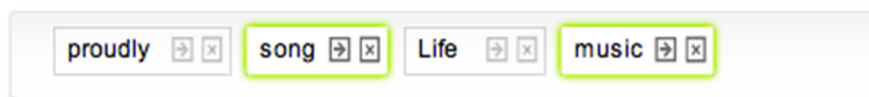
⁶ <http://www.iseek.com/iseek/home.page>

⁷ <http://www.mnemo.org/>



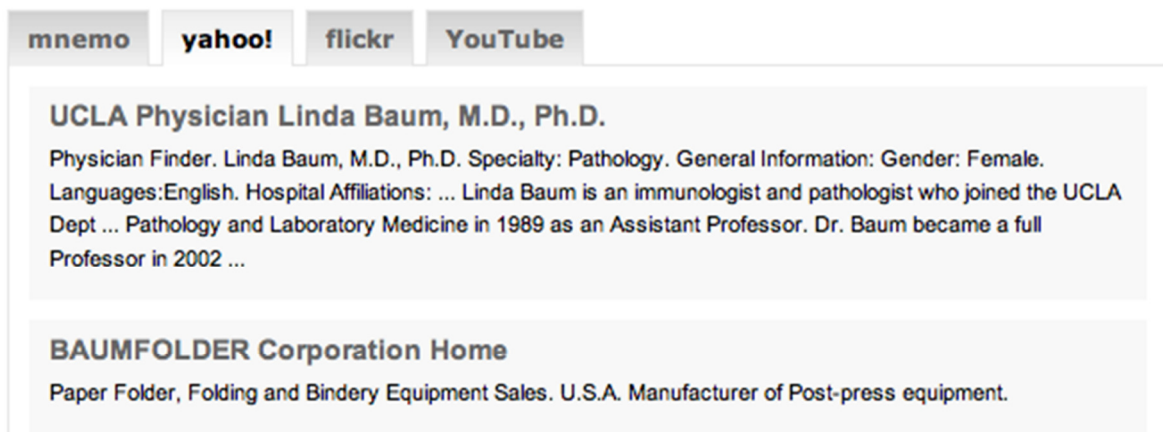
Slika 1.7 1.1.5. MnemoMap – Atomic Tree graf za primjer „Garden“

- Query List - Popis upita, moguće je postaviti status svake riječi na “aktivno” ili “pasivno”, gdje “aktivno” označava da se uključuje u pretraživanje, a “pasivno” da ne pripada.



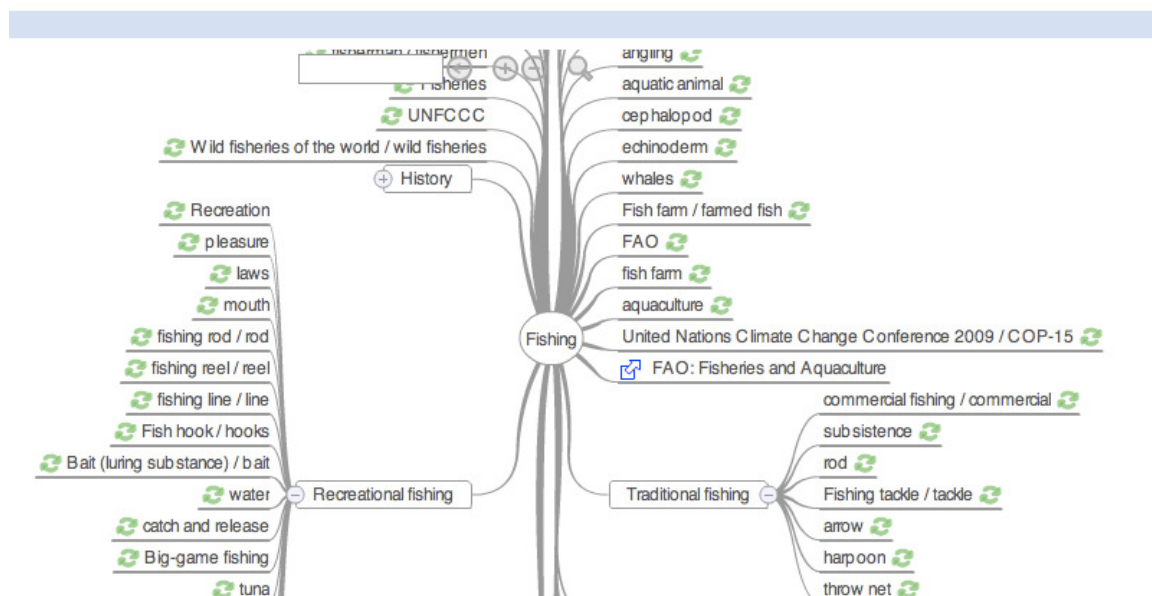
Slika 1.8 MnemoMap – Query List primjer

- Tabs - Svaki tab predstavlja popis rezultata pretrage sa određenog izvora pretraživanja.



Slika 1.9 MnemoMap – Tabs primjer

Domene pretraživanja alata MnemoMap su Web, images, digg, del.icio.us i youtube, a dostupni jezici engleski, njemački, francuski, finski, švedski, talijanski, estonski, nizozemski, katalonski, norveški i turski.

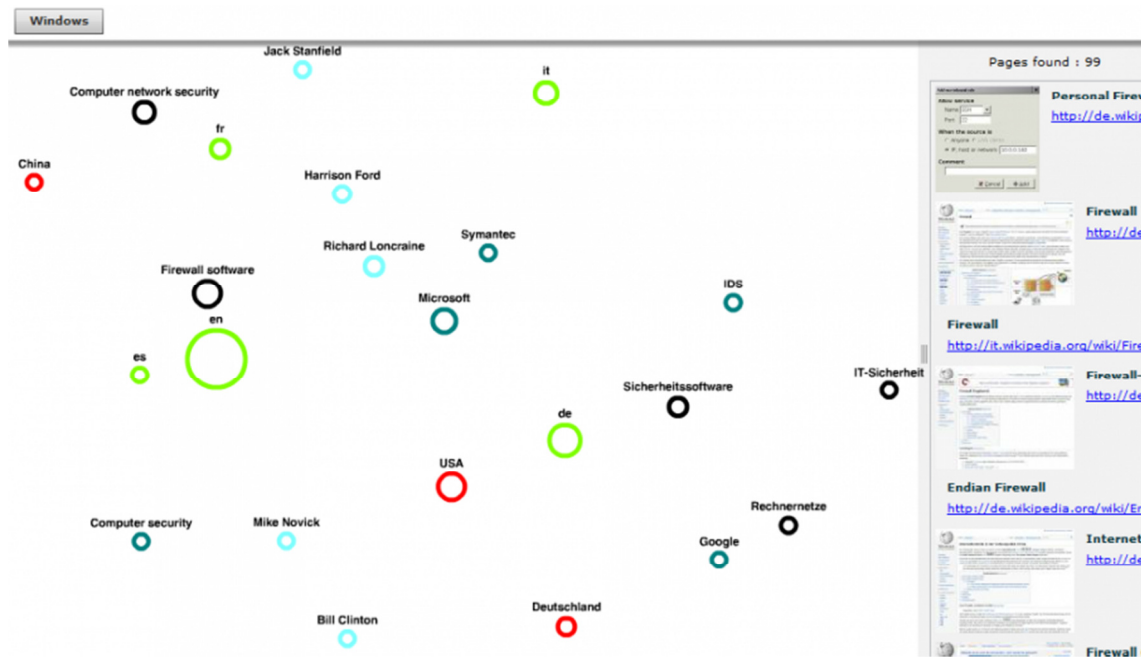


Slika 1.11 WikiMindMap – vizualizacija za pojam „Fishing“

1.1.8. VisionLink Search

VisionLink Search¹² je tražilica te alat za analizu i mapiranje informacija. Rezultat pretrage je vizualni i intuitivni prikaz rezultata i text-mininga prikazan pomoću interaktivnih relacijskih grafova. VisionLink Search omogućava ispis web-stranica te interaktivnog grafa koji predstavlja mrežu povezanih pojmova s entitetima država, jezik, osoba, organizacija i Wiki-kategorija. Prelaskom miša preko entiteta prikazuju se veze među entitetima koje je moguće odabrati, sakriti i filtrirati. Svi grafovi koje generira VisionLink temelje se na Flash tehnologiji te se mogu prikazati interaktivno unutar web-preglednika. Domena pretraživanja je Wikipedia, a podržani jezici su engleski, njemački i francuski.

¹² <http://www.pikko-software.com/information-visualization-tool.html>



Slika 1.12 VisionLink Search – vizualizacija za pojam „firewall“

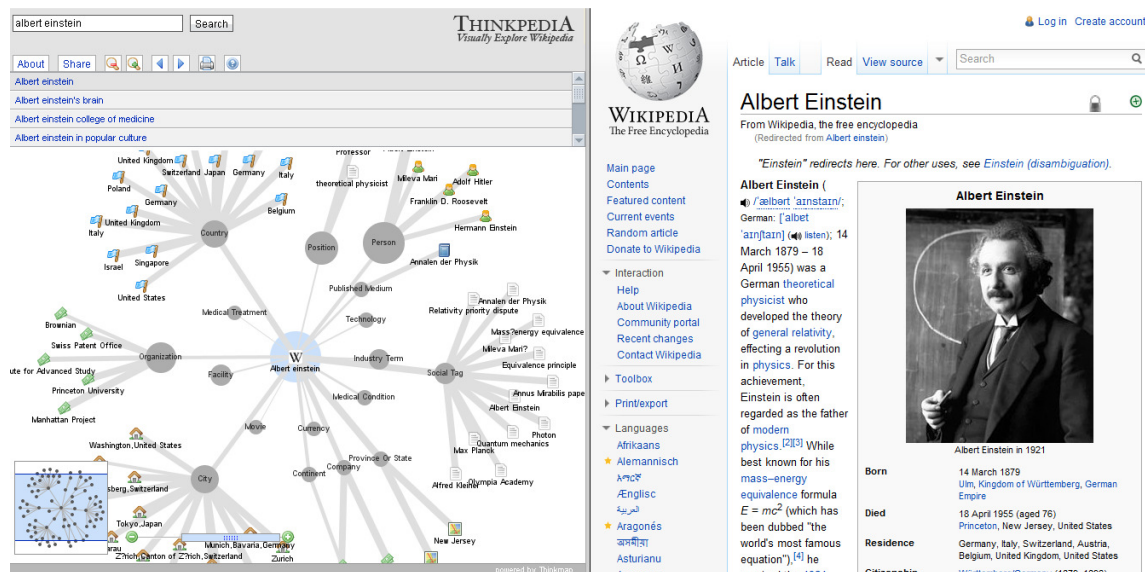
Rezultat pretrage su kružići i web-stranice prikazane u obliku poveznica i slika. Prelaskom miša preko kružića iscrtavaju se veze sa ostalim kružićima. Različite vrste rezultata su određene bojom, a veličina označava “broj” pronađenih rezultata. Npr. domene su označene svijetlozelenom bojom, zemlje crvenom, osobe svijetloplavom, organizacije/udruge/tvrtke tamnoplavom bojom, a pojmovi crnom bojom. Moguće je isprobati demo verziju tražilice¹³.

1.1.9. Thinkpedia

Thinkpedia¹⁴ je tražilica za vizualno pretraživanje Wikipedije. Rezultat pretraživanja je interaktivni graf povezanih pojmova gdje su određene grane na grafu grupirane po zemljama, tvrtkama, društvenim oznakama, organizacijama, pozicijama, osobama itd. Pretraživanje je moguće usredotočiti, a odabirom entiteta graf se proširuje.

¹³ <http://demo.pikko-software.com/exalead/>

¹⁴ <http://thinkpedia.cs.auckland.ac.nz/>

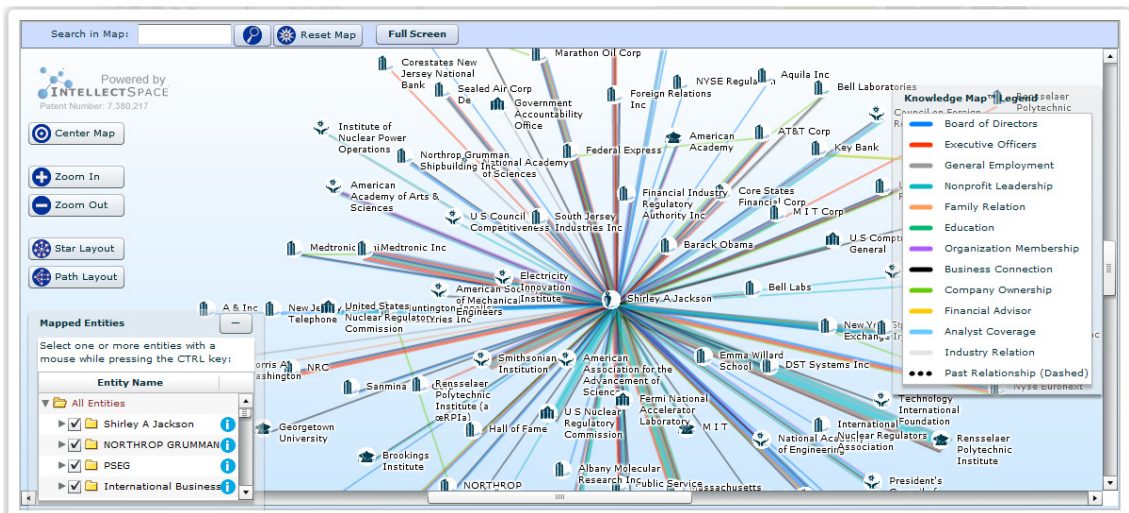


Slika 1.13 Thinkpedia – vizualizacija za pojam „Albert Einstein“

1.1.10. Thinkbase

Thinkbase¹⁵ je tražilica slična Thinkpediji za vizualno pretraživanje Wikipedije. Alat je još u fazi razvoja, a omogućava pretragu po temama na engleskom jeziku. Rezultat pretrage je interaktivni graf u kojem grane sadrže semantičku oznaku veze između središnjeg pojma i pojmova na listovima na grafu. Rezultate je moguće filtrirati po domenama, a pretraživanje je moguće usredotočiti odabirom entiteta. Graf je moguće proširiti odabirom opcije „Expand“ na nekom od entiteta.

¹⁵ <http://thinkbase.cs.auckland.ac.nz/>



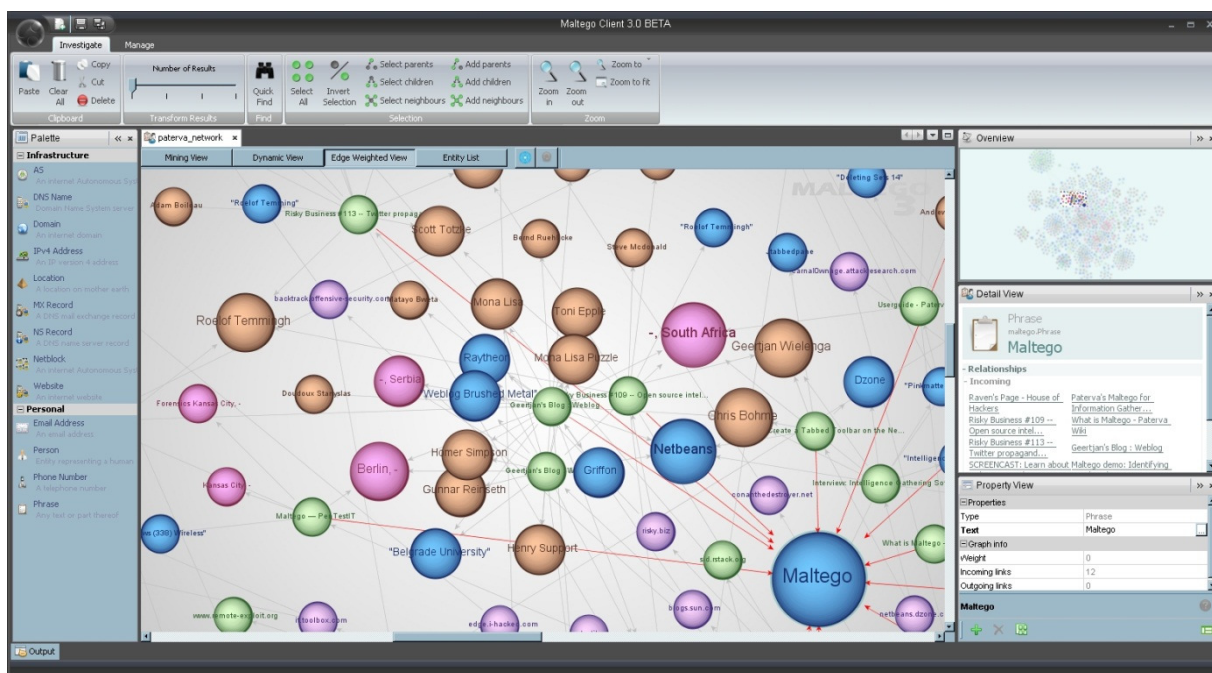
Slika 1.16 MarketVisual – primjer vizualizacije¹⁷

1.1.13. Bibex Social Network

Bibex Social Network¹⁸ je demo verzija alata Bibex (koji je opisan kasnije u radu) i dostupan je za pretraživanje autora i tema pri čemu je rezultat pretrage interaktivni graf koji predstavlja mrežu povezanih autora. Pretragu je moguće ograničiti vremenski (od-do, u godinama).

¹⁸ <http://188.121.62.42/bibex-webapp/bibex/index.action>

besplatno, a aplikacija je kompatibilna s operativnim sustavima Microsoft Windows, Mac i Linux.



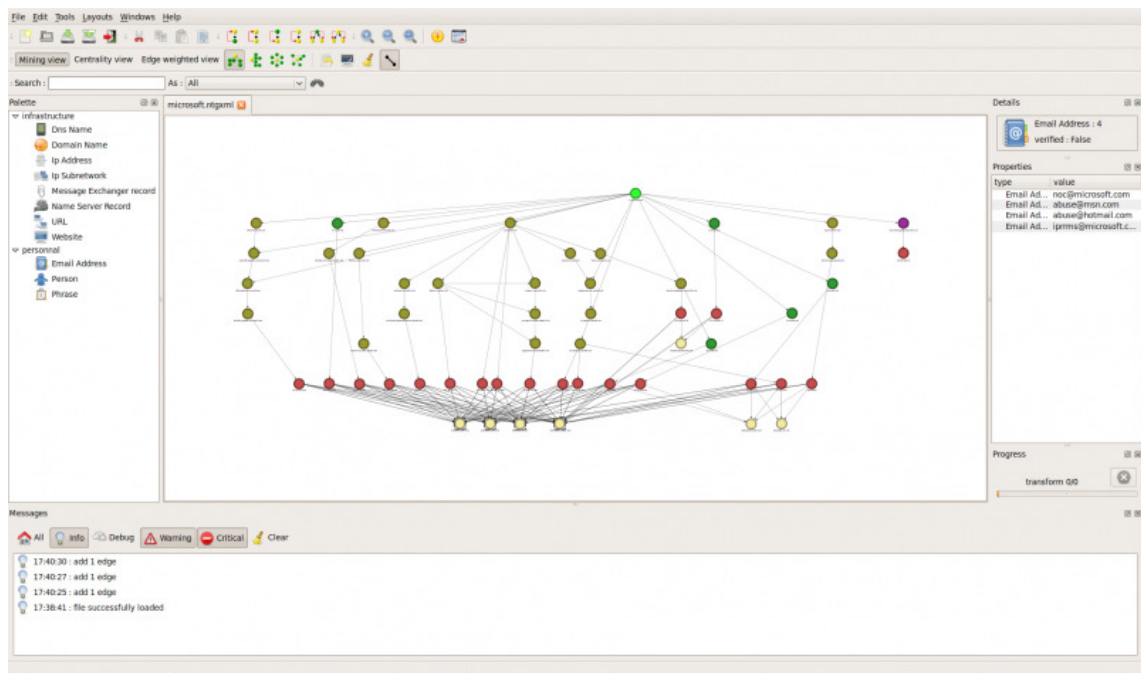
Slika 1.18 Maltego – izgled sučelja i primjer vizualizacije²⁰

1.2.2. NetGlub

NetGlub²¹ je također aplikacija otvorenog koda za prikupljanje informacija i računalnu forenziku koja pruža sučelje za prikupljanje informacija za prikaz u jednostavnom i razumljivom obliku. NetGlub je još u beta verziji i poprilično je sličan Maltegu. Preuzimanje NetGluba je besplatno, a alat je dostupan za operativni sustav Linux.

²⁰ http://download.idg.bg/software/9284_maltego

²¹ <http://www.netglub.org/>

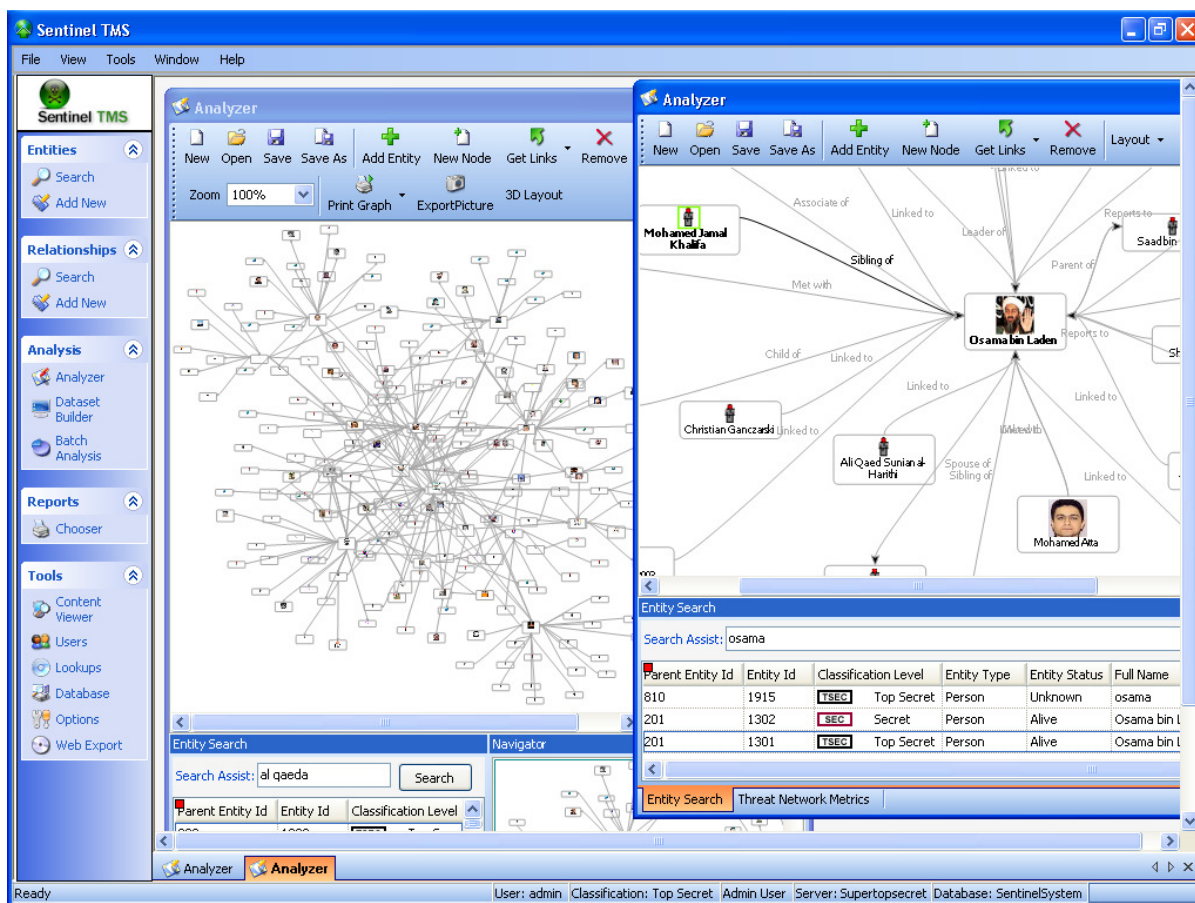


Slika 1.19 NetGlub – izgled sučelja i primjer vizualizacije²¹

1.2.3. Sentinel Visualizer

Sentinel Visualizer²² je alat za vrlo naprednu analizu društvenih mreža. Njegove mogućnosti obuhvaćaju pronalazak skrivenih odnosa među entitetima, prepoznavanje klastera i uzoraka, brzo obavljanje *ad-hoc* analize, testiranje teorija i scenarija te organiziranje složenih mreža. Dostupna izdanja alata su *standard*, *professional* i *premium*, a cijene za područje Sjeverne Amerike su \$2499, \$2999 i \$3999 respektivno, dok je cijena za internacionalne narudžbe nepoznata. Alat je namijenjen pružateljima usluga saveznoj, državnoj ili lokalnoj samoupravi, a alat je moguće koristiti na operativnim sustavima Microsoft Windows 7, Microsoft Windows Vista i Microsoft Windows XP (Service Pack 2 ili naprednije).

²² <http://www.fmsasg.com/Products/SentinelVisualizer/>

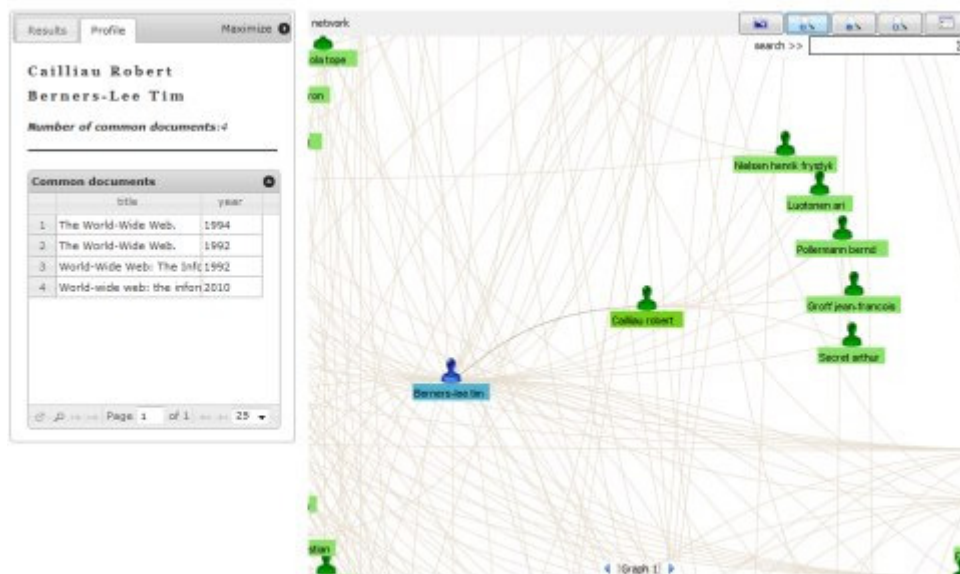


Slika 1.20 Sentinel Visualizer – izgled sučelja i primjer vizualizacije²²

1.2.4. Bibex – Bibliographic Exploration Tool

Bibex²³ je inovativni alat za istraživanje bibliografskih repozitorija i dokumenata te proizvod pet godina istraživanja i razvoja na UPC-u (Politehničkom sveučilištu u Kataloniji). Prvenstveno je namijenjen za bibliografska istraživanja, a njegove mogućnosti su izvođenje složenih bibliografskih upita i prikaz rezultata upita kao kombinaciju grafova i tekstova te istraživanje odnosa između autora, ključnih riječi i radova. Besplatna demo verzija Bibex Social Network je tražilica koja obuhvaća pretraživanje autora i vizualizaciju mreže povezanih autora, a opisana je u prethodnom poglavlju. Bibex je kompatibilan s operativnim sustavima Windows, MAC i Linux.

²³ <http://www.dama.upc.edu/technology-transfer/bibex>



Slika 1.21 Bibex – izgled sučelja i primjer vizualizacije²³

1.2.5. HistCite

HistCite²⁴ je alat koji predstavlja programsku implementaciju historiografskog²⁵ algoritma, a razvio ga je Eugene Garfield, osnivač Instituta za znanstvene informacije i izumitelj važnih alata za pronalaženje informacija, kao što su Current Contents²⁶ i Science Citation Index²⁷. HistCite je programski paket za provođenje bibliometrijskih analiza i vizualizaciju rezultata pretraživanja baze Web of Science²⁸ (Web of Knowledge). Uz pomoć alata moguće je otkriti predmetnu strukturu, povijest i međusobne veze objavljene literature, a namijenjen je historiografskim istraživanjima. Besplatna probna verzija programskog paketa dostupna je na web-stranicama alata, a za cjelokupnu funkcionalnost potreban je korisnički račun na Web of Science. HistCite se pokreće u Internet Exploreru, a

²⁴ <http://en.wikipedia.org/wiki/Histcite>

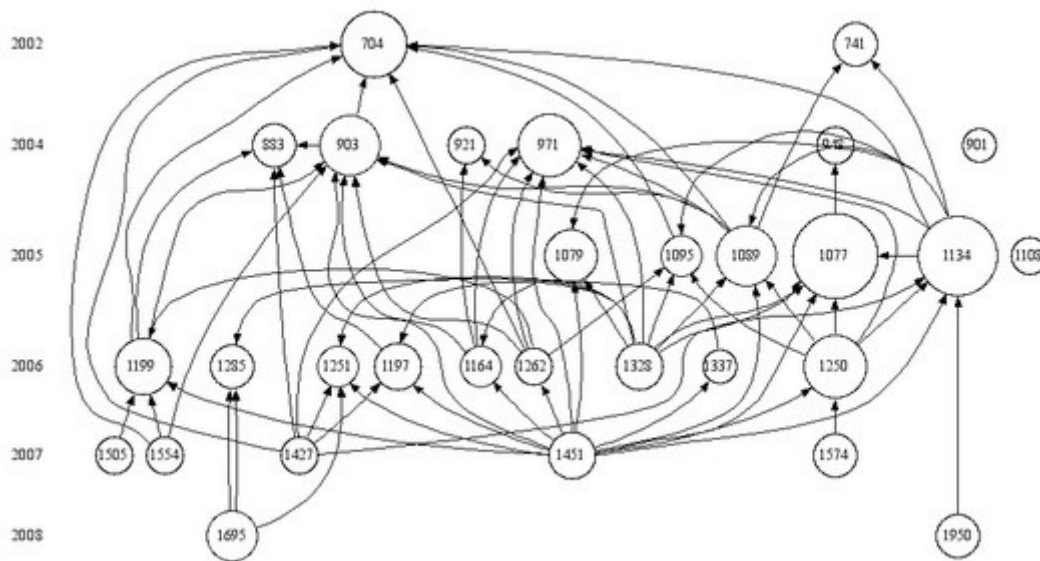
²⁵ Historiografija je znanost koja se bavi opisivanjem prošlosti ljudskog društva prema izgrađenoj metodi; historijska znanost.

²⁶ Current Contents - bibliografska baza podataka koju održava Institute for Scientific Information (ISI, Philadelphia, USA). Baza se sastoji od sadržaja brojeva časopisa i bibliografskih zapisa radova iz više od 7600 vodećih svjetskih časopisa iz svih područja znanosti te više od 2000 knjiga i zbornika skupova.

²⁷ Science Citation Index - http://thomsonreuters.com/products_services/science/science_products/a-z/science_citation_index/

²⁸ Web of Science je bibliografska baza podataka koju održava Institute for Scientific Information (ISI), od 1991. do danas. Web of Science obuhvaća oko 8700, prema ocjeni međunarodne istraživačke zajednice vodećih svjetskih časopisa iz svih područja znanosti.

probna verzija alata je testirana na jednom primjerku bibliografskog zapisa. Alat je kompatibilan s operativnim sustavom Microsoft Windows.



Slika 1.22 HistCite – primjer grafa²⁹

1.2.6. Cogito Intelligence Platform

Cogito Intelligence Platform³⁰ je napredan programski paket i predstavlja svojevrsno rješenje za Open-source Intelligence (OSINT)³¹. Alat ima mogućnost semantičke analize podataka, a namijenjen je kao podrška u radu analitičara i osoba koje se bave područjem upravljanja znanjem te prikupljanja podataka uz mogućnost otkrivanja uzoraka u informacijama i međusobnih veza te njihove vizualizacije. Alat je razvila udruga Expert System³² koja se bavi proizvodnjom programskih alata vezanih uz semantičke tehnologije. Besplatnu demo verziju alata moguće je zatražiti na web-stranici alata³³, a originalnu verziju s cjelokupnim funkcionalnostima je potrebno naručiti.

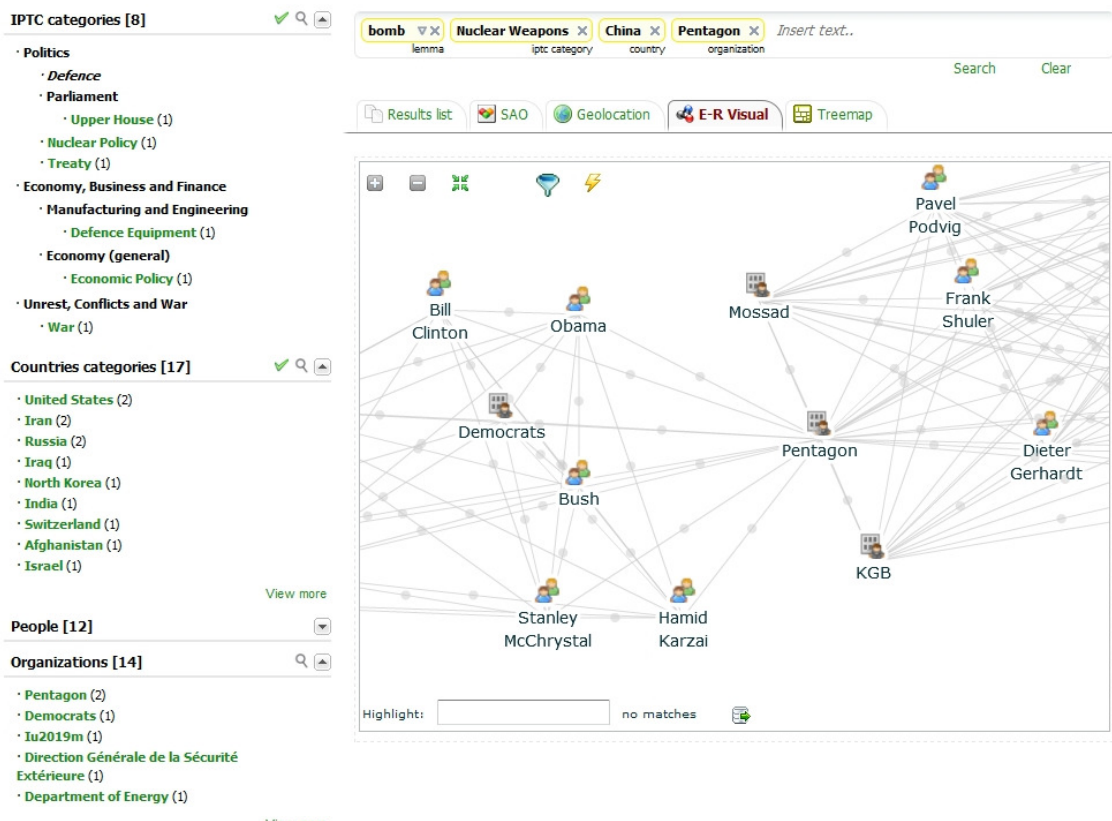
²⁹ <http://blog.sciencenet.cn/home.php?mod=space&uid=304685&do=blog&id=383399>

³⁰ <http://www.osint.it/english/cogito-intelligence-platform-osint.asp>

³¹ Open-source Intelligence (OSINT) oblik je prikupljanja i upravljanja podacima koji uključuje pronalazak, odabir i pribavljanje informacija s javno dostupnih izvora te analizu podataka u svrhu oblikovanja novih saznanja. Definicija „Open source intelligence“ se mijenjala tijekom vremena što se tiče organizacija i država. Najjednostavnije rečeno ona je „neklasificirana informacija“. Još jedna upotrebljiva definicija je „informacija s potencijalom mogućeg značaja za obavještajne službe koja je dostupna javnosti“.

³² Expert System - <http://www.expertsystem.net/>

³³ http://www.expertsystem.net/demo_prodotti.asp



Slika 1.23. Cogito Intelligence Platform – izgled sučelja i primjer grafa³⁰

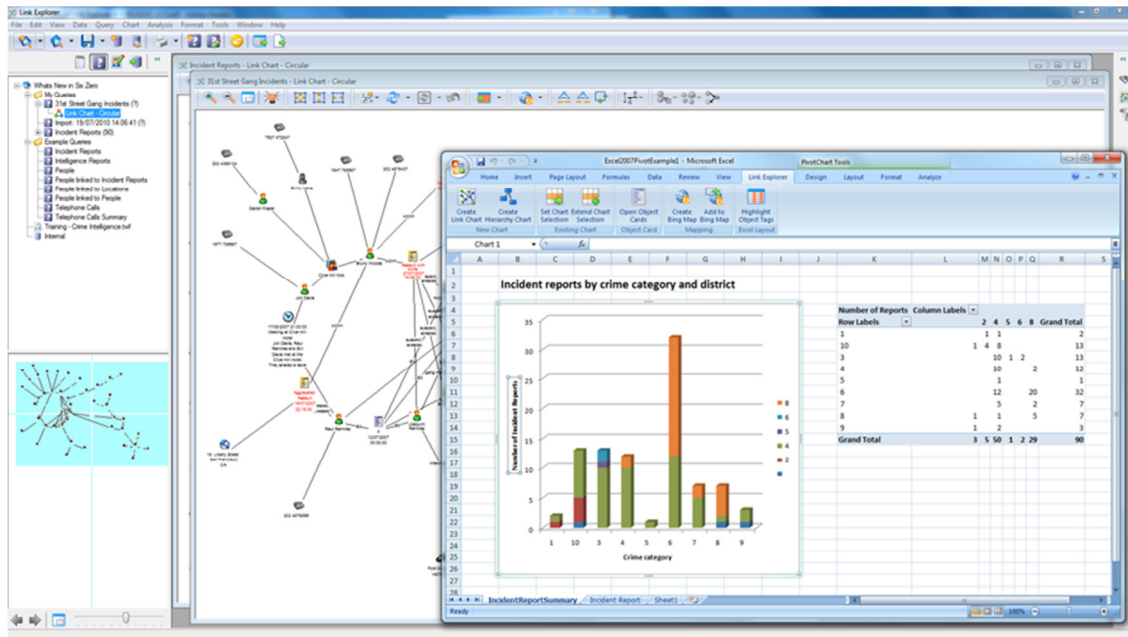
1.2.7. XANALYS Link Explorer

XANALYS Link Explorer³⁴ je *desktop* aplikacija koja ima mogućnost modeliranja podataka kao skupa objekata (ljudi, mjesta, događaja) i veza među tim podacima. XANALYS Link Explorer obrađuje podatke s raznih izvora pri čemu je moguće preuzeti podatke iz baza podataka kao što su ODBC baze podataka³⁵, uvesti podatke iz Microsoft Excel tablice ili uvesti vlastite podatke u internu bazu podataka XANALYS-a. U programskom alatu svi se podaci interpretiraju kao skup objekata i poveznica, a korisnici mogu definirati način na koji se podaci prevode u objekte i poveznice. Alat na temelju uređenih podataka na ulazu iscrtava grafikone i mreže povezanih pojmova. Besplatna verzija dostupna je na 30 dana procjene, a originalnu verziju je potrebno naručiti. Alat je

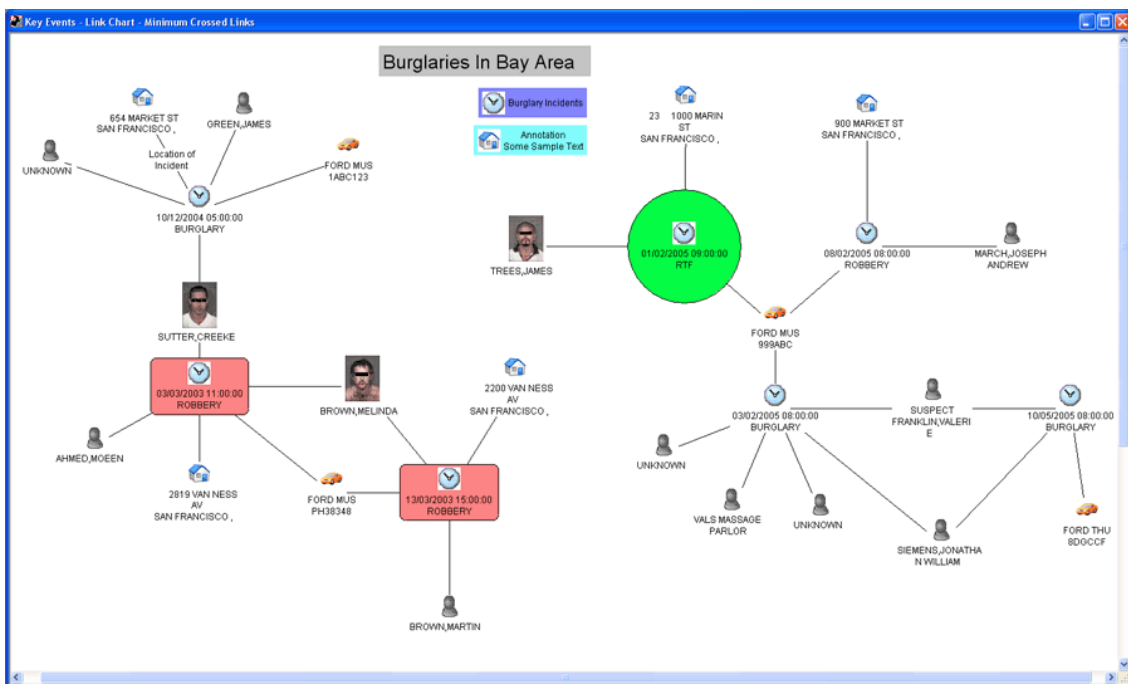
³⁴ <http://www.xanalis.com/solutions/linkexplorer.html>

³⁵ Open Database Connectivity (skraćena: ODBC) je protokol koji se koristi za povezivanje baze podataka programa Microsoft Access s vanjskim izvorom podataka kao što je Microsoft SQL Server.

kompatibilan s operativnim sustavima Windows 7, Windows Vista i Windows XP Professional (Service Pack 2 ili 3).



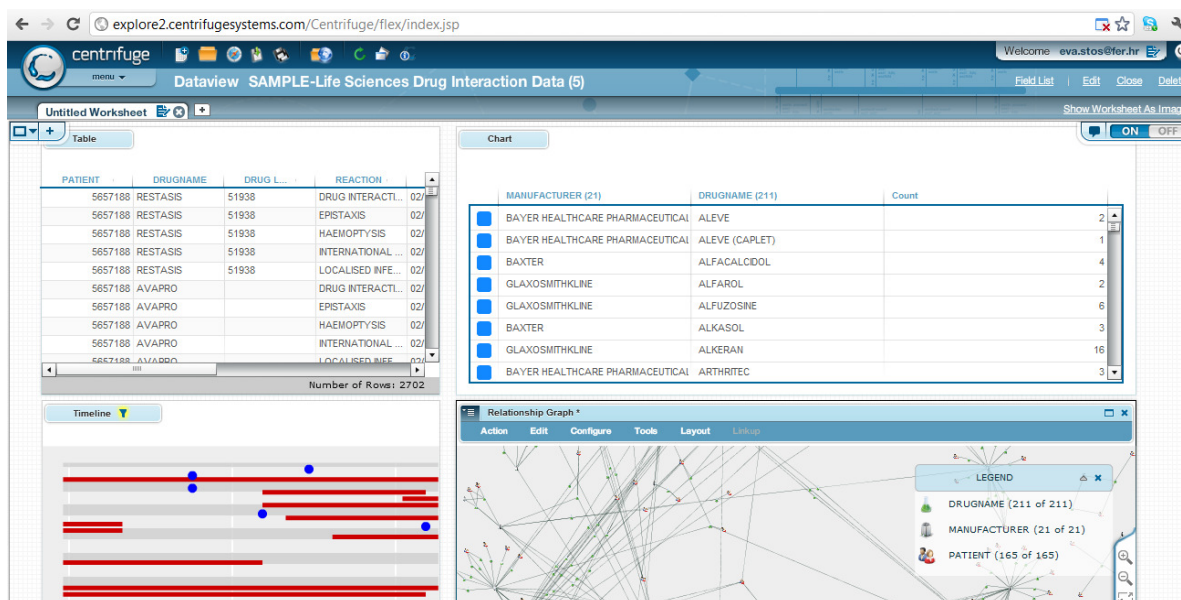
Slika 1.24 XANALYS Link Explorer – izgled sučelja³⁴



Slika 1.25 XANALYS Link Explorer – primjer vizualizacije³⁴

1.2.8. Centrifuge Visual Analytics Network

Centrifuge Visual Analytics Network³⁶ je web-aplikacija koja omogućava otkrivanje uvida u uzorke i odnose skrivene u javnim, *cloud* i društvenim mrežama i podacima vezanim uz poduzeća. Besplatna verzija je dostupna na 14 dana procjene dok je originalnu verziju potrebno naručiti. Visual Analytics Network koristi se pokretanjem web-preglednika, povezivanjem s bilo kojim brojem izvora podataka i interaktivnom vizualizacijom rezultata u obliku grafova, tablica i geoprostornih pogleda. Izvori ulaznih podataka s kojima je moguće povezati alat su Microsoft Excel, Microsoft Access, Microsoft SQL Server, Oracle Thin, PostgreSQL, Text i XML. Za pokretanje aplikacije potrebno je imati instaliran Adobe Flash Player 10, a podržani web-preglednici su Internet Explorer 7, Chrome 5 i Mozilla Firefox 3.



Slika 1.26 Centrifuge Visual Analytics Network – izgled sučelja i primjer vizualizacije

1.2.9. IQeXplore

IQeXplore³⁷ je napredna web-aplikacija čije se funkcionalnosti temelje na semantičkoj tehnologiji, a obuhvaćaju analizu podataka iz raznih izvora, modeliranje veza među entitetima, semantičko pretraživanje, izgradnju vlastite aplikacije, naprednu vizualizaciju

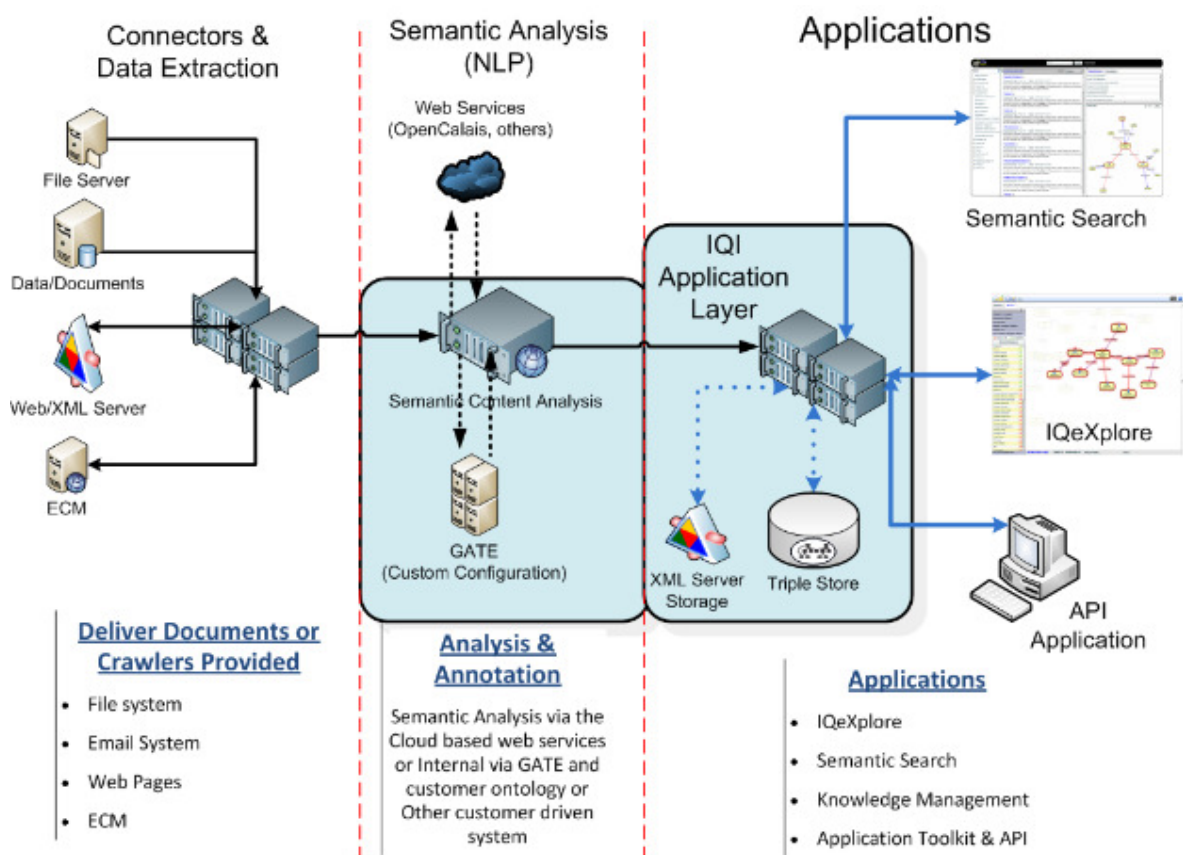
³⁶ <http://www.centrifugesystems.com/index.php>

³⁷ <http://innovativequery.com/products-and-services/iqexplore.html>

te stvaranje grafova povezanih pojmova itd. Alat je namijenjen rudarenju podataka³⁸ te složenoj analizi podataka korištenjem semantičkih tehnologija. Ne postoji besplatna verzija za procjenu alata i za detaljnije informacije o funkcionalnosti alata potrebno ga je naručiti.

“Što se vanjskih izvora podataka tiče, možemo dodati mogućnost preuzimanja podataka s bilo kojeg tekstualnog izvora - Web stranica, rezultata pretraživanja, RSS³⁹ feedova, blogova, PDF ili Word datoteka, sustava za upravljanje dokumentima, portala i plaćenih informacijskim uslugama.” [2]

Kao primjer analize raznih izvora podataka i modeliranja veza između entiteta opisan je jedan *use-case* scenarij korištenja IQeXplore-a. [23]

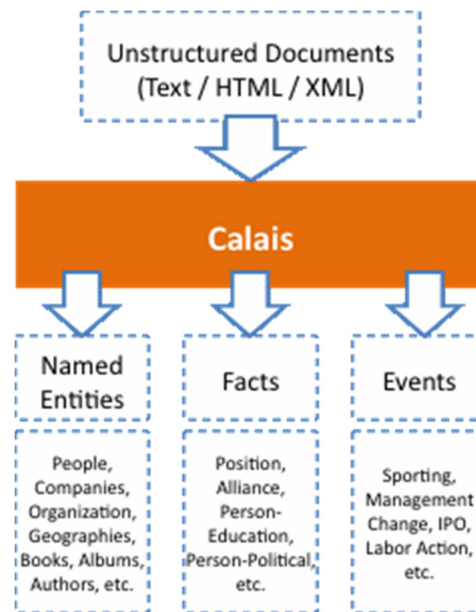


Slika 1.27 IQeXplore – konceptualna shema alata³⁷

³⁸ Rudarenje podataka (engl. *Data mining*) je sortiranje, organiziranje ili grupiranje velikog broja podataka i izvlačenje relevantnih informacija.

³⁹ RSS (skraćena od *RDF Site Summary*, više poznat kao *Really Simple Syndication* - stvarno jednostavne vijesti) je skup web-formata rabljenih za web-stranice koje se često osvježavaju.

pronalazi entitete u njemu. Osim klasične identifikacije entiteta također vraća i povezane činjenice te događaje skrivene unutar teksta. Alat je besplatan za komercijalnu i nekomercijalnu upotrebu, a za njegovo korištenje potrebno je zatražiti API⁴³ ključ.



Slika 1.29 OpenCalais – konceptualna shema alata⁴¹

pregledavanja, prijenosa i dokumentiranja informacijskog sadržaja. U digitalnom smislu to su strukturirani podatci koji opisuju, objašnjavaju, lociraju ili na neki drugi način omogućavaju lakše upravljanje resursima.

⁴³ Aplikacijsko programsko sučelje (engl. *Application programming interface, API*) ili sučelje za programiranje aplikacija je skup određenih pravila i specifikacija koje programeri slijede tako da se mogu služiti uslugama ili resursima operacijskog sustava ili nekog drugog složenog programa kao standardne biblioteke rutina (funkcija, procedura, metoda), struktura podataka, objekata i protokola.

Topics:
Business Finance 100%

Social Tags:
Business Finance ☆☆☆
Novartis ☆☆☆
Nestlé ☆☆☆
Alcon ☆☆☆

Entities:

- City**
 - London, Greater London, United Kingdom
 - ZURICH
- Company**
 - Alcon Inc
 - Aventis
 - GlaxoSmithKline PLC
 - Jefferies Quarterdeck
 - Kraft Foods Inc.
 - Mead Johnson Nutrition
 - Nestlé S.A.
 - Novartis AG
- Country**
 - Switzerland
 - United States
- Currency**
 - CHF
 - USD
- Industry Term**
 - babyfood
 - consumer healthcare
 - eye care
 - food
 - healthcare
- Person**
 - Ben Hirschler
 - Daniel Vasella
 - Erica Billingham
 - Jeffrey Holford
 - Paul Bulcke

Novartis plans to buy rest of Alcon for \$39 billion

ZURICH (Reuters) - **Switzerland's Novartis plans to get full control of eye care group Alcon** for a total **\$39.3 billion**, to diversify away from prescription drugs, after agreeing to acquire a majority stake from **Nestlé**.

Deals

Novartis said on Monday it would boost its stake in Alcon to 77 percent, exercising an option to buy an extra 52 percent stake from the world's **largest food group Nestlé for \$28.1 billion** after buying an initial 25 percent stake in 2008.

Novartis also plans to buy out the 23 percent stake in **Alcon** held by minority shareholders at an additional cost of **\$11.2 billion**, ending uncertainty as to whether or not it would go for full control.

The fixed exchange ratio proposal, however, of 2.80 **Novartis** shares for each remaining **Alcon** share is less generous to minorities than the price agreed with **Nestlé** of **\$180** per share, since it amounts to **\$153** per share.

Jeffrey Holford, an analyst at **stockbroker Jefferies**, in **London** said **Novartis** would likely end up paying more than **\$153** to buy out the minorities but wanted to keep the cost as low as possible.

"What we are seeing here is the starting point of a negotiation. It think they'll end up paying more for it, but they are trying to not pay more than they are paying for the **Nestlé** stake," **he** said.

Alcon said its **independent director** committee was reviewing **the Novartis offer**.

Novartis Chief Executive Daniel Vasella **told reporters he** was confident that **Alcon** minority shareholders would accept the offer.

Drugmakers like **Novartis**, **GlaxoSmithKline** and **Sanofi-Aventis** are pushing into areas like **consumer healthcare** and generics as they face the biggest loss of patent protection in history.

"The addition of **Alcon** will strategically strengthen our **healthcare** portfolio and our position in eye care, a sector with dynamic growth due to the increasing patient needs of an aging population," said **Vasella**.

Novartis said it expected to complete the deal in the second half of the year, funding it from available cash resources and up to **\$16 billion** of external debt financing.

It will also ask its shareholders to approve the issuance of 98 million new shares to finance the purchase of the **Alcon** minority shares, together with 107 million shares held in treasury.

Novartis said it expects about **\$200 million** of annual pretax cost synergies within three years after closing with the 77 percent stake through shared service agreements, collaborations, joint ventures and other business arrangements.

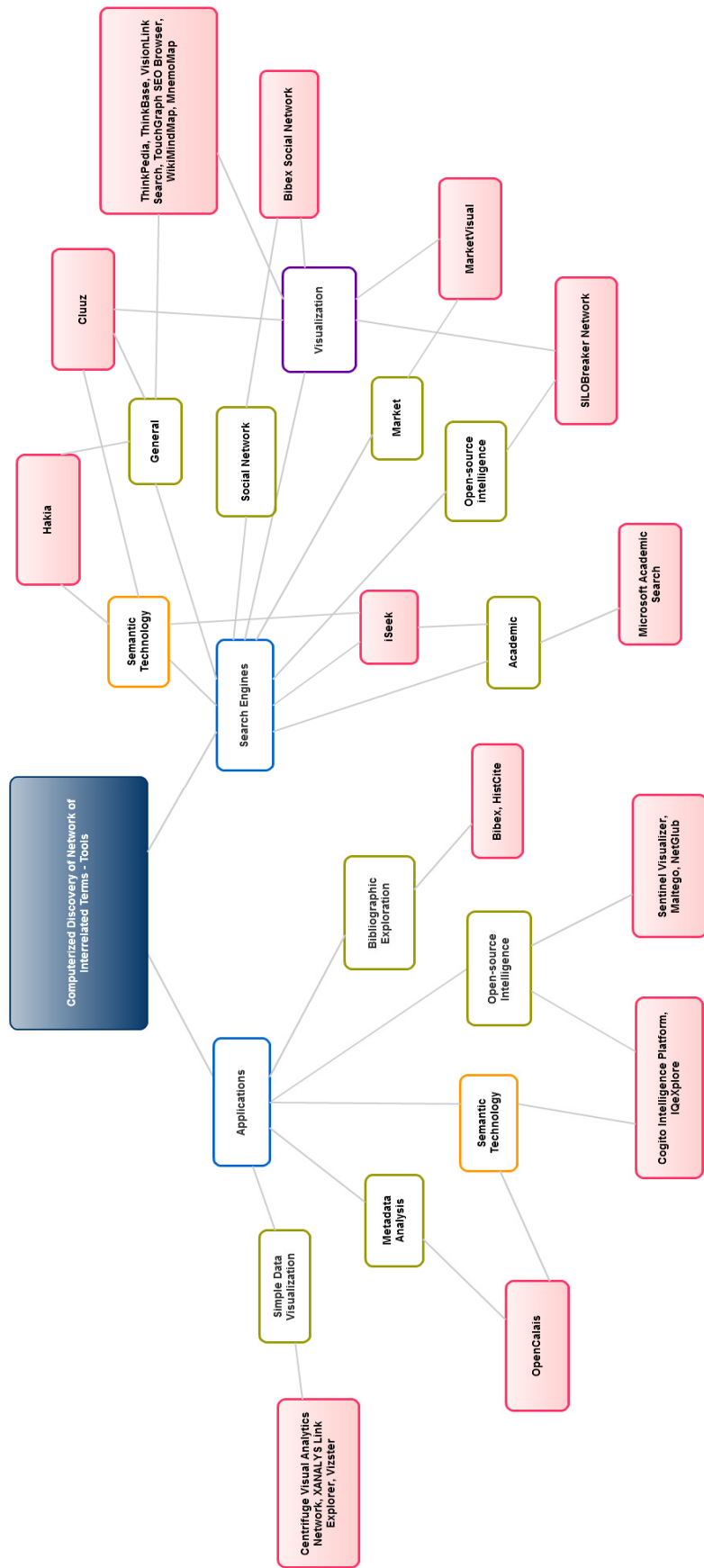
CASH-RICH **NESTLÉ**

Slika 1.30 Open Calais – primjer analize teksta

1.3. Sažetak programskih alata

Osim navedenih alata, u sklopu istraživanja je pronađeno još alata koji nisu opisani u radu zbog toga što se ne mogu koristiti pri akademskom pretraživanju ili nemaju mogućnosti otkrivanja veza među podacima već se koriste samo za programsko ostvarenje vizualizacija. Neki od takvih alata se koriste u području poslovne inteligencije⁴⁴ (kao što je RapidMiner) te pri stvaranju vizualizacija u web-preglednicima (programske knjižnice kao što je Cytoscape Web). Konačno, radi opširnijeg pregleda područja važni alati su prikazani u grafu identificiranih alata prikazanom na slici 1.31.

⁴⁴ Poslovna inteligencija (engl. *Business Intelligence*, *BI*) skup je metodologija i koncepata za prikupljanje, analizu i distribuciju informacija uz pomoć različitih programskih alata. Poslovna inteligencija jedna je od tehnika poslovnog izvještavanja, koja omogućuje pronalaženje informacija potrebnih za lakše i točnije donošenje poslovnih odluka. Neke od metoda poslovne inteligencije uključuju rudarenje podataka (engl. *Data Mining*), skladištenje podataka (engl. *Data Warehousing*) i OLAP mrežnu analitičku obradu podataka.



Slika 1.31 Graf postojećih alata

Legenda pojmova grafa na slici 1.31:

Tools – Programski alati

Applications – Primijenjena programska podrška/aplikacije

Search Engines – Web-tražilice

Open-source Intelligence - Označava prikupljanje „neklasificiranih informacija“, odnosno, informacija s potencijalom mogućeg značaja za obavještajne službe koja je dostupna javnosti.

Bibliographic Exploration – Bibliografsko istraživanje: pronalaženje i pretraživanje autora, publikacija, knjiga u bibliografskim izvorima u kojima se mogu naći informacije o pojedinim zapisima.

Metadata Analysis - Analiza metapodataka, označava i prikupljanje metapodataka iz digitalnih datoteka (npr. tekstualnih). Metapodaci su podaci koji opisuju karakteristike nekog izvora u digitalnom obliku. Korisni su kod pregledavanja, prijenosa i dokumentiranja informacijskog sadržaja. U digitalnom smislu to su strukturirani podaci koji opisuju, objašnjavaju, lociraju ili na neki drugi način omogućavaju lakše upravljanje resursima.

Simple Data Visualization – Grafički prikaz strukturiranih skupova podataka i veza među podacima pohranjenim u određenom datotečnom obliku (bazi podataka, XML-zapisu, MS Excel-datoteci...). Ne omogućava naknadno otkrivanje veza među podacima.

Semantic Technology – Označava upotrebu nekih semantičkih tehnologija. Semantika se kao grana lingvistike bavi proučavanjem značenja jezičnih znakova, dok s druge strane informacijske znanosti nisu usredotočene na značenje informacija. Međutim, područja umjetne inteligencije (engl. *Artificial intelligence*) kao što su obradba prirodnog jezika (engl. *Natural language processing*), strojno učenje (engl. *Machine learning*) i strojno prevođenje (engl. *Machine translation*) zanima obradba prirodnog jezika na svim jezičnim razinama, pa i na semantičkoj. Semantičke tehnologije ne zanima struktura poveznica među različitim mrežnim sjedištima, odnosno dokumentima, nego odnosi među elementima i njihovim svojstvima. Utvrđivanjem odnosa među elementima i njihovim svojstvima s pomoću metapodataka omogućava se strukturiranje nestrukturiranih ili polustrukturiranih podataka na mreži. Semantičke tehnologije uključuju alate za automatsko prepoznavanje tema i koncepata, ekstrakciju značenja i informacija te kategorizaciju podataka.

U nastavku se nalaze tablice svih alata sa navedenim bitnim karakteristikama:

Naziv	Predmet pretrage	Izgled rezultata	Vizualizacija	Mogućnosti	Namjena	Domena	Dostupnost	Jezik
Hakia	Pojmovi ili izrazi	Ispis web-stranica, prikaz rezultata po područjima: službene stranice, biografije, slike, vijesti i intervjui, fan stranice, nagrade, govori, mitovi, polemike, resursi, inovacije, statistike, literatura i filmografija	Nema	Koristi mehanizme kojima pokušava oponašati ljudske kognitivne sposobnosti prepoznavanja onog što je bitno u web-stranicama	Za sve korisnike	Web	Besplatno	Engleski
Microsoft Academic Search	Publikacije i autori	Ispis, prikaz rezultata po područjima	Academic Map, CFP Calendar, Domain Trend, Organization Comparison, Co-author Graph, Co-author Path	Moguće je ograničiti pretragu po web-područjima i tražiti posebno po Author, Conference, Journal, Organization, Year, DOI te koristiti navedene vizualizacije	Pretraživanje akademskih publikacija i autora u akademske svrhe	Web	Besplatno, dobro dokumentirani API-ji	Engleski
Cluuz	Pojmovi: tvrtke/organizacije, telefonski brojevi, e-mail adrese, adrese, domene, datumi, geografija, ostalo	Istaknuti bitni pojmovi i slike, popis važnih web-stranica i prikaz vezanih pojmova u obliku semantičkih grafova	Semantički interaktivni grafovi	Pretraživanje je moguće prilagoditi; moguće je kliknuti na bilo koji entitet te dodatno usredotočiti pretraživanje	Za sve korisnike	Web ili News	Besplatno	Engleski
iSeek	Pojmovi	Popis web-stranica, povezanih pojmova, poveznica na poznate udruge i sveučilišta koja se bave traženom temom te prikaz po sljedećim skupinama: Source, Topics, Subject, Resource Type, Grade Level, Places, People, Organizations, Standards by State	Nema	Svaka skupina rezultata sadrži pojmove sortirane od najvećeg broja podudarnosti u web-stranicama do najmanjeg	Za sve korisnike, pretraživanje u akademske svrhe	Web, Education	Besplatno	Engleski
MnemoMap	Pojmovi	Atomic-Tree, Query List, Tabs	Atomic tree: Ispisuje sinonime, tagove i prijevode. Svaku od pronađenih riječi moguće je postaviti kao centar atomskog stabla ili dodati kao ključnu riječ u pretraživanje	Još u početnom stadiju razvoja, nema sve mogućnosti implementirane	Za sve korisnike	Web, images, digg, del.icio.us, youtube	Besplatno	Engleski, njemački, francuski, finski, švedski, talijanski, estonski, nizozemski, katalonski, norveški, turski
TouchGraph SEO Browser	Pojmovi	Ispis povezanih pojmova, top rezultata pretraživanja, top domena pretraživanja, interaktivni graf - mreža povezanih pojmova	Interaktivni graf	Pretraživanje na grafu je moguće usredotočiti odabirom entiteta na grafu; grafove je moguće preuzeti	Za sve korisnike	Web	Besplatno	Engleski
WikiMindMap	Pojmovi	Ispis rezultata u obliku interaktivne kognitivne mape	Interaktivni graf	Pretraživanje na grafu je moguće usredotočiti odabirom entiteta na grafu; grafove je moguće preuzeti u formatu FreeMind	Za sve korisnike	Wikipedia	Besplatno	Podržava desetak jezika, među njima i engleski
VisionLink Search	Pojmovi	Ispis web-stranica te interaktivnog grafa - mreže povezanih pojmova s entitetima: država, jezik, osoba, organizacija i Wiki-kategorija	Interaktivni graf	Pretraživanje preko entiteta prikazuju se veze među entitetima koje je moguće odabrati, sakriti i filtrirati	Za sve korisnike	Wikipedia	Besplatna demo verzija	Engleski, njemački, francuski
Thinkpedia	Pojmovi	Interaktivni graf - mreža povezanih pojmova, grane sa pojmovima su grupirane po zemljama, tvrtkama, društvenim oznakama, organizacijama, pozicijama, osobama itd.	Interaktivni graf	Moguće je usredotočiti pretraživanje, odabirom entiteta graf se proširuje	Za sve korisnike	Wikipedia	Besplatno	Engleski
Thinkbase	Teme i pojmovi	Interaktivni graf - mreža povezanih pojmova, grane sadrže semantičku oznaku - vezu između središnjeg pojma i pojmova na listovima na grafu	Interaktivni graf	Rezultate je moguće filtrirati po domenama i usredotočiti pretraživanjem, odabirom entiteta graf se proširuje	Za sve korisnike	Wikipedia	Besplatno	Engleski
Silobreaker Network	Entiteti sadržani u člancima objavljenim na domeni silobreaker.com	Interaktivni graf - mreža povezanih pojmova, obuhvaća tvrtke, organizacije, ljude, gradove, ključne riječi i proizvode	Interaktivni graf	Moguće je usredotočiti pretraživanje pomoću grafa te otkriti članke u kojima je opisana veza između entiteta te pretraživati dokumente određenog tipa, ograničiti vrijeme objavljivanja i jezik pretraživanja	Za sve korisnike, može se koristiti u političke, znanstvene ili poslovne svrhe	Silobreaker.com, poddomene: Global Issues, Technology, Science, Business, Energy, Countries	Besplatno	Arapski, nizozemski, engleski, francuski, njemački, španjolski, švedski
MarketVisual	Profesionalne osobe, moguće pretražiti po imeni, tvrtki ili titulu	Profesionalni životopisi, uključuju detaljne položaje, datume i prateću interaktivnu vizualizaciju mreže pojmova	Interaktivni graf	Moguće je odrediti vrstu veze između entiteta te vidjeti podatke vezane uz pojedini entitet	Za tržišne stručnjake	Nepoznato	Besplatna demo verzija, originalna verzija \$19,99 mjesečno	Engleski
Bibex Social Network	Autori (i teme)	Interaktivni graf - mreža povezanih autora	Interaktivni graf	Pronalazi povezane autore, moguće je vremenski organizirati pretragu od-do (u godinama)	Za sve korisnike	Nepoznato	Besplatna demo verzija	Engleski

Tablica 1.1 Svojstva tražilica

Ime	Vrsta	Mogućnosti	Namjena	Dostupnost	Podržane platforme
Maltego	Open-source aplikacija	Pružajući sučelje za prikupljanje informacija za prikaz u jednostavnom i razumljivom obliku, omogućava identificiranje veza među informacijama	Prikupljanje informacija i računalna forenzika	Besplatno	Microsoft Windows, Mac i Linux
NetGlub	Open-source aplikacija	Pružajući sučelje za prikupljanje informacija za prikaz u jednostavnom i razumljivom obliku, omogućava identificiranje veza među informacijama	Prikupljanje informacija i računalna forenzika	Besplatno	Linux, ostalo - nepoznato
Sentinel Visualizer	Aplikacija	Pronalazak skrivenih odnosa među entitetima, prepoznavanje klastera i uzoraka, brzo obavljanje ad-hoc analize, testiranje teorija i scenarija, organiziranje složenih mreža	Analiza društvenih mreža	Izdanja: standard, professional, premium; cijene su \$2,499, \$2,999 i \$3,999 za područje Sjeverne Amerike, nepoznata cijena za internacionalne narudžbe	Microsoft Windows 7, Microsoft Windows Vista, Microsoft Windows XP, Service Pack 2 ili naprednije
Bibex – Bibliographic Exploration Tool	Aplikacija	Izvođenje složenih bibliografskih upita i prikaz rezultata upita kao kombinaciju grafova i tekstova, istraživanje odnosa između autora, ključnih riječi i radova	Bibliografska istraživanja	Besplatna demo verzija – pretraživanje autora i vizualizacija mreže autora	Nepoznato
HistCite	Programska implementacija historografskog algoritma	Vizualizacija rezultata pretraživanja Web of Science	Historiografska istraživanja	Besplatno, potreban pristup Web of Science za cjelokupnu funkcionalnost	Microsoft Windows
Cogito Intelligence Platform	Programski paket	Primjenjuje semantičku analizu podataka te predstavlja podršku za rad u području upravljanja znanjem i prikupljanja podataka uz mogućnost otkrivanja uzoraka u informacijama i međusobnih veza s mogućnošću vizualizacije	Open-source intelligence, prikupljanje informacija	Moguće je zatražiti besplatnu demo verziju, originalnu verziju je potrebno naručiti	Nepoznato
XANALYS Link Explorer	Desktop-aplikacija	Služi za modeliranje podataka kao skupa objekata (ljudi, mjesta, događaja) i veza između njih	Modeliranje veza među podacima	Besplatna verzija dostupna na 30 dana procjene, originalnu verziju je potrebno naručiti	Windows 7, Windows Vista, Windows XP Professional (Service Pack 2 ili 3)
Centrifuge Visual Analytics Network	Web-aplikacija	Omožava otkrivanje uvida u uzorke i odnose skrivene u javnim, cloud i društvenim mrežama i podacima vezanim uz poduzeća	Analiza veza među podacima	Besplatna verzija dostupna na 14 dana procjene, originalnu verziju je potrebno naručiti	Potrebno je imati Adobe Flash Player 10, podržani preglednici su Internet Explorer 7, Chrome 5, Mozilla Firefox 3
IQeXplore	Semantička web-aplikacija	Analiza podataka iz raznih izvora, modeliranje veza između entiteta, semantičko pretraživanje, izgradnja vlastite aplikacije, napredna vizualizacija, stvaranje grafova povezanih pojmova itd.	Data-mining, analiza podataka	Potrebno naručiti	Nepoznato
Cytoscape Web	Open-source projekt	Sadrži komponente/knjižnice koje omogućavaju vizualizaciju mreža te ugradnju unutar HTML-dokumenata	Vizualizacija mreža u web-preglednicima	Besplatno	Potrebno je imati Adobe Flash Player plugin, a radi na zadnjim verzijama preglednika Chrome, Firefox, Safari i Internet Explorer
Vizster	Desktop-aplikacija	Omožava vizualizaciju društvenih mreža; podaci se mogu učitati iz XML-datoteke koja opisuje mrežu ili prilagođene baze podataka MySQL	Vizualizacija društvenih mreža	Besplatno	Potrebno je imati instaliranu Javu
OpenCalais	Web-aplikacija	Stvara bogatu semantiku metapodataka za uneseni sadržaj; analizira dokument i pronalazi entitete u njemu. Osim klasične identifikacije entiteta također vraća i povezane činjenice i događaje skrivene unutar teksta.	Analiza metapodataka	Besplatna verzija za komercijalnu i nekomercijalnu upotrebu; potrebno je zatražiti API ključ	Nepoznato

Tablica 1.2 Svojstva aplikacija

Zaključci istraživanja:

1. U sklopu istraživanja identificirano je više različitih područja tehničkih znanosti koja se bave problematikom računalnog otkrivanja mreže povezanih pojmova. Sukladno tome identificirani su i različiti alati s različitim namjenama.
2. Potencijalno najkorisniji alati su tražilice koje imaju mogućnost izgradnje interaktivnih vizualizacija i napredne aplikacije koje imaju mogućnost otkrivanja veza među entitetima te njihove vizualizacije.
3. Dobra svojstva web-tražilica je to što su besplatne i neovisne o platformama na kojima se izvode jer su vezane samo uz web-preglednike. Međutim, problem je što su njihove domene pretraživanja ili previše općenite (Web) ili previše ograničene

(Wikipedia) i u tom slučaju nije moguće usredotočiti pretraživanje na akademske domene koje su od interesa za ovaj rad. Isto tako, u slučaju tražilica pojmove po kojima se odvijaju pretrage i ispis rezultata nije moguće prilagoditi.

4. Što se aplikacija tiče, jedino Bibex ima namjenu za akademska pretraživanja (pretraživanje akademskih izvora), dok je većina ostalih alata koji imaju potrebne mogućnosti nepogodna za korištenje u tu svrhu. Takve je alate većinom potrebno naručiti i cijena im je vrlo visoka, pogotovo zbog toga što se koriste kao profesionalni alati (nerijetko u svrhe nacionalne sigurnosti i kao vladini alati pri otkrivanju uzoraka među javno dostupnim informacijama) i prenapredni su za korištenje pri svakodnevnom samostalnom istraživanju.
5. Svrha ovog istraživanja bila je pronaći alat koji je moguće prilagoditi potrebama akademskog istraživanja te kao pomoć pri učenju. Također, alat bi trebao biti besplatan ili bi trebalo postojati besplatno, nekomercijalno izdanje alata. Nadalje, trebalo bi biti moguće jednoznačno odrediti izvore informacija (domene pretraživanja). Isto tako, ispis entiteta u grafovima te način na koji se identificiraju i ispisuju veze među entitetima bi se trebali moći prilagoditi.
6. Uzimajući u obzir zahtjeve, alati koji ispunjavaju sve ili gotovo sve zadane uvjete su aplikacije **Maltego**, **NetGlub** te **Sentinel Visualizer**. Međutim, NetGlub je dostupan samo za operativni sustav Linux i izdanje je još u beta verziji, a Sentinel Visualizer je potrebno naručiti. **Maltego**, s druge strane, ima *community* izdanje koje je besplatno i kompatibilno s više operativnih sustava. Iako je namjena alata Maltego prvenstveno za računalnu forenziku, detaljnijim istraživanjem utvrđeno je da se alat može prilagoditi za razne svrhe pa tako i za akademska istraživanja. U narednim poglavljima ovog rada stoga su pobliže istražena svojstva i mogućnosti upravo ovog alata.

2. Maltego

Kao što je navedeno ranije u radu, Maltego je aplikacija otvorenog kôda koja pruža sučelje za prikupljanje informacija te grafički prikaz u jednostavnom i razumljivom obliku. Originalna namjena alata je korištenje arhitekture klijent/poslužitelj u svrhu prikupljanja podataka za određivanje odnosa u području računalne forenzike. Alat ima mogućnost grafičkog prikaza međusobnih veza među podacima kao što su:

- Osobe
- Grupe ljudi (socijalne mreže)
- Tvrtke
- Organizacije
- Web-stranice
- Internet infrastruktura (poput domena, DNS imena, IP adresa i sl.)
- Izrazi
- Članstva
- Dokumenti i spisi

Navedene mogućnosti već su ugrađene u Maltego od strane razvijatelja aplikacije i dostupne su svim korisnicima. Međutim, Maltego omogućava korisniku da na sličan način samostalno ostvari nove funkcionalnosti. U nastavku rada opisane su osnove koje je potrebno znati o Maltego te pobliže objašnjeni načini na koje se to može učiniti.



Slika 2.1 Maltego Community Edition 3.1.1 logo

2.1. Entiteti i transformacije

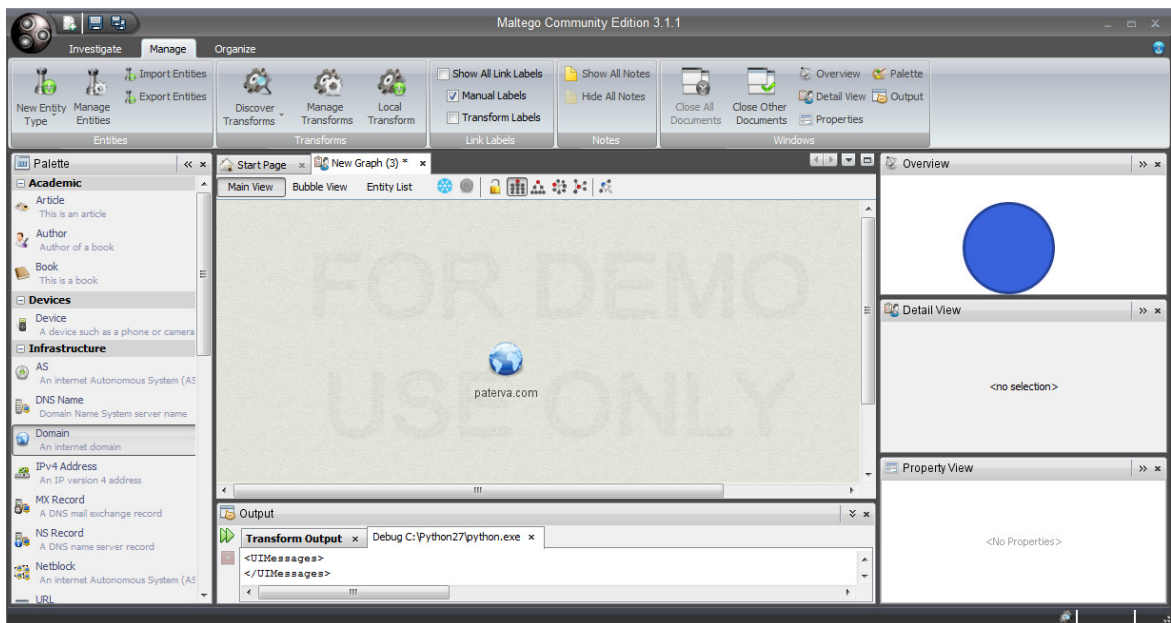
Osnovni pojmovi vezani uz Maltego su **entiteti** (engl. *Entities*) i **transformacije** (engl. *Transforms*).

Općenito, entiteti su grafičke reprezentacije pojmova pri čemu svaki entitet ima svoj naziv te dodatna svojstva. Maltego ima mogućnost iscrtavanja veza među povezanim entitetima, a za svaku vrstu entiteta moguće je napisati transformacije.

Transformacije su programski moduli koji se izvode nad entitetima i kao ulazni parametar primaju podatak o entitetu na kojemu se izvode (naziv i svojstva). Smisao transformacije je korištenje tog podatka (npr. za pronalazak povezanih pojmova po nekom kriteriju) i preslikavanje pronađenih pojmova u nove entitete koji se vraćaju kao izlazni parametar u Maltego. Pri pokretanju transformacije Maltego iscrtava nove entitete u grafu zajedno s vezama prema početnom entitetu.

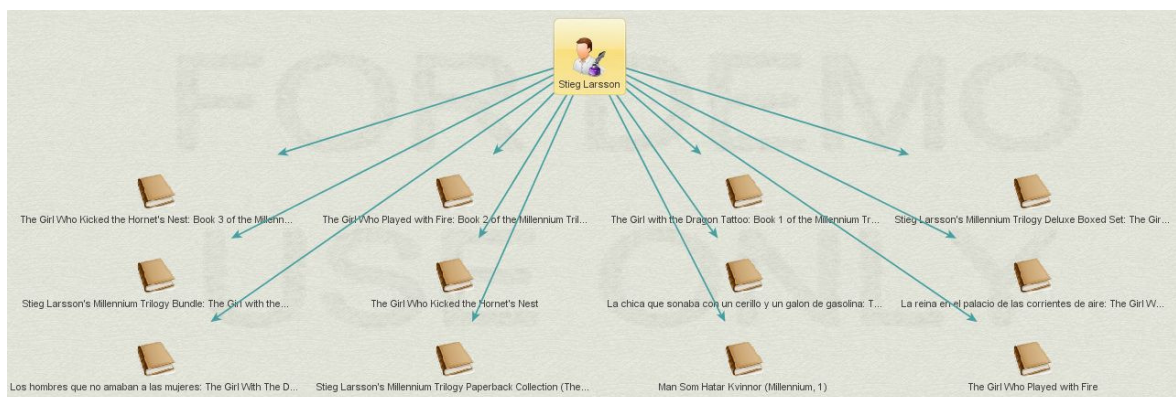
Osim postojećih entiteta moguće je stvoriti proizvoljan broj novih entiteta te im dodati nova svojstva. Također, moguće je dodati proizvoljan broj transformacija za svaku od vrsta entiteta.

Izgled sučelja može se vidjeti na slici. Rad u Maltegu započinje stvaranjem novog grafa te dodavanjem novog entiteta na graf po *drag and drop* principu iz palete na lijevoj strani sučelja. Nakon toga je moguće desnim klikom miša na entitet pokrenuti transformaciju/e. Maltego, osim pokretanja transformacija i stvaranja grafova nudi i mnogo ostalih opcija za prilagodbu i uređivanje grafova te pretraživanje na Googlu i Wikipediji i otvaranje URL-ova. Sve mogućnosti Maltega opisane su u službenim korisničkim uputama.



Slika 2.1 Maltego – izgled sučelja

Kao primjer dodavanja nove funkcionalnosti u Maltego napisana je nova transformacija koja se izvodi nad entitetom vrste *autor*. Programski kod transformacije napisan je tako da se pokretanjem transformacije izvršava pretraga knjiga na Amazonu za tog autora, a rezultati pretrage preslikavaju u nove entitete vrste *knjiga*. Nakon izvođenja transformacije Maltego iscrtava nove entitete zajedno s vezama prema početnom entitetu (slika 2.2).



Slika 2.2 Početni entitet (autor Stieg Larsson) i vezani entiteti (knjige koje transformacija pronalazi na Amazonu)

2.2. Transform Distribution Server (TDS)

Paterva TDS⁴⁵ (skraćena od *Transform Distribution Server*) je poslužitelj čija je svrha dijeljenje, upravljanje i implementacija lokalnih/prilagođenih transformacija diljem Weba. TDS je ustvari web-aplikacija koja omogućava kontrolu *seedova* i postavki transformacija. Zahvaljujući *seedovima* ažuriranje transformacija može se odvijati samo na lokalnom poslužitelju (slika 2.3).

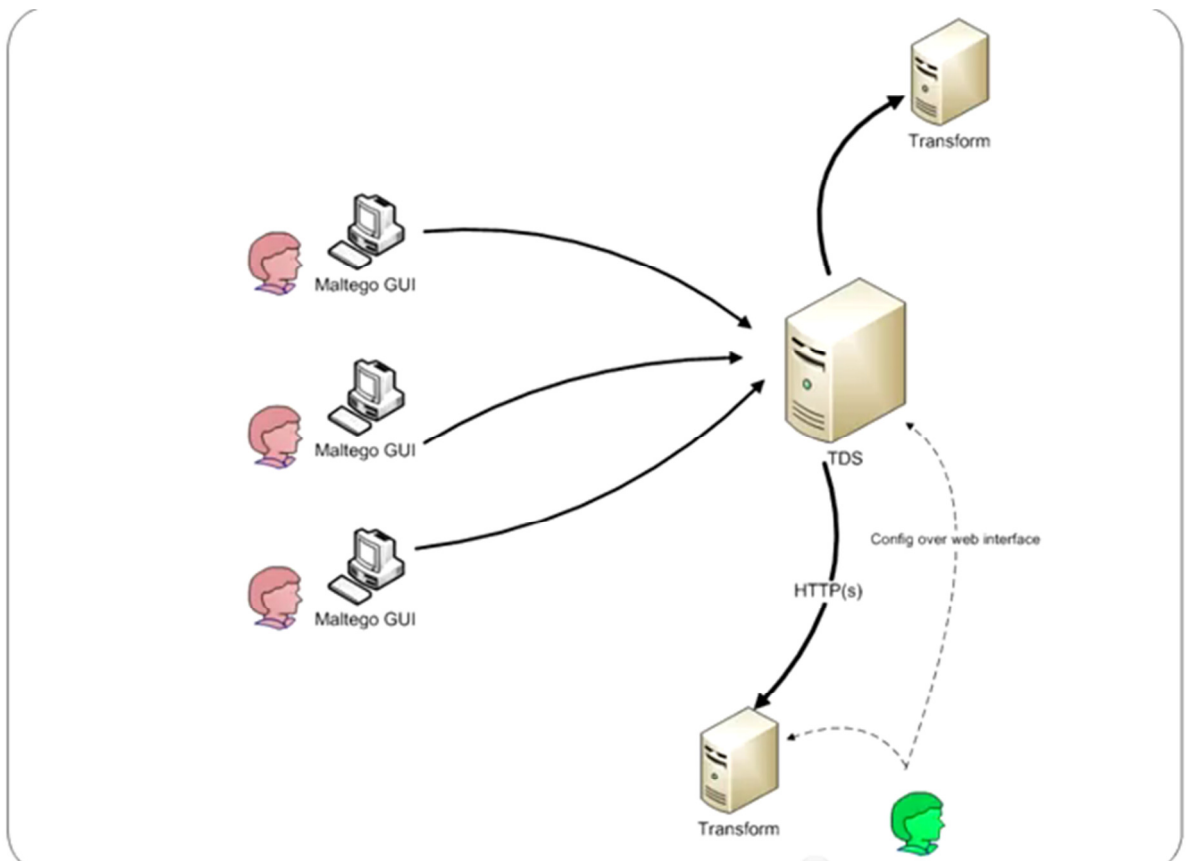
Transformacije je moguće pisati lokalno (skripta se može nalaziti bilo gdje na računalu) ili dodati u vlastiti *TransformSeed* na Paterva TDS poslužitelju. Da bi to bilo moguće, transformacija se mora nalaziti na web-poslužitelju u *cgi-bin* direktoriju. Dodavanje transformacije u vlastiti *seed* omogućava drugim korisnicima korištenje transformacije iz Maltego GUI-a⁴⁶ bez prethodne konfiguracije alata, međutim, to nije nužno za lokalno testiranje.

Zašto se koristi TDS [7]:

- Lokalne transformacije nije moguće uvijek osposobiti:
 - Potrebno je imati sve programske module i knjižnice na svakom od računala
 - Nije moguće udaljeno ažuriranje transformacija
 - Potrebno je dijeliti programski kod (ili osjetljive informacije)
 - Potrebno je kopirati datoteke za konfiguraciju

⁴⁵ <https://cetas.paterva.com/TDS/>

⁴⁶ GUI (engl. *Graphical user interface*) - grafičko korisničko sučelje je način interakcije čovjeka s računalom kroz manipulaciju grafičkim elementima i dodacima uz pomoć tekstovnih poruka i obavijesti. http://hr.wikipedia.org/wiki/Grafi%C4%8Dko_korisni%C4%8Dko_su%C4%8Delje



Slika 2.3 Shema pristupa TDS-u [7]

2.3. Dostupnost i podržane platforme

Maltego postoji u dva izdanja: *community* i *commercial* izdanju.

Maltego *community* izdanje je besplatno, a ograničenja su sljedeća:

- Pojedina transformacija može vratiti maksimalno 12 entiteta
- Nije moguće zalijepiti više od 50 entiteta odjednom
- Izvođenje transformacija je nešto sporije
- Nije moguće napraviti uvoz iz tablica, napraviti izvještaj ili ispisati sadržaj na pisaču
- Na pozadini se vidi natpis „for demo use only“

Maltego *commercial* izdanje je moguće platiti, a plaćanje se odvija svake godine. Cijena za prvu godinu korištenja je \$650 USD, a cijena za naredne godine je \$320 (slika).

Product	Year 1	Year (1+n)
Maltego	\$650 USD	\$320 USD
Maltego CaseFile	\$200 USD	\$200 USD

Slika 2.4 Maltego Commercial Edition – cijene [7]

Plaćanje *commercial* izdanja Maltega podrazumijeva:

1. Licencni ključ koji vrijedi jednu godinu na jednom poslužitelju
2. Neograničen pristup svim transformacijama na komercijalnom poslužitelju
3. E-mail podršku
4. Integriranu dokumentaciju
5. Slobodno korištenje svih vanjskih transformacija i njihovih poslužitelja

Što se podržanih platformi tiče, Maltego je kompatibilan sa platformama Microsoft Windows, Mac i Linux, a zahtijeva Java Runtime Environment 1.6 iz Oracla (prijašnjeg Sun Microsystems).

3. Pisanje transformacija za Maltego

Cilj ovog rada je omogućiti pretraživanje knjiga, članaka, autora itd. na nekima od važnijih izvora na Webu te stvaranje mreže povezanih pojmova. Pošto je pokazano da Maltego ima mogućnost stvaranja mreže pojmova uz pomoć entiteta i transformacija, nastala je ideja da bi se Maltego mogao iskoristiti kao alat za učenje. Unošenjem jednog pojma (entiteta) moglo bi biti moguće pronaći udruge, knjige, časopise, predmete, sveučilišta, autore koji su vezani uz pojam i ispisati ih u Maltegu te na taj način stvarati mreže pojmova. Da bi se to ostvarilo, trebalo bi postojati više transformacija (kao u primjeru za pretraživanje Amazona) koje pretražuju razne baze podataka.

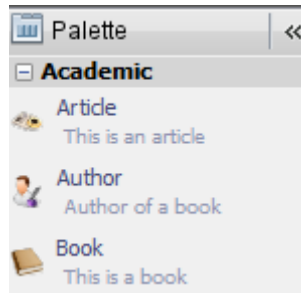
Pri tome se uzimaju u obzir svi pojmovi koje je potrebno istražiti da bi se obuhvatila veća slika nekog područja pri istraživanju. Ugrubo to bi mogli biti sljedeći pojmovi:

- Knjige
- Autori
- Članci
- Projekti
- Sveučilišta
- Udruge
- Video-materijali
- Itd.

Pošto je moguće napisati transformacije za pretraživanje bilo kojeg izvora na Webu, bilo bi idealno kada bi se mogla ostvariti pretraga svih prethodno navedenih pojmova na što više relevantnijih izvora. Da bi se ta metoda pokazala izvedivom, načinjene su transformacije koje omogućavaju pretraživanje knjiga, članaka i vezanih autora na nekim područjima na Webu. U nastavku je pobliže opisan način na koji je to izvedeno.

3.1. Entiteti

U Maltegu su stvoreni novi entiteti *Article*, *Author* i *Book* i dodani u set koji je nazvan *Academic* (slika 3.1). Transformacije koje su opisane u nastavku koriste i ugrađeni entitet *Phrase*.



Slika 3.1 Set *Academic* i entiteti



Slika 3.2 Entitet *Book*

Book	Recreative Minds: Imagination in Philosophy and Psychology
Full Title	Recreative Minds: Imagination in Philosophy and Psychology
Publisher	Oxford University Press, USA
Authors	Gregory Currie, Ian Ravenscroft
ISBN	9780198238096
Release Date	None
Price and Currency	50.0 USD

Slika 3.3 Dodatna svojstva entiteta *Book*

3.1.1. Pregled entiteta

Pregled stvorenih entiteta te njihova svojstva moguće je vidjeti u tablici 3.1:

Naziv entiteta	Dodatna svojstva
Article	Full Title URL
Author	First Name Last Name
Book	Full Title Publisher Authors ISBN Release Date Price and Currency

Tablica 3.1 Svojstva entiteta

3.2. Transformacije

Svaka je transformacija definirana svojim nazivom, programskim jezikom u kojem je napisana, pozivajućim modulom te vrstom entiteta na kojoj se izvodi. Pojedina se transformacija može izvoditi na jednoj vrsti entiteta, a može vratiti nula ili više entiteta različitih vrsta. Transformacije mogu biti pisane u proizvoljnom programskom jeziku (najčešće Perlu ili Pythonu).

Jedini uvjeti pri pisanju koda transformacije odnose se na komunikaciju pozivajućeg modula s Maltegom koja se odvija putem standardnog ulaza/izlaza. Pri pokretanju transformacije, Maltego ispisuje podatke o početnom entitetu na standardni izlaz u propisanom XML-formatu te ih modul mora moći pročitati i parsirati. Nakon izvođenja koda transformacije, modul treba ispisati podatke o novostvorenim entitetima u određenom XML-formatu da bi ih Maltego mogao prepoznati i iscrutati. Također, Maltego dozvoljava ispis samo ASCII-znakova i svi podaci koji se vraćaju kao ispis Maltegu trebaju biti normalizirani i kodirani u ASCII⁴⁷.

Za pisanje transformacija odabran je skriptni programski jezik Python zbog postojećih pomoćnih modula za Maltego, jednostavnosti pretraživanja web-sjedišta te općenito mnoštva korisnih pomoćnih modula na Webu. Instalirano je izdanje Pythona je 2.7.2 [13] te razvojno okruženje Eclipse sa pluginom za Python PyDev [14].

Također, kao pomoć pri pisanju korišteni su sljedeći dokumenti i forumi:

- Specifikacije lokalnih transformacija [8]
- Dokument o opcijama integracije s Maltegom [10]
- Službeni forum Maltega [12]

Napisane su transformacije koje pretražuju sljedeća područja: Amazon⁴⁸, Library of Congress⁴⁹, knjižnični sustav Fakulteta elektrotehnike i računarstva⁵⁰ (FER Kohu⁵¹),

⁴⁷ ASCII (skraćenica od *American Standard Code for Information Interchange*, još i kao Američki standardni znakovnik za razmjenu informacija, Američki standardni znakovnik za razmjenu obavijesti) način je kodiranja znakova temeljen na engleskoj abecedi.

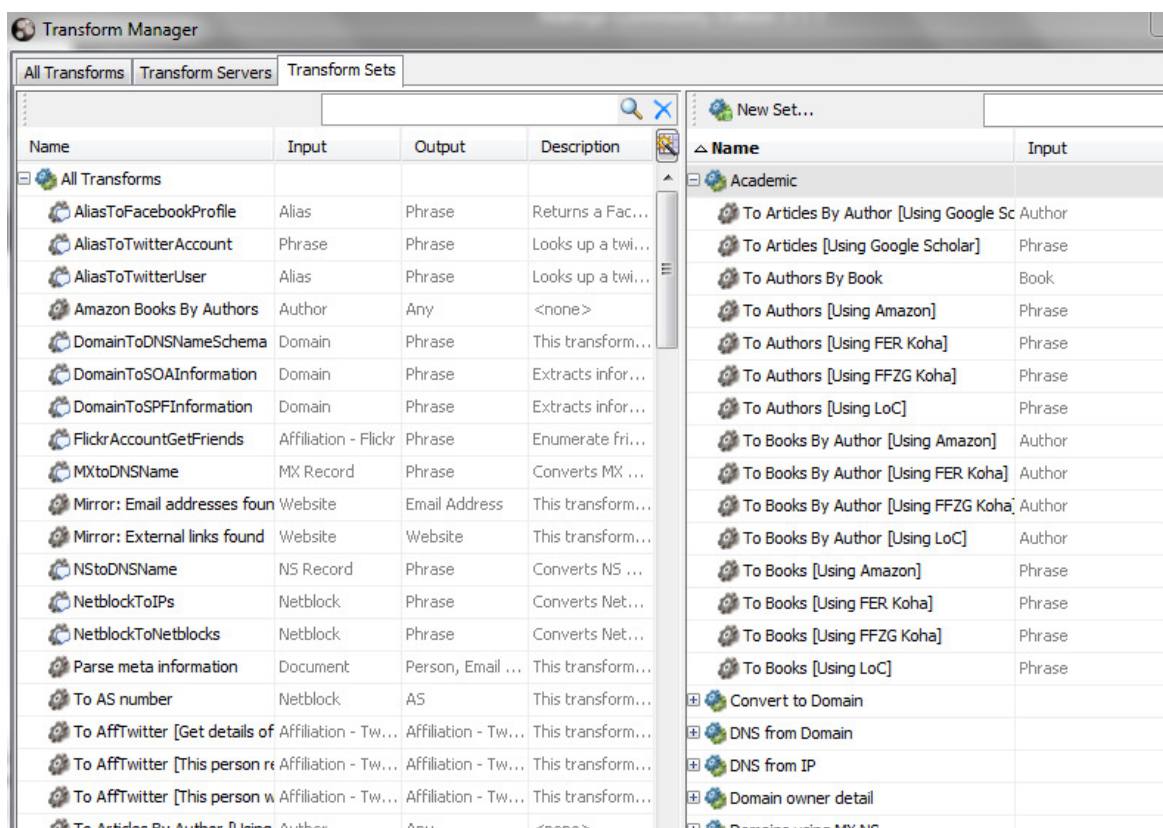
⁴⁸ <http://www.amazon.com/>

⁴⁹ <http://www.loc.gov/index.html>

⁵⁰ <http://lib.fer.hr/>

⁵¹ <http://www.koha.org/>

knjižnični sustav Filozofskog fakulteta u Zagrebu⁵² (FFZG Kohu⁴⁷) za knjige te Google Scholar⁵³ za članke. Također, dodana je općenita transformacija koja ispisiuje sve autore određene knjige. Sve transformacije nalaze se u setu *Academic* (slika).

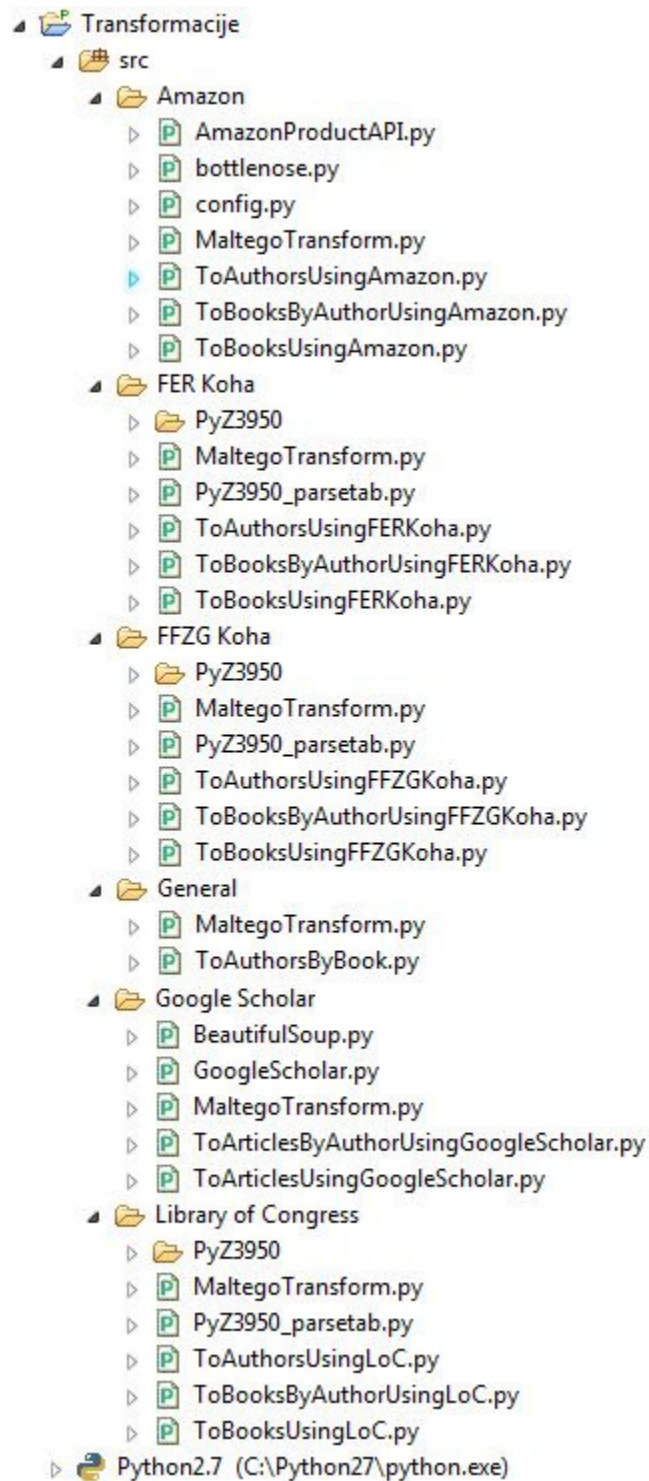


Slika 3.4 Maltego – Transform Manager

Svaka transformacija ima definiran pozivajući modul koji se izvršava pokretanjem transformacije. Organizacija programskih modula u razvojnom okruženju može se vidjeti na slici:

⁵² <http://koha.ffzg.unizg.hr/>

⁵³ <http://scholar.google.hr/>



Slika 3.5 Organizacija modula u razvojnom okruženju PyDev Eclipse

3.2.1. Amazon

Transformacije koje pretražuju **Amazon** pristupaju podacima putem njegovog novog API-ja od studenog 2011. godine. Python Simple Product Amazon API je modul u Pythonu koji funkcionira za novi API Amazona⁵⁴.

Pri pisanju korištena je dokumentacija Amazon API-ja, a za pristup Amazonu bilo je potrebno registrirati se na njihovoj stranici i kao Amazon Associate te pribaviti tri koda za pristup (Access Key ID, Secret Access Key, Unique Associates ID). Kodovi su navedeni u *config.py* i koristi ih modul *AmazonProductAPI.py*.

Transformacije:

- To Books [Using Amazon]: *ToBooksUsingAmazon.py*
- To Authors [Using Amazon]: *ToAuthorsUsingAmazon.py*
- To Books By Author [Using Amazon]: *ToBooksByAuthorUsing Amazon.py*

3.2.2. Library of Congress, knjižnični sustavi FER-a i FFZG-a

Za pristup **Library of Congress** te **knjižničnim sustavima FER-a i FFZG-a** bilo je potrebno programski uspostaviti Z39.50⁵⁵ klijenta. Instalirane su pomoćne knjižnice PyZ3950 (implementacija Z39.50 klijenta u Pythonu) te PLY (Python Lex-Yacc). Za spajanje na Z39.50 poslužitelj od Library of Congress pronađeni su potrebni podaci na Webu, a za spajanje na poslužitelje FER i FFZG Kohe zatraženi su potrebni podaci od ovlaštenih osoba. Bibliografski zapisi na Z39.50 poslužitelju nalaze se u formatu MARC21, a parsiranje zapisa ostvareno je uz pomoć dokumentacije MARC21 zapisa na domeni Library of Congress.

Transformacije (Library of Congress):

- To Books [Using LoC]: *ToBooksUsingLoC.py*
- To Authors [Using LoC]: *ToAuthorsUsingLoC.py*

⁵⁴ <http://webservices.amazon.com/AWSECommerceService/AWSECommerceService.wsdl>

⁵⁵ Z39.50 je klijentsko-poslužiteljski protokol za traženje i dobivanje informacija iz udaljenih računalnih baza podataka. Pokriven je standardima ANSI / NISO Z39.50 i ISO 23950. Library of Congress zaslužan je za održavanje Z39.50 standarda.

- To Books By Author [Using LoC]: *ToBooksByAuthorUsingLoC.py*

Transformacije (FER Koha):

- To Books [Using FER Koha]: *ToBooksUsingFERKoha.py*
- To Authors [Using FER Koha]: *ToAuthorsUsingFERKoha.py*
- To Books By Author [Using FER Koha]: *ToBooksByAuthorUsingFERKoha.py*

Transformacije (FFZG Koha):

- To Books [Using FFZG Koha]: *ToBooksUsingFFZGKoha.py*
- To Authors [Using FFZG Koha]: *ToBooksUsingFFZGKoha.py*
- To Books By Author [Using FFZG Koha]: *ToBooksUsingFFZGKoha.py*

3.2.3. Google Scholar

Pošto **Google Scholar** nema definiran API, za pristup je korišten improvizirani HTML-parser za pretragu članaka po ključnim riječima te pretragu članaka po autoru.

Transformacije:

- To Articles [Using Google Scholar]: *ToArticlesUsingGoogleScholar.py*
- To Articles By Author [Using Google Scholar]:
ToArticlesByAuthorUsingGoogleScholar.py

3.2.4. Općenito

Napisana je i **općenita** transformacija koja ispisuje navedene autore knjige pri čemu su autori navedeni u entitetu *Book* u dodatnom svojstvu *Authors*. Transformacija je važna zbog načina na koji se u Maltegu iscrtavaju veze među entitetima.

Transformacija:

- To Authors By Book: *ToAuthorsByBook.py*

3.2.5. Pregled transformacija

Sve transformacije se sistematizirane u tablici 3.2. Naziv transformacije odgovara nazivu u Maltegu, ulazni entiteti označavaju vrstu entiteta na kojoj se izvršava transformacija, a izlazni entiteti označavaju vrstu entiteta koje transformacija vraća u Maltego.

Naziv transformacije	Područje pretraživanja	Način pristupa	Ulazni entitet	Izlazni entitet(i)	Opis
To Books [Using Amazon]	Amazon	Amazon API	Phrase	Books	Pronalazi knjige vezane uz određeni izraz
To Books [Using LoC]	Library of Congress	Z39.50			
To Books [Using FER Koha]	FER Koha				
To Books [Using FFZG Koha]	FFZG Koha				
To Authors [Using Amazon]	Amazon	Amazon API	Phrase	Authors	Pronalazi autore vezane uz određen izraz
To Authors [Using LoC]	Library of Congress	Z39.50			
To Authors [Using FER Koha]	FER Koha				
To Authors [Using FFZG Koha]	FFZG Koha				
To Books By Author [Using Amazon]	Amazon	Amazon API	Author	Books	Pronalazi knjige vezane uz određenog autora
To Books By Author [Using LoC]	Library of Congress	Z39.50			

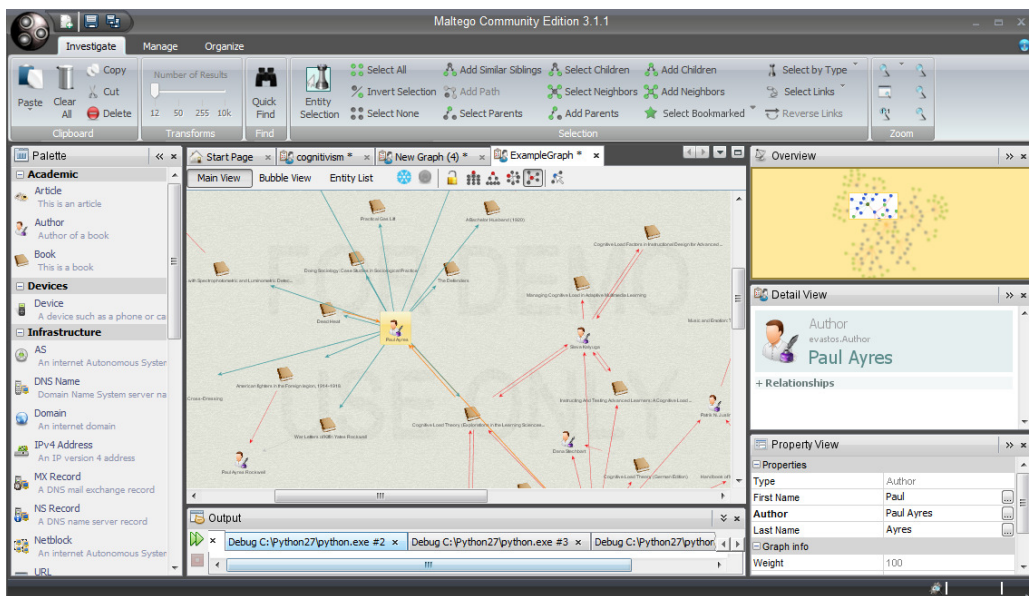
To Books By Author [Using FER Koha]	FER Koha				
To Books By Author [Using FFZG Koha]	FFZG Koha				
To Articles [Using Google Scholar]	Google Scholar	HTML- parser	Phrase	Articles	Pronalazi članke vezane uz određeni izraz
To Articles By Author [Using Google Scholar]			Author	Articles	Pronalazi članke vezane uz određenog autora
To Authors By Book	-	-	Book	Authors	Ispisuje autore određene knjige

Tablica 3.2 Svojstva transformacija

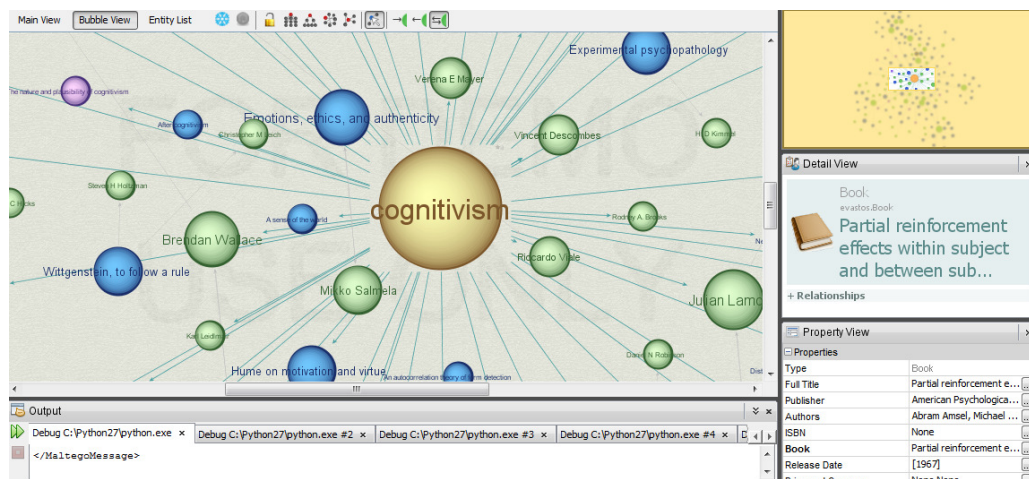
Pisanjem ovih transformacija pokazano je da je pomoću Maltega zaista moguće ostvariti sustav naprednog vizualnog pretraživanja. U nastavku su prikazani neki primjeri izvođenja transformacija i stvaranja bogatih grafova (slike 3.6. – 3.10). Međutim, potrebno je ukazati na sljedeće stvari:

- Iako je ovime obrađeno više izvora informacija za knjige te jedan izvor za akademske članke, za kvalitetnu obradu bilo kojeg područja pretraživanja potrebno je uložiti napor pri stvaranju novih transformacija koje pretražuju ostale izvore na Webu.
- Najznačajniji problemi pri pisanju transformacija vezani su uz parsiranje podataka preuzetih sa raznih izvora na Webu. Da bi Maltego mogao isti podatak preuzet sa više različitih web-sjedišta prepoznati kao jedan te isti pojam, taj se podatak mora preslikati u isti entitet. Stoga je važno da svi podaci koji se vraćaju u Maltego budu u istom formatu. Međutim, zbog krivog ili nepostojanog formatiranja podataka na izvoru pretraživanja to nije uvijek moguće ostvariti.

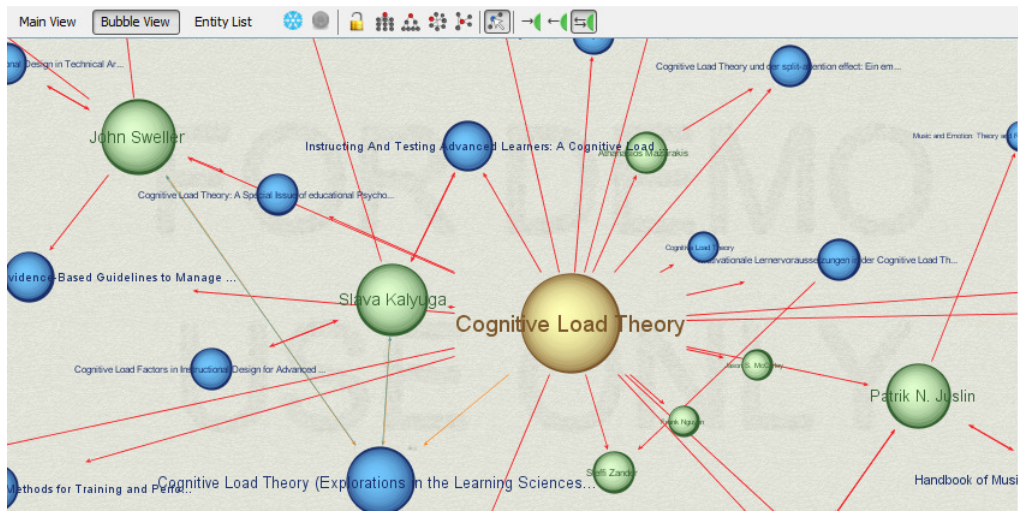
- Također, nije moguće garantirati da će transformacije vratiti u potpunosti ispravne podatke jer ispravnost podataka ne ovisi o njima nego o web-sjedištima koja se koriste za pretraživanje.
- U konačnici, potrebno je uzeti u obzir da ovakvo pretraživanje služi korisniku samo kao pomoć te svojevrsna referenca pri daljnjem istraživanju i u tom slučaju ovakav način pretraživanja može biti vrlo koristan.



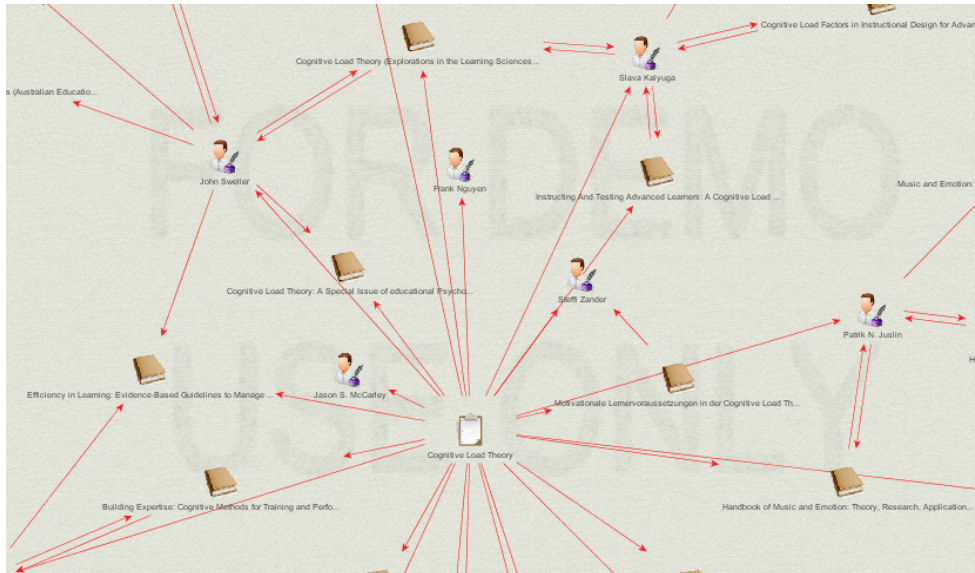
Slika 3.6 Izgled sučelja i primjer grafa



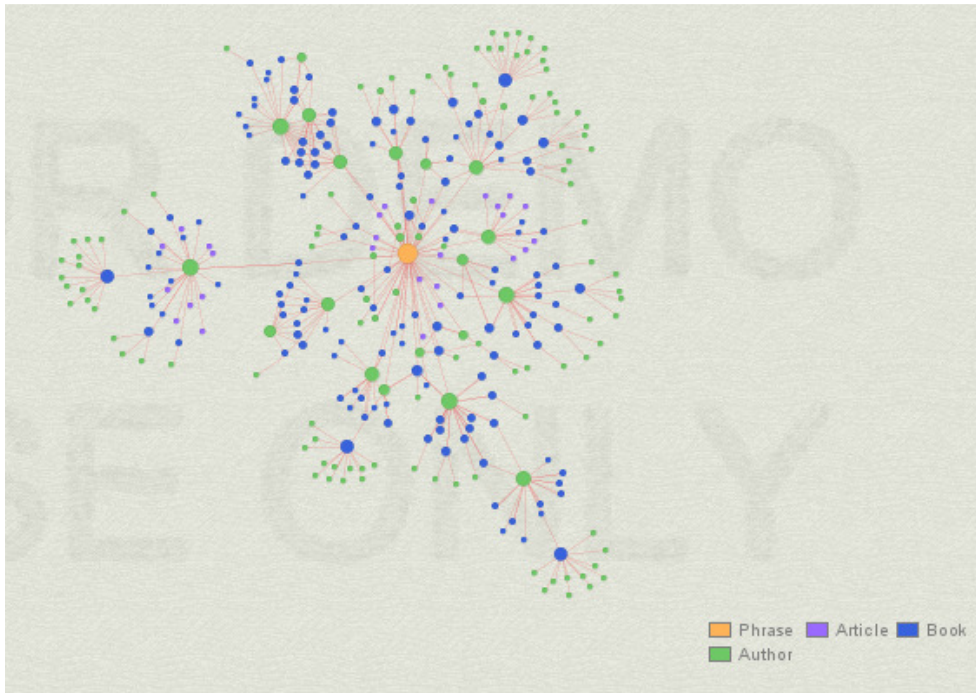
Slika 3.7 Primjer grafa za pojam „cognitivism“ i dio sučelja



Slika 3.8 Primjer grafa za „Cognitive Load Theory“ - Bubble View



Slika 3.9 Primjer grafa za „Cognitive Load Theory“ – Main View



Slika 3.10 Primjer grafa u Maltegu – cijeli graf

4. Upute za pisanje transformacija

Prilikom rada na novim transformacijama potrebno je obratiti pažnju na sljedeće:

1. Potrebno je odabrati pojam koji se želi obraditi (članci, sveučilišta, projekti, video zapisi...)
2. Istražiti sustave i tražilice na Webu pomoću kojih je moguće pretražiti dotični pojam te odabrati one koji su najrelevantniji za pretraživanje u tom području
3. Otkriti na koji način je moguće pristupiti sustavu:
 - a. Može biti putem API-ja (ako postoji)
 - b. Pomoću protokola Z39.50 (provjeriti koji od sustava imaju postavljen Z39.50 poslužitelj)
 - c. Pomoću improviziranih modula koji pretražuju web-sustave parsiranjem HTML-a⁵⁶ ili putem HTTP-a⁵⁷ (nije najbolje rješenje)
 - d. Otkriti neki drugi način
4. U skladu sa prethodnim odabrati programski jezik u kojem će biti pisane transformacije te:
 - a. Vidjeti postoji li već pomoćni modul za Maltego u tom programskom jeziku (na službenom forumu Maltega)
 - b. Ako ne postoji, napisati novi modul po uzoru na postojeće
5. U Maltegu stvoriti entitet koji odgovara odabranom pojmu te odrediti njegova dodatna svojstva
6. Vidjeti na koji način je novi entitet povezan sa prethodno napravljenim entitetima (npr. svaka knjiga ima autore) te može li pretraživanje web-sjedišta otkriti neke nove korisne entitete

⁵⁶ HTML (skraćenica od *HyperText Markup Language*) je prezentacijski jezik za izradu web-stranica. Hipertekst dokument stvara se pomoću HTML-jezika.

⁵⁷ HTTP (skraćenica od *HyperText Transfer Protocol*) je glavna i najčešća metoda prijenosa informacija na Webu. Osnovna namjena ovog protokola je omogućavanje objavljivanja i prezentacije HTML-dokumenata, tj. web-stranica.

7. Otkriti sve odnose među entitetima te za svaki od njih napisati odgovarajuću transformaciju (npr. odnosi autor-knjige, ključna riječ-knjige, ključna riječ-autori, knjiga-autori)
8. Prije samog pisanja transformacija korisno je proučiti sljedeće dokumente:
 - a. Video-zapise u kojima je objašnjeno pisanje transformacija [7]
 - b. Specifikacije lokalnih transformacija [8]
 - c. Upute za transformacije [9]
 - d. Dokument o opcijama integracije s Maltegom [10]
9. Pri pisanju transformacija paziti na konzistentnost pri parsiranju zapisa koji su prikupljeni sa istog sustava na Webu
10. Normalizirati ne-ASCII znakove te kodirati u ASCII sve podatke koje transformacija vraća Maltegu
11. U slučaju da je potrebno ostalim korisnicima osigurati jednostavno korištenje tih transformacija, postaviti transformacije na lokalni poslužitelj te ih potom dodati u vlastiti *TransformSeed* na Paterva TDS-u

Zaključak

Većina sustava pretraživanja na Webu nudi linearan ispis rezultata. Pri učenju i proučavanju nekog područja korisnik mora proći veliki broj rezultata da bi dobio krupnu grubu sliku nekog područja te razlučio važne pojmove od nevažnih. Stoga bi bilo korisno kada bi postojao sustav koji bi obavljao taj posao za korisnika. U zadnje vrijeme javlja se trend poboljšavanja tražilica te razvoja novih vizualnih tražilica sa mogućnosti ispisa rezultata u obliku interaktivnih grafova. Međutim, bilo bi idealno kada bi postojao sustav koji bi se mogao u potpunosti prilagoditi zahtjevima ovog rada.

Istraživanje je pokazalo da aplikacija Maltego u tom slučaju zadovoljava najviše kriterija. U Maltego je moguće implementirati neograničen broj načina pretraživanja te definirati načine stvaranja veza među entitetima na grafu. Osim toga, Maltego posjeduje i mnoštvo opcija za uređivanje grafova te razne načine prikaza tih grafova. Stoga je pokazan način na koji se Maltego može iskoristiti kao koristan alat za stvaranje mreže povezanih pojmova uz visoku interaktivnost s korisnikom.

Kao i svaki, i ovakav sustav ima svoje nesavršenosti, međutim, uzimajući u obzir da takav način pretraživanja služi samo kao pomoć i svojevrsna referenca korisnicima, sustav se pokazao prilično korisnim i jednostavnim za korištenje. Također, pošto u sklopu ovog rada nisu obrađena sva moguća područja pretraživanja, u radu su priložene smjernice za budući rad na unaprjeđivanju sustava pretraživanja u Maltegu.

Literatura

- [1] MATEŠIĆ, M., Semantički Web, <http://www.mensa.hr/glavna/misli-21-stoljeca/652-semanticki-web>, *SEMANTIČKI WEB*, 16.6.2012.
- [2] SPRYLOGICS INTERNATIONAL INC., Cluuz, <http://www.cluuz.com/About.aspx>, *Cluuz*, 16.6.2012.
- [3] SUMNER, C., Social Networking Special OPS: Extending data visualization for faster pwnage, 19.8.2010., <http://www.defcon.org/images/defcon-18/dc-18-presentations/Suggmeister/DEFCON-18-Suggmeister-Social-Net-Special-Ops-WP.pdf>, *Social Networking Special OPS: Extending data visualization for faster pwnage*, 16.6.2012.
- [4] Open-source intelligence, 29.5.2012., http://en.wikipedia.org/wiki/Open_source_intelligence, *Open-source intelligence*, 16.6.2012.
- [5] KDNUGGETS, Social network analysis, Link Analysis, and Visualization, <http://www.kdnuggets.com/software/social-network-analysis.html>, *Social network analysis, Link Analysis, and Visualization*, 16.6.2012.
- [6] KDNUGGETS, Visualization and Data Mining Software, <http://www.kdnuggets.com/software/visualization.html>, *Visualization and Data Mining Software*, 16.6.2012.
- [7] PATERVA, Maltego, <http://www.paterva.com/web5/>, *Maltego*, 16.6.2012.
- [8] PATERVA, Local Transforms Specification, 9.1.2009., <http://www.paterva.com/web5/documentation/localTransforms-SpecIII.pdf>, *Local Transforms Specification*, 16.6.2012.
- [9] PATERVA, Maltego transforms: A reference guide, 17.1.2011., <http://www.paterva.com/malv3/303/M3GuideTransforms.pdf>, *Maltego transforms: A reference guide*, 16.6.2012.
- [10] PATERVA, Integration options with Maltego v3, 12.8.2010., <http://www.paterva.com/web5/server/Integration%20options%20with%20Maltego%20v3.pdf>, *Integration options with Maltego v3*, 16.6.2012.
- [11] PATERVA, Maltego Version 3 User Guide: Using the GUI, 12.6.2012., <http://www.paterva.com/malv3/303/M3GuideGUI.pdf>, *Maltego Version 3 User Guide: Using the GUI*, 16.6.2012.
- [12] PATERVA, Maltego forum, <http://www.paterva.com/forum/>, *Maltego forum*, 16.6.2012.
- [13] PYTHON SOFTWARE FOUNDATION, Python, <http://www.python.org/>, *Python Programming Language – Official Website*, 16.6.2012.
- [14] THE ECLIPSE FOUNDATION, Eclipse, <http://www.eclipse.org/>, *Eclipse*, 16.6.2012.
- [15] APPCELERATOR INC., PyDev, 10.4.2012., <http://pydev.org/>, *appcelerator PyDev*, 16.6.2012.

- [16] AVIRAM, Y., Python Amazon Simple Product API 1.0.0, <http://pypi.python.org/pypi/python-amazon-simple-product-api/1.0.0>, *python-amazon-simple-product-api 1.0.0*, 16.6.2012.
- [17] Z39.50, 28.5.2012., <http://en.wikipedia.org/wiki/Z39.50>, *Z39.50*, 16.6.2012.
- [18] LAV, A., PyZ3950, <http://www.panix.com/~asl2/software/PyZ3950/>, *PyZ3950*, 16.6.2012.
- [19] BEAZLEY, D., PLY (Python Lex-Yacc), <http://www.dabeaz.com/ply/>, *PLY (Python Lex-Yacc)*, 16.6.2012.
- [20] LIBRARY OF CONGRESS, MARC 21 Format for Bibliographic Data, 16.4.2012., <http://www.loc.gov/marc/bibliographic/ecbdhome.html>, *MARC 21 Format for Bibliographic Data*, 16.6.2012.
- [21] KREIBICH, C., A parser for Google Scholar, written in Python, 1.2.2012., <http://www.icir.org/christian/scholar.html>, *A parser for Google Scholar, written in Python*, 16.6.2012.
- [22] INNOVATION QUERY, INC., Innovative Query, Inc. <http://innovativequery.com/>, *Innovative Query, Inc.*, 16.6.2012.
- [23] INNOVATION QUERY, INC., Air Force Acquisitions Business Intelligence Case Study, 30.8.2009., http://innovativequery.com/Case_Studies/AirForceAcquisitions.pdf, *Air Force Acquisitions Business Intelligence Case Study*, 17.6.2012.

Sažetak

Računalno otkrivanje mreže povezanih pojmova

Na početku učenja ili istraživanja nekog područja ključno je obuhvatiti krupnu, grubu sliku područja kako bi se daljnji napori na učenju i istraživanju pravilno usmjeravali. Većina raspoloživih web-alata za pretraživanje pruža linearan ispis rezultata pri čemu je nemoguće uvidjeti međuovisnosti rezultata te važnost pojedinačnih rezultata u području pretraživanja. Stoga bi prilikom jednog takvog pretraživanja veliku pomoć pružila mogućnost automatskog, računalnog pronalaženja povezanih pojmova i definiranje njihove mreže.

U ovom radu istraženo je stanje znanosti i tehnike u području automatiziranog prikazivanja znanja pomoću generirane mreže povezanih pojmova važnih za neko područje te su evaluirani raspoloživi programski alati. Naglasak je stavljen na jednostavnost upotrebe za korisnike s minimalnim informatičkim vještinama te na dobavljalivost i troškove korištenja alata.

Istraživanje je pokazalo da alat Maltego u tom slučaju zadovoljava najviše kriterija. U Maltego je moguće implementirati neograničen broj načina pretraživanja te definirati načine stvaranja veza među pojmovima na grafu. Osim toga, Maltego posjeduje i mnoštvo opcija za uređivanje grafova te razne načine prikaza tih grafova. Stoga je pokazan način na koji Maltego može iskoristiti kao koristan alat za stvaranje mreže povezanih pojmova uz visoku interaktivnost s korisnikom. Također, uz upute za korištenje dosadašnjih rezultata rada, navedene su i upute za budući rad u Maltegu.

Ključne riječi: mreža povezanih pojmova, računalno otkrivanje, sustav pretraživanja, vizualno pretraživanje, upravljanje znanjem, open-source intelligence, Maltego, transformacije

Summary

Computerized discovery of network of interrelated terms

At the beginning of a study or research in a certain area it is crucial to get the big picture of an area and thereby make sure that further efforts in study and research are properly directed. Most of the available web-search tools provide only linear display of results from which it is impossible to recognize the interdependence and the importance of specific results in a research area. Therefore, a great help in such research would be the possibility of automatic, computerized discovery of related terms and defining their network.

This paper includes an investigation on present state of science and technology in the field of automated presentation of knowledge by generated network of interrelated terms relevant to a certain area and the evaluation of available software tools. Focus is placed on simplicity of use for users with minimal IT skills as well as on availability and low cost of the software tool.

Investigation has showed that application Maltego meets the highest criteria in this case. It is possible to implement an unlimited number of search methods in Maltego and define own ways of creating links between terms in graphs. In addition, Maltego has plenty of options for graph editing and various display modes. Therefore, it is shown how Maltego can be used as a useful tool for generating a network of interrelated terms with high user-interactivity. Also, this paper includes instructions for using current achievements as well as directions for any future work with Maltego.

Keywords: network of interrelated terms, computerized discovery, search system, visual search, knowledge management, open-source intelligence, Maltego, transforms

Skraćenice

API	<i>Application programming interface</i>	Sučelje s aplikacijskim programima
ASCII	<i>American Standard Code for Information Interchange</i>	Način kodiranja znakova temeljen na engleskoj abecedi
DNS	<i>Domain Name System</i>	Distribuirani hijerarhijski sustav poslužitelja na Internetu
DOI	<i>Digital Object Identifier</i>	Digitalni identifikator objekta
HTML	<i>HyperText Markup Language</i>	Prezentacijski jezik za oblikovanje hiperteksta
HTTP	<i>HyperText Transfer Protocol</i>	Protokol za prijenos informacija na Webu
IP	<i>Internet Protocol</i>	Mrežni protokol za prijenos podataka na Internetu
ODBC	<i>Open Database Connectivity</i>	Protokol koji se koristi za povezivanje baze podataka programa Microsoft Access s vanjskim izvorom podataka
PDF	<i>Portable Document Format</i>	Format za zapis dvodimenzionalnih dokumenata
RSS	<i>RDF Site Summary</i>	Skup web-formata rabljenih za web-stranice koje se često osvježavaju
SEO	<i>Search Engine Optimization</i>	Optimizacija za tražilice
TDS	<i>Transform Distribution Server</i>	Poslužitelj čija je svrha dijeljenje, upravljanje i implementacija transformacija diljem Weba
URL	<i>Uniform Resource Locator</i>	Web-adresa određenog resursa na Internetu
XML	<i>EXtensible Markup Language</i>	Jezik za označavanje podataka

Privitak

Uz rad je priložen CD-ROM s izvornim kodom te ovim radom u elektroničkom obliku.

Instalacija programske podrške

Za lokalno postavljanje transformacija potrebno je:

- Instalirati Maltego *Community Edition* sa službenog web-sjedišta⁵⁸ i registrirati se na istom sjedištu
- Instalirati Python verziju 2.7.2⁵⁹
- Instalirati *lxml* za Python 2.7⁶⁰ (radi na operativnim sustavima Microsoft Windows)

Upute za korištenje programske podrške

Za korištenje i pokretanje lokalnih transformacija potrebno je:

- Kopirati direktorij *Maltego* sa Python modulima transformacija u lokalni direktorij na računalu
- Uvesti konfiguracijsku datoteku *eva-konfiguracija.mtz* u Maltego (*Maltego -> Import -> Import configuration...*) ili samo pokrenuti *.mtz* datoteku
- Uvesti entitete *eva-entiteti.mtz* u Maltego (*Maltego -> Import -> Import entities...*) ili samo pokrenuti *.mtz* datoteku

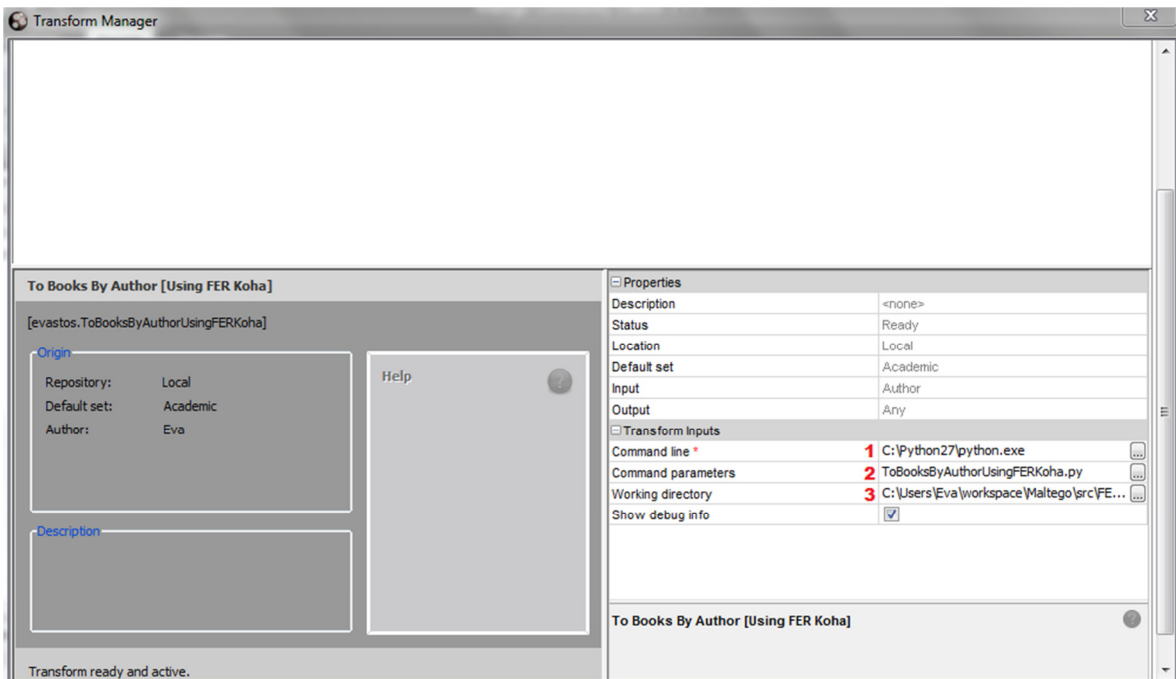
Definicije transformacija bi sada trebale postojati u Maltegu. Međutim, u definicijama transformacija se nalaze krivi putevi do datoteka. Pod *Manage -> Manage Transforms -> All Transforms* potrebno je pronaći svaku od transformacija i ažurirati sljedeće puteve do datoteka (slika 0.1):

- 1.** Put do *python.exe* datoteke. Pri standardnoj instalaciji je „*C:\Python27\python.exe*“
- 2.** Naziv pozivajućeg Python modula. Poklapa se s izvornim nazivom modula i nije ga potrebno mijenjati
- 3.** Put do lokalnog direktorija u kojemu se nalazi pozivajući modul. Potrebno ga je promijeniti. Npr. „...*Maltego\src\Amazon\ToAuthorsUsingAmazon.py*“

⁵⁸ <http://www.paterva.com/web5/client/download.php#Community>

⁵⁹ <http://www.python.org/getit/releases/2.7/>

⁶⁰ <http://www.lfd.uci.edu/~gohlke/pythonlibs/#lxml>



Slika 0.1 Transform Manager – putevi do datoteka