

Travica, Izidor

Undergraduate thesis / Završni rad

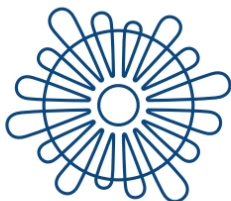
2017

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zadar / Sveučilište u Zadru**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:162:012874>

Rights / Prava: [In copyright](#) / [Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2024-04-25**



Sveučilište u Zadru
Universitas Studiorum
Jadertina | 1396 | 2002 |

Repository / Repozitorij:

[University of Zadar Institutional Repository](#)



Sveučilište u Zadru

Odjel za ekonomiju

SVEUČILIŠNI PREDDIPLOMSKI STUDIJ MENADŽMENTA



Izidor Travica

Web inteligencija

Završni rad

Zadar, 2017.

Sveučilište u Zadru

Odjel za ekonomiju
SVEUČILIŠNI PREDDIPLOMSKI STUDIJ MENADŽMENTA

Web inteligencija

Završni rad

Student/ica:

Izidor Travica

Mentor/ica:

Dr. sc. Ante Panjkota

Zadar, 2017.



Izjava o akademskoj čestitosti

Ja, **Izidor Travica**, ovime izjavljujem da je moj **završni** rad pod naslovom **Web inteligencija** rezultat mogega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na izvore i radove navedene u bilješkama i popisu literature. Ni jedan dio mogega rada nije napisan na nedopušten način, odnosno nije prepisan iz necitiranih radova i ne krši bilo čija autorska prava.

Izjavljujem da ni jedan dio ovoga rada nije iskorišten u kojem drugom radu pri bilo kojoj drugoj visokoškolskoj, znanstvenoj, obrazovnoj ili inoj ustanovi.

Sadržaj mogega rada u potpunosti odgovara sadržaju obranjenoga i nakon obrane uređenoga rada.

Zadar, 6. listopada 2017.

Sažetak:

Ovaj rad obrađuje različite načine primjene web inteligencije u poslovanju. Radom su tako obuhvaćene metode i načini na koje se web inteligencija primjenjuje u različitim segmentima poslovanja poput marketinga, CRM-a, upravljanja tvrtkom i korisničkoj podršci. Poseban fokus stavljen je na primjere iz prakse kako bi se pokazala uspješnost primjene web inteligencije i ukazalo na važnost koju ima sama primjena. Razmotreno je i na koje načine tvrtke mogu ostvariti prednost primjenom web inteligencije u donošenju poslovnih odluka i edukaciji zaposlenika. Posebno je stavljen naglasak i na važnost rudarenja weba korištenjem raznih metoda i algoritama poput sentiment analize kako bi se prikupilo što više podataka o kupcima i pokušalo razumjeti njihovo ponašanje prilikom kupnje. Rezultati do kojih smo došli ukazuju na to da primjena web inteligencije u poslovanju pridonosi boljem razumijevanju potrošača, boljoj mogućnosti prilagođavanja potrošačima i razvoju boljeg odnosa sa potrošačima.

Ključne riječi

Web inteligencija, semantički web, rudarenje web-a, web agenti, web ekspertni sustavi, sentiment analiza

SADRŽAJ:

1	UVOD	1
2	Tehnološka osnova za Web inteligenciju	3
2.1	Internet.....	3
2.2	Web servisi	7
2.3	ETL proces i alati	10
2.3.1	Faza ekstrakcije podataka (<i>Extract</i>)	10
2.3.2	Faza transformiranja podataka (<i>Transform</i>).....	11
2.3.3	Faza učitavanja podataka (<i>Load</i>).....	11
2.3.4	ETL alati.....	12
2.4	Skladišta i spremišta podataka (<i>Data Warehouse</i> i <i>Data Marts</i>)	12
3	Paradigma semantičkog weba i web inteligencija.....	14
3.1	Arhitektura semantičkog weba	15
3.2	Standardi semantičkog weba	17
3.3	Primjeri web inteligencije zasnovane na paradigmi semantičkog weba	19
3.3.1	Intelligentni pretraživač online usluga javne uprave grada Zaragoze	19
3.3.2	Kontekstualno pretraživanje - Volkswagen	20
3.3.3	Predlaganje povezanih ideja u stvarnom vremenu - financijski sektor	21
3.3.4	Web aplikacija CRUZAR za personalizirane rute	22
4	Unaprjeđenje web sjedišta primjenom web inteligencije.....	23
4.1	Intelligentni dizajn web sjedišta	23
4.2	Dinamičko web sjedište.....	24
4.3	Primjeri unaprjeđenja web sjedišta tvrtke pomoću web inteligencije	25
4.3.1	Netflix.....	25
4.3.2	Facebook	26
4.3.3	Fits.me Virtual Dressing Room.....	26
5	Web inteligencija zasnovana na rudarenju podataka	29
5.1	Metoda asocijacije	30
5.2	Metoda klasifikacije	30
5.3	Metoda klasteriranja	31
5.4	Metoda stabla odlučivanja	32
5.5	Primjena rudarenja weba u poslovanju.....	32
5.5.1	Personalizirano korisničko iskustvo u B2C e-trgovini – Amazon	32
5.5.2	Metoda klasteriranja kao način segmentacije tržišta.....	33

5.5.3	Asocijacijska pravila za analizu potrošačke košarice u web trgovinama.....	35
6	Web agenti kao elementi web inteligencije.....	37
6.1	Inteligentni digitalni agent Alice – podrška pacijentima.....	38
6.2	Online aukcijski sustav temeljen na web agentima	38
6.3	Inteligentni osobni asistent – Bixby	39
7	Web inteligencija u upravljanju znanjem.....	41
7.1	Web ekspertni sustavi.....	41
7.1.1	Primjer Web ekspertnog sustava - WITS	42
7.2	Web sustavi za stjecanje novih znanja	43
7.2.1	Primjer Web sustava za stjecanje novih znanja - Dow Chemical-ov trening za intervju	44
7.2.2	Adobe Captivate Prime	44
8	Web inteligencija i društvene mreže	46
8.1	Sentiment analiza.....	46
8.1.1	Upotreba sentiment analize u praksi	47
8.2	Rudarenje podataka na društvenim mrežama	47
8.2.1	Rapleaf Inc. – rudarenje web-a	48
	Rasprava	49
	Zaključak	52
	Literatura	53
	Popis slika	61
	Summary	62

1 UVOD

Snažan razvoj interneta i *world wide web-a*, doveo je do novog informatičkog doba. Web predstavlja sasvim novi medij za komunikaciju i suradnju. Web je značajno revolucionirao način na koji se informacije prikupljaju, pohranjuju, prikazuju, dijele i koriste. Razvoj web-a ponudio je nove mogućnosti i izazove za mnoga područja poput poslovanja, trgovine, marketinga, financija, edukacije, istraživanja i razvoja. Razvojem web-a razvile su se i mnoge popratne tehnologije od kojih je jedna i Web inteligencija. Web inteligencija koristi umjetnu inteligenciju i naprednu informacijsku tehnologiju na različite web i internetske servise.¹ Web inteligencija je jako široko područje u informacijskoj tehnologiji i uz to je relativno novo područje. Kroz ovaj rad dan je pregled nekih područja web inteligencije, te je kroz karakteristične primjere objašnjena njihova primjena u poslovnoj praksi.

Primjena informacijskih tehnologija postala je neizostavan dio svakog poslovanja. Informacijske tehnologije, naročito internet, omogućavaju tvrtkama širenje svojih tržišta na globalnoj razini, ukratko cijeli svijet postaje tržište. Osim toga informacijske tehnologije pomažu tvrtkama u poboljšanju poslovanja i stjecanju konkurentske prednosti. Problem koji ćemo istraživati u ovom radu biti će kako primijeniti web inteligenciju na poslovanje kako bi tvrtke ostvarile konkurentsku prednost i općenito unaprijedile segmente vlastitog poslovanja.

Cilj ovog istraživanja je pokazati prednost korištenja web inteligencije u poslovanju, a naročito u e-poslovanju. Potom pokazati načine primjene web inteligencije u poslovanju, odnosno koje metode i pristupi web inteligencije se koriste u tu svrhu i na koji način. Pokazat ćemo u kojim područjima i aspektima poslovanja je moguće primijeniti web inteligenciju.

Svrha ovog istraživanja je pokazati zašto bi poduzeća trebala primjenjivati web inteligenciju u svom poslovanju i koje najvažnije koristi proizlaze iz primjene web inteligencije u e-poslovanju.

¹ Y. Y. Yao, et. al., *Web Intelligence (WI) Research Challenges and Trends in the New Information Age*, Springer, 2001., p. 1-2

Istraživačka pitanja kojima ćemo se baviti u okviru ovog rada su:

1. Što je web inteligencija, od čega se sastoji, kada se javlja i zbog čega?
2. Koje metode i pristupi se koriste u web inteligenciji?
3. Kako se web inteligencija primjenjuje u CRM-u, kako u marketingu, upravljanju tvrtkom, podršci klijentima?
4. Koje prednosti ostvaruju tvrtke korištenjem web inteligencije?
5. Koji su uspješni primjeri primjene web inteligencije u praksi?
6. Koliko je uopće moguće vjerovati inteligentnim sustavima pri donošenju poslovnih odluka?

Analizom relevantne literature, proveli smo istraživanje o primjeni web inteligencije u poslovanju. Povijesnu metodu koristili smo za dobivanje uvida u kronološki redoslijed razvoja i pojave web inteligencije, potom smo koristili deskriptivnu metodu kako bi opisali pojam web inteligencije i ono što ona predstavlja. Metodu analize koristili smo u raščlanjivanju web inteligencije na zasebne elemente od kojih se ona sastoji, poput web agenata, web ekspertnih sustava, rudarenja weba, te semantičkog weba. Metodom kompilacije potkrijepili smo elemente na koje smo raščlanili web inteligenciju, korištenjem tuđih radova, rezultata i opažanja. Korištena je metoda sinteze kako bi na temelju istraživanja drugih autora došli do nekih važnijih zaključaka o prednostima korištenja web inteligencije, a potom smo metodom dedukcije došli do zaključaka o prednostima primjene web inteligencije u poslovanju. Korištena je i metoda studije slučajeva kako bi pokazali konkretne koristi od primjene web inteligencije u poslovanju.

2 Tehnološka osnova za Web inteligenciju

Web inteligencija je relativno novo multi-disciplinarno znanstveno i stručno područje, kao takvo je još uvijek u fazi razvoja. Općenito, da bi tvrtke mogle uspješno primjenjivati web inteligenciju, najprije moraju zadovoljiti određene tehnološke zahtjeve. Među brojnim zahtjevima navest ćemo samo najvažnije. Svakako je nužni zahtjev postojanje interneta (internetske mreže), popratnih web servisa i protokola, potom rješenja za pripremu i skladištenje podataka, usvajanje standarda semantičkog weba, primjena sigurnosnih standarda, suvremenih kriptografskih rješenja itd. Upravo ova cjelina obrađuje osnove navedenih tehnoloških zahtjeva za Web inteligenciju.

2.1 Internet

Moderan svijet je nezamisliv bez postojanja interneta. Internet koristimo na računalima, pametnim telefonima, tablet računalima, pa čak i na TV-u. Nadalje, suvremeno poslovanje gotovo je nezamislivo bez interneta, a bez njega neke tvrtke ne bi ni postojale. Primjer takvih tvrtki su Amazon, eBay i AliExpress, koje uglavnom posluju preko interneta.

Počeci interneta vezani su za povijesni događaj lansiranja „Sputnik-a“, prvog objekta izrađenog od strane ljudi koji je poslan u orbitu. Amerikanci su postali svjesni činjenice kako podosta zaostaju u znanosti i tehnologiji. Ponukan time, tadašnji predsjednik Eisenhower 1958. osniva *Advanced Research Projects Agency* (skraćeno ARPA), agenciju koja je trebala osigurati da komunisti nikada ne pobijede Ameriku u bilo kojoj utrci u tehnologiji. Jedan od ARPA-inih ureda bio je *Information Processing Techniques Office* (skraćeno IPTO), koji je financirao istraživanja u računalnim znanostima. Njihova istraživanja bila su vrlo uspješna, a istraživanja u segmentu umrežavanja dovela su do pojave Interneta. Godine 1969. nastaje projekt ARPAnet pod vodstvom Larryja Robertsa, voditelja IPTO ureda.²

ARPAnet je funkcionirao na teoriji razmjene paketa, a teorija je opisana u radu objavljenom od strane Leonarda Keirrocka s MIT-a, koji je kasnije prvi objavio knjigu na istu temu.³

² L. Kleinrock, 'History of the Internet and Its Flexible Future', *IEEE Wireless Communications*, vol. 15, no. 1, 2008., p. 10

³ B. M. Leiner et. al., *Brief History of the Internet*, [web stranica], 1997.,: <http://www.internetsociety.org/internet/what-internet/history-internet/brief-history-internet> (pristupljeno: 05.07.2017.)

Od tada zapravo počinje razvoj interneta koji traje još i danas. Iako je nastao 1969. godine, internet postaje dostupan za javnost tek 1991. godine i to razvojem World Wide Web-a.

World Wide Web, ili skraćeno www, izumio je Tim Berners-Lee, istraživač koji je radio u CERN-u. Naime, on je tražio način na koji fizičari mogu dijeliti informacije preko velikih udaljenosti bez toga da svi koriste isti tip *hardware-a* i *software-a*. Njegova ideja kulminirala je u radu koji je predlagao veliku *hypertext* bazu podataka s *linkovim-a*.⁴

Povijest razvoja interneta je opsežna, a gore su navedena dva važnija događaja iz te povijesti. Dakle, ARPAnet je pružio temelje za daljnji razvoj interneta, a pristup internetu je zapravo omogućen preko World Wide Web-a (skraćenica WWW), prefiks www sadržan je u većini web stranica na internetu. Web stranica ne mora nužno sadržavati prefiks www.

Internet je zapravo najveća računalna mreža na svijetu koja povezuje milijune računala. Mreža je grupa od dvoje ili više računala koja su zajedno povezana.

Isto tako, postoje dvije glavne vrste računalnih mreža od kojih razlikujemo LAN i WAN.

LAN (*Local Area Network*) se odnosi na dvoje ili više međusobno povezanih računala koji dijele određene resurse u manjem geografskom području. Primjerice kućna mreža ili uredska mreža.

WAN (*Wide Area Network*) se tipično sastoji od dva ili više LAN-a. Računala su udaljenija, a povezana su preko telefonske mreže ili radio valova. Internet je zapravo najveći WAN u povijesti.⁵ No, interesantno je kako internet zapravo funkcionira.

Funkcioniranje interneta zasniva se na izmjeni podatkovnih paketa od jednog računala do drugog primjenom više protokola od kojih je najvažniji takozvani TCP/IP protokol. Prvenstveno, da bi računala mogla međusobno komunicirati potrebne su im IP adrese. IP adrese su unikatne za svako računalo koje je povezano na internet, a možemo je usporediti i sa telefonskim brojem. IP adrese su obično u obliku nnn.nnn.nnn.nnn, gdje n zamjenjuje broj u rasponu od 0 do 255. Kratica IP stoji za *Internet Protocol*, kojeg ćemo obraditi malo kasnije u ovom poglavlju. Zasad ćemo se dotaknuti verzija IP-a jer su one vezane za IP adrese. Dakle, razlikujemo IPv4 i IPv6 verzije. IPv4 verzija je prva verzija internetskog protokola koja se upotrebljavala, a koristi se i dan danas. Postoji preko 4 milijarde IPv4 adresa, i iako ih je puno,

⁴ M. Bryant, *20 years ago, the World Wide Web opened to the public*, The Next Web (INSIDER), [web stranica], 2011., https://thenextweb.com/insider/2011/08/06/20-years-ago-today-the-world-wide-web-opened-to-the-public/#.tnw_5o4nxfsx (pristupljeno: 05.07.2017.)

⁵ *Internet 101: What is Internet?*, GCF LearnFree.org, Goodwill Community Foundation Inc., [web stranica], 2013. <https://www.gcflearnfree.org/internetbasics/what-is-the-internet/1/> (pristupljeno: 05.07.2017.)

nije ih dovoljno da bi zadovoljile potrebe rastućeg broja korisnika i pružatelja sadržaja spojenih na internet. IPv6 verzija je noviji sistem numeriranja koji pruža veći broj adresa nego IPv4 verzija. Najveća razlika ove dvije verzije je broj IP adresa. IPv4 broji nešto preko 4 milijarde adresa (2^{32}), dok IPv6 broji 2^{128} adresa. Tehnička funkcionalnost Interneta je ista s obje verzije, a verzije mogu raditi paralelno jedna s drugom. Mreže koje koriste IPv6 verziju podržavaju i IPv4 adrese i IPv6 adrese. Oblik IPv4 adrese je isti kao i onaj koji smo naveli gore, dok je oblik IPv6 adrese u heksadecimalnom obliku, odnosno zapisan u vidu 32 heksadekadske znamenke.⁶

Web serveri na internetu imaju jedinstvenu IP adresu, pa se postavlja pitanje zašto onda web stranice vidimo u obliku www.primjer.hr? Razlog tomu je jednostavnost primjene simboličkih ekvivalenata umjesto pripadajućih numeričkih IP vrijednosti što nas dovodi do tzv. DNS-a.

DNS (eng. *Domain Name System*) predstavlja servis za prevođenje simboličkih adresa u pripadajuće numeričke IP adrese i obrnuto. Dakle, DNS mapira stranicu primjer.hr u IP adresu poslužitelja 90.120.220.140, i to je samo jedan od primjera kako taj servis radi.

Svaki javni server ima svoju dodijeljenu IP adresu, i da upišemo IP adresu web servera Sveučilišta u Zadru, u obliku koji je naveden gore, u internet tražilicu učitalo bi nam web stranicu Sveučilišta.

U narednim pasusima ćemo detaljnije govoriti o TCP/IP protokolima, a kasnije u ovom poglavlju ćemo se kratko osvrnuti i na ostale protokole poput UDP-a, HTTP-a, FTP-a, SMTP-a, te POP/IMAP-a, koji su važni za korištenje interneta.

IP protokol (eng. *Internet Protocol*) je skup tehničkih pravila koja definiraju kako računala komuniciraju preko interneta. Svi podaci koji dolaze preko TCP-a, UDP-a, ICMP-a i IGMP-a prenose se kao IP *datagram*-i. IP pruža najbolju moguću uslugu dostave *datagram-a* bez spajanja (eng. *connectionless*). Pod „najboljom mogućom“ uslugom podrazumijeva se da nema garancije da će IP *datagram* uspješno stići do odredišta. Pojam „bez spajanja“ znači da IP ne održava nikakve informacije o stanju veze o povezanim *datagram-ima* unutar mrežnih elemenata (primjerice unutar usmjernika – eng. *router*). Svakim *datagram-om* se rukuje neovisno od svih ostalih.⁷

TCP protokol (eng. *Transmission Control Protocol*) upravlja stanjem veze, odnosno provjerava informacije o vezi koju održavaju obje strane. Prije negoli ijedan kraj može slati podatke

⁶ *IPv4 and IPv6*, [web stranica], http://biotech.law.lsu.edu/blog/ipv4_ipv6.pdf (pristupljeno: 05.07.2017.)

⁷ K. R. Fall i W. R. Stevens, *TCP/IP Illustrated*, Addison-Wesley Professional, 2011., p.196

drugom, veza između njih mora biti uspostavljena. TCP veza sastoji se od dvije IP adrese i dva broja vrata (eng. *porta*). Pomoću IP adrese i broja *port-a* identificira se svaka strana.⁸

UDP protokol (eng. *User Datagram Protocol*) dio je TCP/IP paketa. On pruža vezu između dva procesa na oba kraja prijenosa. To je sloj TCP/IP modela, koji pruža uslugu potpunog prijenosa podataka aplikacijama.⁹

HTTP protokol (eng. *Hypertext Transfer Protocol*) je još jedan od protokola za prijenos podataka. Započeo je kao jednostavan protokol koji je bio baziran na jednostavnom principu; klijent (internet preglednik) bi izdao mali set komandi i primao bi povratno odgovarajuće informacija od strane servera, odnosno web stranice. Originalna verzija bila je HTTP 0.9, a od 1999. službena verzija je 1.1, koja se koristi i dan danas. U današnje vrijeme, HTTP služi za prijenos web stvari, poput Java Applet-a, ASP-a (eng. *Active Server Pages*), PERL skripta¹⁰, itd.¹¹

FTP protokol (eng. *File Transfer Protocol*) je protokol za prijenos podataka. Nalazi se u takozvanom „aplikacijskom sloju“. FTP prebacuje kopiju datoteke, a original je obično i dalje prisutan na izvornom *host-u*, omogućujući tako kopiranje nanovo prema zahtjevima korisnika.¹²

SMTP protokol (eng. *Simple Mail Transfer Protocol*) pruža mehanizam za prijenos internetske pošte (*mail*). Pošta se prenosi od pošiljatelja do primatelja kada su oba spojena na isti prijenosni servis ili kada su spojeni preko jednog ili više SMTP-servera kada pošiljatelj i primatelj nisu spojeni na isti prijenosni servis.¹³

POP (eng. *Post Office Protocol*) je protokol koji omogućuje pristup elektroničkoj pošti. POP radi na principu da elektronička pošta poslana od strane pošiljatelja dolazi k primatelju i pohranjuje se na primateljevo računalo, a ne na server. Zadane postavke su obično konfigurirane da se elektronička pošta pohranjuje lokalno (računalo, smartphone, tablet) i nigdje drugdje. Dok IMAP (eng. *Internet Message Access Protocol*) protokol omogućuje da se elektronička pošta pohranjuje na server i da se njoj pristupa putem autentikacije korisničkim

⁸ Op. Cit., p.569

⁹ B. M. Wilamowski i J. D. Irwin, *Industrial Communication Systems*, CRC Press, 2011., p.59-2

¹⁰ PERL – programski jezik opće namjene kreiran od strane Larryja Wellsa

¹¹ W. Goralski, *The Illustrated Network*, Massachusetts, Morgan Kaufmann Publishers, 2009., p. 569-570

¹² Op. Cit. p. 509

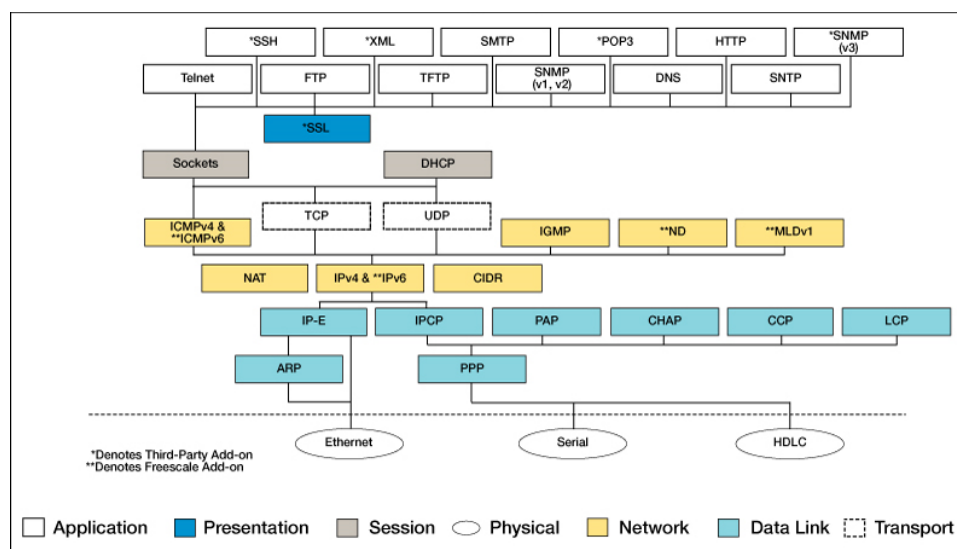
¹³ J. B. Postel, *Simple Mail Transfer Protocol*, University of Southern California, Information Sciences Institute, 1982. p. 2-3

imenom i lozinkom. Elektronička pošta se privremeno pohranjuje, ali se ne sprema na uređaj koji se koristi.¹⁴

Nakon protokola, osvrnut ćemo se i na sigurnosni sloj, odnosno SSL.

SSL (eng. *Secure Sockets Layer*) je namijenjen pružanju sigurnosti u komunikaciji internet klijent-server. SSL je podijeljen na dva sloja: *record layer* i *SSL handshake protocol*. *Record layer* pruža povjerljivost, autentikaciju i zaštitu od reprodukcije preko protokola poput TCP-a. *SSL handshake protocol*, je protokol za razmjenu takozvanih ključeva koji pokreće i sinkronizira kriptografsko stanje na oba kraja. Nakon što je protokol razmjene ključeva kompletiran, osjetljivi podaci se mogu poslati preko SSL record layer-a.¹⁵

Slika 1. prikazuje sve protokole TCP/IP modela. Za potrebe rada obradili smo samo neke koji su najvažniji za primjenu Web inteligencije u poslovanju.



Slika 1 Prikaz svih protokola u TCP/IP modelu¹⁶

2.2 Web servisi

Web servise je WWW konzorcij (skraćeno W3C) definirao na sljedeći način:

¹⁴ R. Holly, *Geek 101: Beginners guide to IMAP vs. PoP*, [web stranica], 2013., <https://www.geek.com/chips/geek-101-pop-vs-imap-1536343/> (pristupljeno: 19.07.2017.)

¹⁵ D. Wagner i B. Schneier, *Analysis of the SSL 3.0 protocol*, University of California, 1997., p 1

¹⁶ MQX Real-Time TCP/IP Communication Suite (RTCS), [online slika], http://www.nxp.com/assets/images/en/block-diagrams/MQX_TCP_IP_BD_IMG.jpg (pristupljeno: 12.08.2017)

„A Web service is a software system designed to support interoperable machine-to-machine interaction over a network. It has an interface described in a machine-processable format (specifically WSDL). Other systems interact with the Web service in a manner prescribed by its description using SOAP messages, typically conveyed using HTTP with an XML serialization in conjunction with other Web-related standards“¹⁷

Iz definicije vidimo da je web servis zapravo *software* koji je dizajniran za interoperabilnu¹⁸ komunikaciju među uređajima priključenim na određeni tip komunikacijske mreže..

Serijalizacija (*serialization*) je proces transformiranja objekta u bajtove koji se tada mogu koristiti u rekonstrukciji originalnog objekta.¹⁹ Objekt može biti varijabla, podatkovna struktura, funkcija ili metoda.

XML (eng. *eXtensible Markup Language*) predstavlja skriptni jezik proširenih mogućnosti koji omogućuje definiranje podataka i njihovih korisnika i/ili kreiranje Web stranica.²⁰ Prije smo govorili da web servis omogućava interoperabilnu komunikaciju između računala, a ta komunikacija moguća je samo ako se podaci prenose u računalima razumljivim jezikom. Upravo tome služi XML.

Web servise čine tri protokola: SOAP (eng. *Simple Object Accessing Protocol*), WSDL (eng. *Web Services Description Language*) i UDDI (eng. *Universal Description, Discovery and Interoperability*). Ova tri protokola građena su na bazi XML-a i HTTP-a. SOAP definira jedinstven način prijenosa XML podataka, a WSDL dozvoljava pružateljima servisa da specificiraju što web servis može raditi, gdje se nalazi i kako mu pristupiti. UDDI pruža

¹⁷ D. Booth et. al., 'Web Services Architecture', *W3C Working Group Note*, 2004., p. 7, <https://www.w3.org/TR/ws-arch/wsa.pdf> (pristupljeno: 09.07.2017.)

¹⁸ Interoperabilnost (eng. *interoperability*) – karakteristika proizvoda ili sistema, čija se sučelja kompletno razumiju, kako bi funkcionirali sa ostalim proizvodima ili sistemima, sadašnjim ili budućim, u implementaciji ili pristupu, bez ikakvih ograničenja.

J. McCreesh i E. Daniel, *Definition of Interoperability*, [web stranica], <http://interoperability-definition.info/en/> (pristupljeno: 15.09.2017.)

¹⁹ K. Woolcock, *Serialization in Java (Binary and XML)*, Iowa State University, 2014., dostupno na: <http://web.cs.iastate.edu/~smkautz/cs430s14/tutorials/current/Serialization%20in%20Java%20Tutorial.pdf> (pristupljeno: 09.07.2017.)

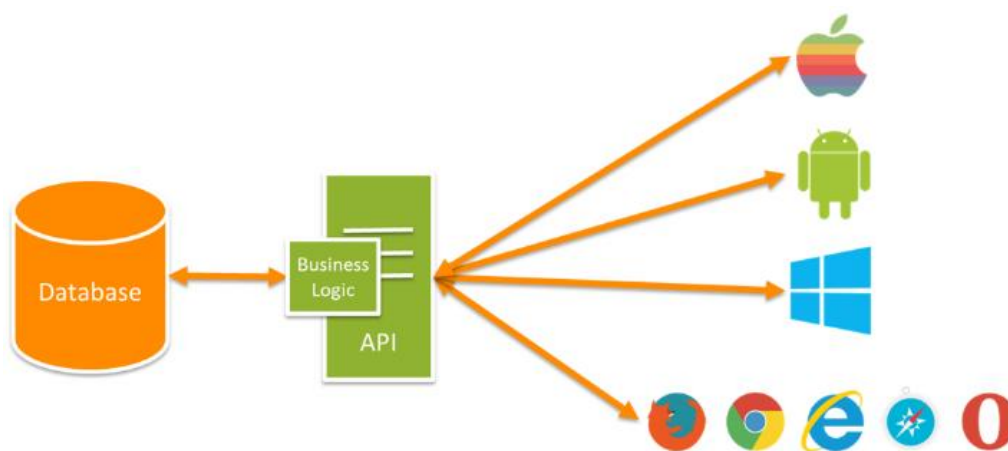
²⁰ 'XML (eXtensible Markup Language)', *Encyclopedia of Information Systems*, 2003., dostupno na: <http://www.sciencedirect.com/science/article/pii/B0122272404001994> (pristupljeno: 09.07.2017.)

mehanizam za klijente kako bi mogli pronaći druge Web servise.²¹

Prednosti Web servisa u poslovanju proizlaze iz XML sučelja i pristupnih opisa u WSDL-u. WSDL opis koristi se za pokretanje servisno orijentiranih arhitektura koje omogućuju integraciju poslovnih aplikacija (eng. *EAI – Enterprise Application Integration*), integraciju *business-to-business (B2B)* aplikacija i *grid computing* (korištenje resursa većeg broja računala na mreži za rješavanje zajedničkog problema).²²

Kod web servisa potrebno je spomenuti i postojanje web API-a (eng. *Application Programming Interface*). API je skup metoda pristupanja podacima u inače zatvorenim sistemima. API daje programerima i developerima alate potrebne za razvoj *softwarea* i servisa s podacima i servisima iz eksternih izvora.²³

Web API je API posebno dizajniran za specifične web aplikacije. Web API je jedan od standarda trenutnog web ekosistema: ekosistema temeljenog na otvorenosti i dijeljenju. Web servisi više nisu zatvoreni sistemi, već rade skupa omogućujući protok podataka među njima. To se ostvaruje kroz web API.²⁴ Na slici 2 možemo vidjeti arhitekturu centralnog web API-ja u kojem su sadržane sve poslovne logike.



Slika 2 Prikaz arhitekture centralnog Web API-ja²⁵

²¹ E. F. Gehringer, *Web services: SOAP, WSDL, UDDI*, CSC/ECE 517 Lecture Notes, North Carolina State University, 2005., dostupno na: <https://people.engr.ncsu.edu/efg/517/f05/common/lectures/lectures/lec24.pdf> (pristupljeno: 09.07.2017.)

²² C. Ferris i J. Farrell, 'What Are Web Services?', *Communications of the ACM*, vol. 46, no. 6, 2003., p.31

²³ J. P. Michel, *Web Service APIs and Libraries*, Chicago, American Library Association, 2013., p. xi

²⁴ J. P. Michel, loc. cit.

²⁵ M. Kearn, *Architecture with a central API*, [online slika], 05.01.2015., <https://blogs.msdn.microsoft.com/martinkearn/2015/01/05/introduction-to-rest-and-net-web-api> (pristupljeno:12.08.2017.)

2.3 ETL proces i alati

U suvremenom poslovanju ključnu ulogu u ostvarivanju konkurentske prednosti imaju prave informacije. Prava informacija, pravoj osobi u pravo vrijeme od velikog je značaja. No kako doći do smislenih informacija?

Ponajprije, važno je znati da je informacija zapravo smisleno transformirani podatak. Primjerice, ako imamo podatke prodanih artikala u 50 trgovina, tada to ne znači apsolutno ništa. Kada te podatke organiziramo i analiziramo i vidimo da nam se u prodavaonicama koje su na jugu zemlje lošije prodaju određeni artikli, a na sjeveru zemlje bolje, tada je riječ o korisnoj informaciji za menadžment koji će krenuti u daljnju analizu uzroka tih oscilacija.

Upravo tome služe nam ETL alati. ETL je skraćeno od *Extract, Transform, Load*. ETL procesi kritični su u uspjehu projekta skladišta podataka (*Data Warehouse*). Proces se sastoji od ekstrakcije podataka iz jednog sistema, transformiranja u skladu sa dizajnom skladišta podataka i učitavanja u skladište podataka.

Drugim riječima, faza ekstrakcije prikuplja podatke iz različitih izvora podataka, zatim ih u fazi transformacije transformira u podatke prema zahtjevima skladišta podataka te ih uspješno pohranjuje, odnosno učitava na odredište (u skladište podataka).²⁶

Treba paziti da se faze unaprijed predefiniraju, jer moramo imati na umu da kapacitet skladišta podataka nije ograničen i da se svako proširenje dodatno plaća, te nas tako troškovno dodatno opterećuje. Stoga treba odrediti koju vrstu i kakav tip podataka treba prikupljati, transformirati i pohranjivati u skladište podataka.

2.3.1 Faza ekstrakcije podataka (*Extract*)

Ovo je prva faza u procesu, u njoj se podaci prikupljaju iz različitih izvora koji sadrže informacije koje trebaju prebaciti u skladište podataka. Neki od izvora mogu biti samo obične datoteke bez integriteta. Prikupljeni podaci u ovoj fazi postaju dostupni za daljnju obradu. Cilj ove faze je prikupiti određene podatke iz izvora korištenjem što je manje moguće resursa. To

²⁶ R. Katragadda, S. S. Tirumala i D. Nandigam, 'ETL tools for Data Warehousing: An empirical study of Open Source Talend Studio versus Microsoft SSIS', *Computing Conference Papers*, 2015., p. 1-2 <http://unitec.researchbank.ac.nz/handle/10652/3366> (pristupljeno: 09.07.2017.)

znači da ova faza treba biti osmišljena na način da ne utječe na izvorišni sistem ni na koji način, u smislu performansi, vremena odaziva ili bilo kakvih smetnji. Izvlačenje podataka može se raditi na više načina, i to kao ažuriranje, inkrementalno izvlačenje ili potpuno izvlačenje. Frekvencija izvlačenje podataka je krucijalna u slučaju inkrementalnog izvlačenja jer veličine podataka mogu iznositi i do 10GB.²⁷

2.3.2 Faza transformiranja podataka (*Transform*)

Faza transformiranja je najkompleksniji dio ETL procesa. Svi podaci u ovoj fazi su već izvučeni iz svojih izvora, no oni mogu izgledati drugačije nego kako je predviđeno u shemi skladišta podataka. Podaci se tada trebaju formatirati na način da odgovaraju zahtjevima skladišta podataka. U nekim slučajevima podaci će biti drugačije prikazani da bi ostvarili svoju svrhu. U transformacijskom procesu, primjenjuje se set pravila kako bi se transformirali podaci i kako bi odgovarali zahtjevima skladišta podataka. Osim toga, u aspekt transformacijskog procesa spada i takozvano čišćenje podataka. Ovo je važan proces unutar ove faze jer osigurava da samo kvalitetni podaci ulaze u skladište podataka. Čišćenje se provodi prema sljedećim pravilima:

- 1) Transformiranje različitih identifikacija u unikatne prikaze. Primjerice kategorije spola u Muško, Žensko, Nepoznato, ili M, Ž, null itd.
- 2) Konvertiranje null vrijednosti u standardiziranu „Nije dostupno“ vrijednost.
- 3) Konvertiranje brojeva telefona, poštanskih brojeva u standardiziranu formu
- 4) Pridružiti polja adresa jedno s drugim. Primjerice (država, grad, poštanski broj, grad, grad, ulica).²⁸

2.3.3 Faza učitavanja podataka (*Load*)

Jednom kada su podaci izvučeni i transformirani prema zahtjevima skladišta podataka, podrazumijeva se da su podaci spremni za učitavanje u skladište. No, prije učitavanja u skladište podataka treba obratiti pozornost na neke od aspekata poput načina na koji će se podaci učitati, utjecaj i implikacija učitavanja, kao i rukovanje implikacijama. Proces učitavanja može utjecati

²⁷ loc. cit.

²⁸ R. Katragadda et. al., loc. cit.

na brzinu obrade kod servera kao i kod analize. Isto tako, krucijalno je izbjegavati modificiranje baze podataka prilikom učitavanja podataka.²⁹

2.3.4 ETL alati

Postoji bezbroj ETL alata koji pomažu u upravljanju bazom podataka i skladištem podataka. Tvrtkama na raspolaganju stoje dvije opcije kada se radi od ETL alatima. Prva opcija je mogućnost razvijanja vlastitog ETL alata koristeći ETL rješenja otvorenog koda (eng. *open-source*), a druga opcija je da se tvrtka odluči na već postojeće i isprobano komercijalno rješenje. Dakako, postoje prednosti i nedostaci korištenja obje opcije, no u to nećemo ulaziti.

Neka od poznatijih komercijalnih rješenja su *Informatica – PowerCenter*, *IBM – Infosphere Information Server* i *Oracle – Data Integrator*, dok su najpoznatiji open-source ETL alati *Talend Open Studio*, *Pentaho Data Integration* i *Clover ETL*.

2.4 Skladišta i spremišta podataka (*Data Warehouse* i *Data Marts*)

U prethodnom poglavlju govorio sam o ETL procesu i alatima. Da bi tvrtke mogle upotrebljavati informacije kao pomoć u donošenju poslovnih odluka, one moraju biti negdje pohranjene i dostupne u svakom trenutku. Upravo tome služe skladišta podataka (eng. *Data Warehouse*) i spremišta podataka (eng. *Data Marts*).

Skladišta podataka započela su se razvijati 1970-ih, kada je Bill Inmon skovao pojam *Data Warehouse*. Mnogi smatraju Billa Inmona „ocem“ skladišta podataka, jer je on prvi počeo razglabati o teorijskim i praktičnim principima skladišta podataka. Njegov rad, kao pionira skladištenja podataka, odrazio se 1990-ih kada je samostalno krenuo u posao. Osnovao je svoju prvu tvrtku *Prism Solutions*, čiji je jedan od glavnih proizvoda bio *Prism Warehouse Manager*, jedan od prvih industrijskih alata za kreiranje i upravljanje skladištem podataka.

Bill Inmon objavio je knjigu 1992. godine pod nazivom *Building the Data Warehouse*, koju većina stručnjaka ubraja u jedno od glavnih djela te industrije. Knjiga je trenutno u četvrtom izdanju i ne prestaje biti važan dio u poslu svakog profesionalca koji se bavi podacima, sa svojom istančanom teoretskom podlogom i svakodnevnim primjerima iz prakse.³⁰

²⁹ Loc. cit.

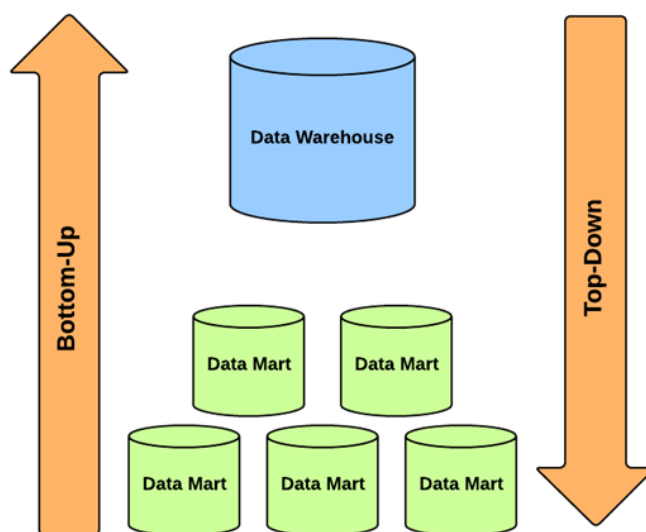
³⁰ P. Williams, *A Short History of Data Warehouseing*, [web stranica], 2012., „<http://www.dataversity.net/a-short-history-of-data-warehousing>“ (pristupljeno: 14.07.2017.)

Bill Inmon je definirao skladište podataka na sljedeći način: „*skladište podataka je subjektno-orijentirana, integrirana, o vremenu ovisna, neizbrisiva kolekcija podataka s namjerom da pomaže menadžmentu u procesu donošenja odluka*“.³¹

Bill Inmon je težio velikim centraliziranim skladištima podataka, a vodio se pristupom „*top-down*“, pristup koji je vidljiv na slici 3.

No osim Bill Inmona, skladištenjem podataka bavio se i Ralph Kimball, kojeg se još smatra i drugim „ocem“ skladištenja podataka.

Dok je Inmonova knjiga pružala jaku teoretsku podlogu za koncepte *Data Warehousea*, Ralph Kimball je objavio knjigu pod nazivom *The Data Warehouse Toolkit*, 1996. godine, u kojoj su bili opisani praktični primjeri za OLAP modeliranje. No, kada se radilo na velikim projektima *Data Warehouse-a*, za tvrtke koje su pružale financijske usluge u kasnim 90-ima uvelike je korištena knjiga Ralpha Kimballa. Njegova knjiga izdana je u više izdanja, a fokus je bio na web-baziranim skladištima podataka, ETL u okruženju skladišta podataka, kao i posebna izdanja koja pokrivaju *Microsoft SQL Server* i *Microsoft Business Intelligence Toolset*.



Slika 3 Bottom-Up i Top-Down pristup skladištenju podataka³²

³¹ B. Jiang, 'Is Inmon's Data Warehouse Definition Still Accurate?', *BeyeNETWORK*, 2012., „<http://www.b-eye-network.com/view/16066>“ (pristupljeno: 14.07.2017.)

³² P. D. Elias, *A abordagem Top-Down e Bottom-up no Data Warehouse*, [online slika], 12.05.2014., <https://canaltech.com.br/infra/a-abordagem-top-down-e-bottom-up-no-data-warehouse-21108/> (pristupljeno: 12.08.2017.)

Ralph Kimbal služio se „bottom up“ pristupom, vidljivim na slici 3, koji je usklađen sa njegovom preferencijom modeliranja *star-shemom*. Kimball se svojim pristupom fokusirao na spremišta podataka, koja manje tvrtke nalaze pogodnijima jer ne opterećuju budžet u tolikoj mjeri koliko skladišta podataka. Inmonov i Kimballov pristup i dalje su srž arhitekture skladištenja podataka i do današnjeg dana.³³

Spremišta podataka su zapravo manja skladišta podataka, i postoji mogućnost da tvrtke koriste pristup više spremišta podataka, koja kasnije mogu spojiti u jedno veliko skladište podataka. Spremišta podataka su specijalizirana mjesta za pohranu podataka. Primjerice, ako imamo tvrtku koja pruža konzultantske usluge, tvrtka može uzeti jedno spremište podataka za računovodstvo i jedno za podatke o klijentima. Dakle nepotrebno je trošiti resurse na veliko skladište podataka ukoliko nema potrebe za tim. Dok s druge strane imamo veliko proizvodno poduzeće, tada će oni trebati razmisliti hoće li koristiti veliko skladište podataka ili će ići pristupom više manjih spremišta podataka. Iako je preporučljivo da velike tvrtke koriste skladišta podataka zbog veličine i količine podataka.

3 Paradigma semantičkog weba i web inteligencija

Do sada smo govorili o razvoju i funkcioniranju interneta gdje smo spomenuli i *world wide web* i njegova tvorca Tim Berners Leeja. Upravo je on začetnik ideje za razvoj semantičkog weba. Pri tome semantički web ne vidi samo kao zaseban *web*, već kao proširenje postojećeg. Ideja je bila stvoriti *web* koji ne samo da povezuje dokumente, već prepoznaje značenje informacija u tim dokumentima, drugim riječima, da računalo prepoznaje i povezuje nešto što ljudi lako povezuju kako bi kooperacija između ljudi i računala bila bolja.³⁴

Tim Berners-Lee kaže: „*The Semantic Web is really data that is procesable by machine*“³⁵

Ovaj citat tvorca semantičkog weba, zapravo najbolje opisuje što je semantički web. On je želio postići da računalo razumije i samo povezuje informacije.

U svom radu, Tim Berners-Lee objašnjava kako bi semantički web trebao funkcionirati i to

³³ P. Williams, *A Short History of Data Warehouseing*, [web stranica], 2012., „<http://www.dataversity.net/a-short-history-of-data-warehousing>“ (pristupljeno: 14.07.2017.)

³⁴ M. Frauenfelder, *A Smarter Web*, [web stranica], 2001., <https://www.technologyreview.com/s/401241/a-smarter-web/> (pristupljeno: 01.08.2017.)

³⁵ S. Hardin, 'Tim Berners-Lee: The semantic Web-Web of machine processable data', *Bulletin of the American Society for Information Science and Technology*, vol. 31, no. 3, 2005., p. 13

korištenjem sustava za reprezentaciju znanja, ontologije³⁶, korištenjem agenata te stalne evolucije znanja čime je postavio dio osnova za web inteligenciju.

Već smo spomenuli kako je web inteligencija multidisciplinarno područje, te kao takvo u sebi sadrži i elemente semantičkog weba. Jedno od temeljnih pitanja web inteligencije je proučavanje semantike weba, koje se naziva semantički web. Semantički web modelira semantiku web informacija kako bi se omogućilo:

- da više sadržaja na web-u postane razumljivo i obradivo računalu (eng. *machine readable and processible*)
- prepoznavanje semantičkog konteksta u korištenju web materijala
- usuglašavanje terminoloških razlika između drugačijih zajednica korisnika

Informacije će biti strojno obradive u načinima da podržavaju inteligentne mrežne servise kao što su informacijski brokeri i agenti za pretraživanje.³⁷

U nastavku rada fokusirat ćemo se na arhitekturu semantičkog weba, standarde za semantički web postavljene od strane W3C, te ćemo spomenuti neke primjere korištenja semantičkog weba u praksi.

3.1 Arhitektura semantičkog weba

Dakle, već smo rekli kako je semantički web vizija weba u kojem informacije imaju smisla i značenja, te ih računalno može automatski procesirati. Na slici 4 prikazana je osnovna arhitektura semantičkog web-a.

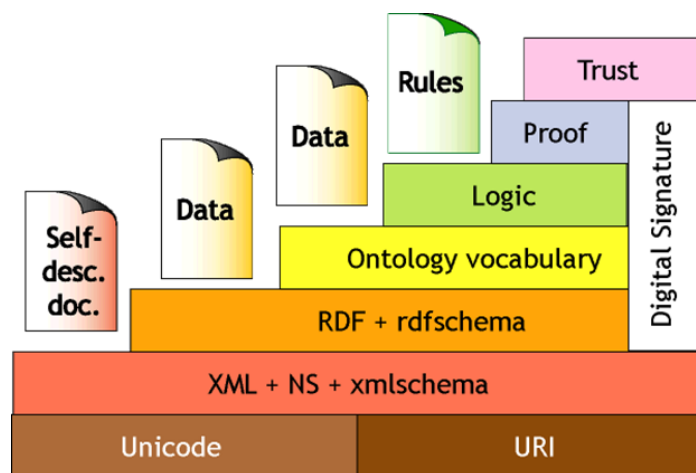
Semantički web se gradi na mogućnostima XML-a u definiranju prilagođenih shema za označavanje i *Resource Description Framework* (skraćeno RDF) fleksibilnosti u prikazivanju podataka. U semantičkom Web-u prvu razinu iznad RDF-a čini ontologijski jezik OWL (eng.

³⁶ Ontologija (eng. ontology) – u kontekstu računalnih i informacijskih znanosti, ontologija definira skup primitivnih reprezentacijskih podataka s kojima se modelira domena znanja ili diskurs. T. Gruber, 'Ontology', u L. Liu i M. Tamer Ozsu (eds.), *Encyclopedia of Database Systems*, 2009., <http://tomgruber.org/writing/ontology-definition-2007.htm> (pristupljeno: 15.09.2017.)

³⁷ N. Zhong, J. Liu i Y. Yao, 'Web Intelligence', *Computer Software and Applications Conference, 2000. COMPSAC 2000. The 24th Annual International*, 2000., p. 5

Web Ontology Language) kojim se opisuje značenje termina korištenih u web dokumentima.³⁸ Kasnije ćemo detaljnije objasniti OWL, a u nastavku osvrnut ćemo se na funkcioniranje komponenti semantičkog weba i veze među njima. Kao što vidimo na slici 4, semantički web sastoji se od slojeva (eng. *Layers*).

XML, koji smo već ranije pojasnili u prethodnom poglavlju rada, pruža sintaksu za strukturirane dokumente, ali ne nameće semantička ograničenja na značenja ovih dokumenata. XML shema se odnosi na osnovne grafične elemente strukture XML dokumenta i mogućnosti proširenja XML-a sa potrebnim tipovima podataka. RDF je podatkovni model za objekte i veze među njima. Pruža jednostavnu semantiku za ovaj podatkovni model, a modeli mogu biti prikazani kao XML sintaksa. RDF shema je rječnik za opisivanje svojstava i klasa RDF resursa, sa semantikom za generalizaciju hijerarhija takvih svojstava i klasa. OWL nadopunjuje rječnik za opisivanje svojstava i klasa: veze između klasa kardinalnosti, jednakost, bogatiji tipovi svojstava, karakteristike svojstava i nabrojane klase.³⁹



Slika 4 Arhitektura semantičkog weba⁴⁰

Sloj za logiku se sastoji od pravila koji omogućuju zaključke. Sloj za dokaze (eng. *Proof layer*) potreban je da bi pružio objašnjenja o odgovorima od strane automatiziranih agenata koji

³⁸ D. L. McGuinness i F. van Harmelen, *OWL Web Ontology Language Overview*, [web stranica], 2004., <https://www.w3.org/TR/owl-features/> (pristupljeno: 03.08.2017.)

³⁹ D. L. McGuinness i F. van Harmelen, loc. cit.

⁴⁰ T. Berners-Lee, *Architecture of Semantic Web*, [online slika], 2000., <https://www.w3.org/2000/Talks/1206-xml2k-tbl/sweb-stack.gif> (pristupljeno: 03.08.2017.)

koriste pružene informacije. Logično je da želimo provjeriti rezultate proizašle od agenata, to zahtjeva prijevod internog mehanizma razmišljanja u jedinstveni jezik za prikaz dokaza. Slojevi za dokaz i povjerenje (eng. *Trust layer*) su još u fazi razvoja, a u trenutnoj fazi razvoja semantičkog weba i dalje predstavljaju problem na kojem se još nije počelo raditi.⁴¹

3.2 Standardi semantičkog weba

Semantički web je vizija W3C-a da stvori takozvani web podataka (eng. *Web of Data*). Kako bi to bilo moguće stvorene su tehnologije za semantički web koje omogućuju ljudima da stvore mjesta za pohranu podataka na webu, naprave rječnike, i pišu pravila za rukovanje podacima.⁴² Tehnologije za semantički web nazivaju se još i standardima semantičkog weba, a one uključuju: RDF, OWL, RDFa, RDFS, SPARQL, SKOS, JSON-LD, GRDDL, RDB2RDF, POWDER, PROV, SAWSDL, RIF.

Resource Description Framework (skr. RDF) je temelj za obradu metapodataka. RDF pruža interoperabilnost između aplikacija koje razmjenjuju računalu razumljive informacije na webu. Općenitije, RDF definira pristupe i metode za opis različitih resursa bez pretpostavki o pojedinačnoj domeni aplikacije, te oslobađa svu semantiku od veze s domenom aplikacije. Time se postiže neutralni mehanizam neovisan o domeni, dok istovremeno mehanizam mora biti prilagođen za opisivanje informacija o bilo kojoj domeni.⁴³

Web Ontology Language (OWL) je jezik dizajniran za semantički web u svrhu reprezentiranja bogatog i kompleksnog znanja o stvarima, grupama stvari i vezama između stvari. Pod stvarima podrazumijevamo pristupe, stilove *software*-ske arhitekture i uzorke programiranja koji omogućuju stvarnim objektima da budu dio weba. OWL je računalni jezik temeljen na logici na način da znanje iskazano u OWL-u može biti korišteno od strane računalnih programa.⁴⁴

RDFa je specifikacija za atribut kako bi prikazivali strukturirane podatke u HTML5, XHTML i u bilo kojoj XML aplikaciji. Alati mogu izvlačiti strukturirane podatke i generirati RDF trojke

⁴¹ 'The Semantic Web – A Web of Metadata', *Protopogenist Blog*, [web blog], 2012., <https://protopogenist.wordpress.com/tag/proof-trust-layer/> (pristupljeno: 03.08.2017.)

⁴² *W3C Semantic Web – Main Page*, [web stranica], 2014., https://www.w3.org/2001/sw/wiki/Main_Page (pristupljeno: 15.08.2017.)

⁴³ O. Lassila i R. R. Swick, *Resource Description Framework (RDF) Model and Syntax Specification*, W3C Proposed Recommendation, [web stranica], 1991., „<https://www.w3.org/TR/PR-rdf-syntax>“ (pristupljeno: 15.08.2017.)

⁴⁴ *Web Ontology Language (OWL)*, [web stranica], 2012., <https://www.w3.org/2001/sw/wiki/OWL> (pristupljeno: 17.08.2017.)

(eng. *Triples*) za daljnju ponovnu upotrebu. RDFa je još jedan u nizu serijalizacijskih formata za RDF.⁴⁵

RDF Schema (RDFS) je jezik opće namjene za reprezentiranje jednostavnih RDF rječnika na web-u. Ostale tehnologije poput OWL-a ili SKOS-a su nadogradnja na RDFS i pružaju jezik za definiranje strukturiranih, web temeljenih ontologija koje omogućuju bogatiju integraciju i interoperabilnost podataka kroz opisne zajednice⁴⁶

SPARQL (RDF Query Language) je upitni jezik za RDF. Koristi se za postavljanje upita kroz različite izvore podataka, bez obzira jesu li podaci pohranjeni kao RDF ili se pregledavaju kao RDF preko *middleware*-a. SPARQL ima mogućnost propitivanja potrebnih i opcionalnih grafikonskih uzoraka sa njihovim konjunkcijama i disjunkcijama. Trenutna verzija je SPARQL 1.1.⁴⁷

SKOS (eng. *Simple Knowledge Organization System*) je standardizirani sustav organizacije znanja za njegovo dijeljenje, povezivanje i klasificiranje putem web-a⁴⁸

JSON Serialization for Linked Data (JSON-LD) sintaksa je dizajnirana kako bi se lakše integrirala u već postojeće sisteme koji koriste JSON i pruža nadogradnju za JSON na JSON-LD. Namijenjena je da bude način korištenja povezanih podataka (eng. *Linked data*) u web baziranim aplikacijama za programiranje, za pravljenje interoperabilnih Web servisa i da pohranjuje povezane podatke u JSON bazirano skladište.⁴⁹

Gleaning Resource Description from Dialects of Languages (GRDDL) je tehnika prikupljanja RDF podataka iz XML dokumenata i iz pojedinih XHTML stranica.⁵⁰

Relational Databases to RDF (RDB2RDF) sastoji se od 2 jezika za mapiranje sadržaja relacijskih baza podataka u RDF. To su *Direct Mapping* i *R2RML*. Jezici se mogu koristiti za prevođenje relacijskih podataka u RDF, ovaj se proces nekada naziva i ETL proces.⁵¹

Protocol for Web Description Resources (POWDER) pruža mehanizam za opisivanje i

⁴⁵ *RDF in Attributes (RDFa)*, [web stranica], 2013., „<https://www.w3.org/2001/sw/wiki/RDFa>“ (pristupljeno: 17.08.2017.), 2004., <https://www.w3.org/2001/sw/wiki/RDFS> (pristupljeno: 17.08.2017.)

⁴⁶ *RDF Vocabulary Description Language 1.0: RDF Schema (RDFS)*, [web stranica],

⁴⁷ *SPARQL Query Language for RDF*, [web stranica], 2013., „<https://www.w3.org/2001/sw/wiki/SPARQL>“ (pristupljeno: 17.08.2017.)

⁴⁸ SKOS [web stranica], 2009., <https://www.w3.org/2001/sw/wiki/SKOS> (pristupljeno: 17.08.2017.)

⁴⁹ *A JSON-based Serialization for Linked Data (JSON-LD)*, [web stranica], 2014., „<https://www.w3.org/2001/sw/wiki/JSON-LD>“ (pristupljeno: 17.08.2017.)

⁵⁰ *Gleaning Resource Description from Dialects of Languages (GRDDL)*, [web stranica], 2007., <https://www.w3.org/2001/sw/wiki/GRDDL> (pristupljeno: 17.08.2017.)

⁵¹ *Relational Databases to RDF (RDB2RDF)*, [web stranica], 2012., <https://www.w3.org/2001/sw/wiki/RDB2RDF> (pristupljeno: 18.07.2017.)

otkrivanje web resursa, te pruža sažet način definiranja bilo kojeg broja predikata za te resurse. Dizajniran je u svrhu pojednostavljivanja inkorporacije informacija u veće RDF sisteme.⁵² PROV je specifikacija koja omogućuje rječniku da razmjenjuje podrijetlo informacija.⁵³ *Rule Interchange Format* (RIF) je format koji ima za cilj definiranje standarda za razmjenu pravila između sistema pravila, posebice web pravila. RIF se fokusira na izmjenjivanje više nego na definiranje jedinstvenog jezika koji odgovara svim pravilima. Pristup koji koristi RIF *Working Group* je dizajniranje „obitelji“ jezika koji se nazivaju dijalekti, s rigorozno specificiranim sintaksama i semantikom. Namjera je da RIF dijalekti budu ujednačeni i proširivi.⁵⁴

3.3 Primjeri web inteligencije zasnovane na paradigmi semantičkog weba

Do sada smo semantički web gledali samo kroz teorijski okvir, kroz njegovu arhitekturu i postavljene standarde, no da ne bismo ostali samo pri šturoj teoriji, okrenut ćemo se i nekim uspješnim primjerima iz prakse. Među mnogim primjerima uspješne implementacije tehnologije semantičkog weba, spomenut ćemo tri slučaja bliža ekonomskoj praksi. Područja primjene u primjerima su turizam, automobilska industrija i financijski sektor.

3.3.1 Inteligentni pretraživač online usluga javne uprave grada Zaragoze

Prvi slučaj je implementiranje načina inteligentnog pretraživanja za online usluge administrativne uprave grada Zaragoze. Ispostavilo se da kada građanin traži određenu uslugu, da je nije baš lako pronaći. Mnoge gradske web stranice samo numeriraju dostupne usluge, i prikazuju ih kao liste organizirane prema kategorijama, dok neke druge stranice nude tradicionalan pretraživač koji izbacuje rezultate prema upitu i opisu usluge. Problem je i što jezik korišten od strane administracije nije isti kao jezik koji koriste građani.

Grad Zaragoza, riješila je probleme primjenom tehnologije semantičkog web-a koja pronalazi online usluge s obzirom na značenje upita i opisa usluge. Značenje usluge definirano je kroz ontologiju za javnu administraciju, što znači da osoba koja posjeduje auto mora plaćati porez,

⁵² *Protocol for Web Description Resources (POWDER)*, [web stranica], 2009., „<https://www.w3.org/2001/sw/wiki/POWDER>“ (pristupljeno: 17.08.2017.)

⁵³ *PROV*, [web stranica], 2013., „<https://www.w3.org/2001/sw/wiki/PROV>“ (pristupljeno: 17.08.2017.)

⁵⁴ *Rule Interchange Format (RIF)*, [web stranica], 2010., „<https://www.w3.org/2001/sw/wiki/RIF>“ (pristupljeno: 17.08.2017.)

da neki obrti mogu uzrokovati buku (poput kafića ili diskoteka) i tome slično.

Ontologija se razlikuje između različitih vrsta agenata, događaja, objekata, procesa, skupa s njihovim međusobnim vezama.

Kada građanin postavi upit u tražilicu, kombinacija procesiranja prirodnog jezika i ontologijskog razumijevanja se koristi kako bi se proširio upit. Ista tehnologija se primjenjuje i na opis servisa, te sve rezultira obogaćenim indeksom.

Neke od ključnih prednosti koje je implementacijom stekao grad Zaragoza uključuju: pružanje boljih usluga građanima, stimuliranje e-uprave pružanjem lakšeg pristupa i smanjenje troškova smanjenjem opterećenja na pozivne centre i fizičke urede. No, nije samo grad Zaragoza profitirao od implementacije, već su neke prednosti stekli i građani, a to su: jednostavna interakcija kroz ključne riječi i prirodni jezik, visoka preciznost, konkretni odgovori umjesto dugih lista i pretraživač koji „razumije“ građanina.⁵⁵

3.3.2 Kontekstualno pretraživanje - Volkswagen

Još jedan slučaj uspješne implementacije semantičkog weba u, okviru web inteligencije, je i veliki koncern automobila Volkswagen. Volkswagen je želio da njihovi kupci brže dolaze do informacija koje su im potrebne, stoga su pokrenuli projekt kontekstualnog pretraživanja. Ovaj projekt označava uspjeh Volkswagenovog sadržaja i podatkovne strategije, te reprezentira tehničku inovaciju u svom pravom obliku i promjenu paradigme u načinu na koji Volkswagen koristi i dijeli podatke preko web-a. Ključne prednosti korištenja semantičke tehnologije u Volkswagenu su: standardizirano sučelje podataka i sadržaja, dostupno developerima različitih vještina, korištenjem različitih tehnologija unutar i izvan organizacije, odvajanje interesa između informacija i aplikacija kako logički, tako i fizički, povećanje vrijednosti, iskoristivosti i pristupa podacima, te sama adaptacija tehnologija nije zahtijevala promjenu procesa ili promjenu u menadžmentu.⁵⁶

⁵⁵ J. F. Ruiz, *Case Study: An Intelligent Search Engine for Online Services for Public Administrations*, [web stranica], June 2007., „<https://www.w3.org/2001/sw/sweo/public/UseCases/Zaragoza>“ (pristupljeno: 17.08.2017.)

⁵⁶ W. Greenly et. al., *Case study: Contextual Search for Volkswagen and the Automotive Industry*, [web stranica], October 2011., <https://www.w3.org/2001/sw/sweo/public/UseCases/Volkswagen/> (pristupljeno: 17.08.2017.)

3.3.3 Predlaganje povezanih ideja u stvarnom vremenu - financijski sektor

Treći slučaj kojeg smo izdvojili, je primjena na financijsku industriju. Financijska industrija postala je iznimno konkurentna, jer se mnogo banki natječu kako bi dobili svoj dio „kolača“. Konkurencija vrši snažan pritisak na stvaranje novih financijskih proizvoda i usluga, a proizvode i usluge koji se pokazuju uspješnima, konkurencija vrlo brzo kopira. Ovaj slučaj odnosi se na financijsku industriju u Španjolskoj, konkretnije na petu najveću banku u zemlji, Bankinter. Tokom godina, u Bankinteru se pojavilo na desetke tisuća ideja od strane zaposlenika. Evaluacija svake ideje postala je velik posao za odjel ljudskih resursa (eng. *Human Resource Management - HRM*). Osim što je potreban ogroman trud, evaluacija ideja predstavlja i trošak, a teško je i otkriti sličnosti i napraviti analizu informacija (najvrjedniji zaposlenik, socijalno umrežavanje, suradnja između odjela i slično).

Bankinter je rješenje pronašao u implementaciji sistema temeljenog na semantičkom tehnologiji koja pomaže HRM-u upravljati s idejama. Kada zaposlenik unese novu ideju, sistem analizira tekst i prepoznaje relevantne koncepte s financijske perspektive. Sve se odvija u stvarnom vremenu, omogućavajući tako korisniku da vidi i druge ideje koje sadrže iste koncepte (ali ne i riječi). Ovakav pristup zaposlenicima pruža mogućnost uvida u inovativnost njihove ideje, ali i predloženih ideja ostalih zaposlenika. Koncepti su definirani i povezani u financijskoj ontologiji, koja uključuje proizvode, kanale, odjele, klijente i tome slično. Na isti način, zaposlenici mogu pretraživati ideje koje sadrže relevantne koncepte, a takvi koncepti su označeni tako da ih zaposlenik može vidjeti u istom trenutku i odmah dobiti povratnu informaciju. Pretraživanje i analiza se može vršiti prema datumima, individualnom zaposleniku, odjelu i slično. Sustav ima mogućnost davanja objašnjenja zašto misli da su određene ideje slične, prikazom semantičkih veza između koncepta nove ideje i postojeće.

Prednosti korištenja semantičkih tehnologija u Bankinteru su bolje upravljanje inovacijama, što dovodi do skraćivanja vremena njihovog stavljanja na tržište, upravljanje talentima, te povećanja zadovoljstva zaposlenika.⁵⁷

⁵⁷ J. L. Bas Uribe, *Case Study: Real Time Suggestion of Related Ideas in Financial Industry*, [web stranica], May 2007., „<https://www.w3.org/2001/sw/sweo/public/UseCases/Bankinter>“ (pristupljeno: 17.08.2017.)

3.3.4 Web aplikacija CRUZAR za personalizirane rute

Zadnji slučaj kojeg smo izdvojili, odnosi se na uporabu tehnologija semantičkog weba na turizam. Prema podacima Svjetske turističke organizacije (eng. *World Tourism Organization – UNWTO*) turizam je posljednje desetljeće doživio kontinuirani razvoj i postao jedan od najbrže rastućih ekonomskih sektora na svijetu. Prema UNWTO-u, svjetski turizam porastao je 2015. godine za 4,6% na 1.184 milijuna dolazaka turista, dok je 2016. turizam porastao za još 3,9% na i dosegao 1.235 milijuna dolazaka.⁵⁸ Turizam je postao velika i profitabilna industrija. Grad Zaragoza odlučio se za svoje turiste napraviti aplikaciju koja bi za svakog turista pravila personalizirane rute posjeta gradu i njegovim znamenitostima. Aplikaciji su dali ime CRUZAR, a napravljena je korištenjem tehnologija semantičkog weba. CRUZAR je web aplikacija koja koristi ekspertno znanje (u formi pravila i ontologija) i repozitorij relevantnih podataka za pravljenje personaliziranih ruta prema profilu svakog posjetitelja. Dok većina gradova ima predefinirane rute, CRUZAR generira bezbroj personaliziranih ruta i tako pronalazi odgovarajuće rute za svaki profil posjetitelja. Aplikacija sadrži točke interesa (eng. *Points of Interest*), veliku bazu podataka i dostupna je u 5 jezika.

Tehnologije semantičkog weba u aplikaciji koriste se za integriranje i organiziranje podataka iz različitih izvora, reprezentiranje i transformiranje korisničkih profila i turističkih resursa, te za dohvaćanje svih informacija o generiranim rutama.⁵⁹

Izdvojili smo ova 4 primjera kao bližima ekonomskoj praksi, no primjera je zaista mnogo, a tehnologije semantičkog weba primjenjuju se još i u zdravstvu⁶⁰, edukaciji⁶¹, energetici⁶² i tako dalje.

⁵⁸ *Why Tourism?*, [web stranica], 2017., <http://www2.unwto.org/content/why-tourism> (pristupljeno: 20.09.2017.)

⁵⁹ M. J. Fernandez, *Case Study: CRUZAR – An application of semantic matchmaking for eTourism in the city of Zaragoza*, [web stranica], August 2008., „<https://www.w3.org/2001/sw/sweo/public/UseCases/Zaragoza-2/>“ (pristupljeno: 25.08.2017.)

⁶⁰ K. Cheung, et. al., 'Semantic Web for Health Care and Life Sciences: a review of the state of the art', *Briefings in Bioinformatics*, vol. 10, no. 2, 2009., „<https://academic.oup.com/bib/article/10/2/111/184475/Semantic-Web-for-Health-Care-and-Life-Sciences-a>“ (pristupljeno: 15.09.2017.)

⁶¹ R. Koper, 'Use of the semantic Web to Solve Some Basic Problem sin Education: Increase Flexible, Distrubuted Lifelong Learning, Decrease Teacher's Workload', *Journal of Interactive Media in Education*, vol. 6, 2004., „<https://www-jime.open.ac.uk/articles/10.5334/2004-6-koper/>“ (pristupljeno: 15.09.2017.)

⁶² K. Wenzel et. al., *Semantic Web Based Dynamic Energy Analysis and Forecasts in Manufacturing Engineering*, Fraunhofer Institute for Machine Tools and Forming Technology IWU, „https://www.greencarbody.de/csdata/download/1/de/2011_02_05_cirp_semantic_web_207.pdf“ (pristupljeno: 15.09.2017.)

4 Unaprjeđenje web sjedišta primjenom web inteligencije

U novije vrijeme postojanje web sjedišta poslovnih subjekata predstavlja neizostavan dio u gotovo svakom području poslovanja. Na svojim web sjedištima tvrtke objavljuju informacije o proizvodima, uslugama, nude mogućnost podrške i slično. To je jedan od najjednostavnijih načina pružanja informacija postojećim, a i potencijalnim klijentima. Shodno tome, uopće nije potrebno govoriti koliku važnost web sjedište ima za poslovanje, a pogotovo uzmemo li u obzir da neke tvrtke poput Amazona, eBaya, AliExpressa, posluju samo preko interneta. Mogućnost prikupljanja podataka, odnosno rudarenje podataka, jedna je od temeljnih načina unaprjeđenja web sjedišta. U ovom poglavlju govorit ćemo o tome kako korištenjem web inteligencije unaprijediti web sjedište tvrtke. Pod unaprjeđenje korištenjem web inteligencije spada inteligentan dizajn web sjedišta, dinamički web te sigurnost, privatnost i integritet podataka koju pruža web inteligencija. Spomenut ćemo i nekoliko uspješnih implementacija u poslovanju.

4.1 Inteligentni dizajn web sjedišta

Inteligentni dizajn web sjedišta ogleda se kroz jednostavnost i preglednost (eng. *usability*) te pristupačnost. Jednostavnost i preglednost, jednom riječju iskoristivost, je opseg u kojem se proizvod, u ovom slučaju web sjedište, može koristiti od strane različitih korisnika kako bi se ostvarili specifični ciljevi sa efektivnošću, efikasnosti i zadovoljstvom u kontekstu korištenja. S druge strane, pristupačnost web-a je atribut kojim ljudi sa raznim poteškoćama mogu spoznati, razumjeti, upravljati i ostvariti interakciju sa web-om, i kako mogu pridonijeti web-u. Pristupačnost je usmjerena na eliminiranje ograničenja kojima se susreću ljudi sa bilo kakvom vrstom poteškoća, uključujući vizualne, auditorne, fizičke, govorne, kognitivne i neurološke poteškoće, osim toga pristupačnost nastoji omogućiti da web sadržaj bude dostupan svima.

Kada govorimo o web-u, iskoristivost (eng. *usability*) se klasificira kao kvalitativan atribut i definira se kroz pet kvalitativnih komponenti:

- 1) lakoća učenja – lakoća s kojom korisnik koji se prvi put služi stranicom može isprobati sve osnovne funkcionalnosti dizajna

- 2) efikasnost – brzina s kojom se korisnik može obavljati svoje zadatke nakon što se privikne na dizajn
- 3) pamtljivost – lakoća s kojom se korisnik može prisjetiti svojih vještina korištenja web sjedišta
- 4) greške – učestalost, ozbiljnost i lakoća oporavljanja od korisničkih grešaka
- 5) zadovoljstvo – uživanje u korištenju dizajna sjedišta⁶³

Dakle, vrlo je važno imati web sjedište koje je jednostavno, pregledno i prije svega pristupačno. Osim toga, nedostatak bilo kojeg segmenta može izazvati zbunjenost i frustraciju kod korisnika, te mijenjanje jednog web sjedišta za neko konkurentnije.

4.2 Dinamičko web sjedište

Dinamičko web sjedište je bilo koje web sjedište koje reagira na ljudske aktivnosti. Aktivnosti spojene s uvjetima koji se trenutno podudaraju će odrediti koje aktivnosti će poduzeti web sjedište. Ponekad će ovo rezultirati tako da server pošalje drugu stranicu dok će u nekom drugom slučaju prikazati animaciju na trenutnoj stranici. Ovo može biti kontradiktorno jer svaki hiperlink može ostvariti jednaki efekt, no glavna razlika je u tome da se odluke o tome što će se prikazati, donositi dinamično.⁶⁴

Dinamičke web stranice generiraju se putem *software*-skih sistema poput *JavaScripta*, *CGI-a*, *ASP-a* i *Servleta*.

JavaScript je *client-side*⁶⁵ tehnologija koja omogućava korisniku da pravi dinamičke web stranice, dodaje funkcije za upravljanje događajima (eng. *events*) i manipulira objektima dokumenta dok god internetski preglednik podržava *Document Objekt Model* (DOM), koji je važan dio DHTML-a.

CGI (eng. *Common Gateway Interface*) je sučelje koje definira kako se podaci prenose i služi za spajanje web servera na druge *software-e*. *CGI* program može biti kodiran u programskih jezicima *Perl*, *ShellScript*, *C++* ili *Java*. Bez obzira na programski jezik, mehanizam koji dijele je isti. Ako web server podržava *CGI*, obrazac podataka je poslan *CGI* programu kroz parametre

⁶³ A. Dingli i S. Cassar, 'An Intelligent Framework for Website Usability', *Advances in Human-Computer Interaction*, vol. 2014, 2014., p. 1-2

⁶⁴ B. Polzar, *Dynamic Web Pages*, University of Wisconsin, 2007., p. 3

⁶⁵ *Client-side* – software-i za programiranje koji imaju korisničko sučelje s kojim korisnik ima interakciju.

programskog okruženja. Nakon što CGI program završi za obradom podataka, njegov izlaz prema konzoli je preusmjeren nazad na internetski preglednik od strane web servera. CGI sučelje je vrlo jednostavno, no iako je jednostavna tehnologija, ona daje nevjerojatno moćno sučelje koje *JavaScript* ne može dati.

ASP (eng. *Active Server Page*) je program koji radi unutar *Microsoft IIS Server*-a. Kada internetski preglednik da zahtjev za ASP datotekom, IIS prenosi zahtjev *ASP engine*⁶⁶. *ASP engine* čita ASP datoteku, red po red, i izvršava skriptu iz datoteke. Konačno, ASP datoteka se vraća u internetski preglednik kao običan HTML. Izvršava se putem *ASP engine*-a u „nitima“. Zato je ASP puno brži i efikasniji nego CGI.

JSP (eng. *Java Server Page*) i *Servlet* su *Java* tehnologije koje su podržane u *Java 2 Enterprise Edition*. JSP je vrlo sličan ASP tehnologiji. Njegova prednost naspram ASP-a je ta da se može prenositi na bilo koje računalo dok god je *Java Virtual Machine* (JVM) dostupan. JSP datoteka je HTML datoteka ugrađena u neke *Java snippet*-e. U *Java Snippet*-u se realizira snaga JDK-a (*Java Development Kit*), koji ima bolje mogućnosti nego ASP.

Servlet je pisan u *Java*-i, odnosno, JSP datoteka je interpretirana kao *Servlet* pomoću JSP *engine*-a prije negoli je izvršena od strane *Servlet engine*-a.⁶⁷

4.3 Primjeri unaprjeđenja web sjedišta tvrtke pomoću web inteligencije

U primjerima koji slijede vidjet ćemo tvrtke koje su unaprijedile svoja web sjedišta primjenom web inteligencije.

4.3.1 Netflix

Popularni online servis za gledanje filmova i tv serija koristi web inteligenciju kako bi personalizirao prijedloge filmova ili tv serija za svakog korisnika posebno. Kada korisnik doda nešto na svoju listu „gledanje kasnije“, Netflix predlaže sličan sadržaj ovisno o tome što je

⁶⁶ *Engine* – žargon koji developeri upotrebljavaju kako bi opisali dio *software*-a koji često upotrebljavaju

⁶⁷ K. Yi, 'Event driven dynamic web pages', Dissertation thesis, Rochester Institute of Technology, 2006., p. 7-12

dodano na listu i o tome kako je korisnik u prošlosti ocjenjivao sličan sadržaj. Netflix je prepoznao trend personalizacije i smatra da je generički sadržaj stvar prošlosti.⁶⁸

4.3.2 Facebook

Najveća društvena mreža, bolje rečeno online zajednica, na svijetu Facebook također je prepoznala prednosti korištenja web inteligencije. Facebook pristupa personaliziranoj prodaji sa dinamičkim sadržajem sa perspektive društvenog oglašavanja. Omogućava ciljane oglase kroz naprednu demografiju, dajući tako oglašivačima mogućnost istančavanja i ciljanja točno određene skupine.⁶⁹

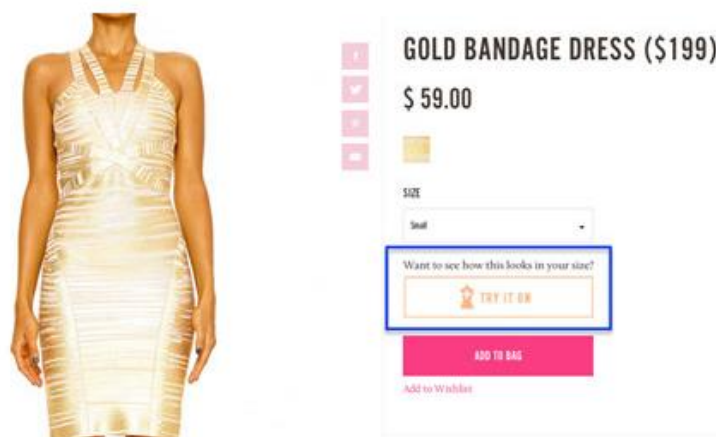
Primjerice, otvarate kopirnicu ciljano za studente, i na želite se oglašavati na Facebooku. Možete dati oglas na Facebooku i točno odrediti da vas zanima ciljna skupna sa godinama između 18 i 25, koji žive u Zadru, u Republici Hrvatskoj i da studiraju na Sveučilištu u Zadru. Facebook će izračunati točno koliki je doseg pojedinih Facebook korisnika kojima će se dani oglas prikazati.

4.3.3 Fits.me Virtual Dressing Room

Mnogi e-dućani nakon što provedu marketinšku kampanju i razna oglašavanja, dođu do točno određenih kupaca koji bi željeli isprobati robu koja se prodaje, a posao e-dućana je uvjeriti kupca da će im roba pristajati. Online kupci često odustaju od kupnja zbog nemogućnosti isprobavanja robe koju žele kupiti. Fits.me cilja na rješenje tog problema omogućujući online prodavačima da pokažu potencijalnim kupcima kako bi oni mogli izgledati, virtualno, u odabranom komadu robe koji je na njihovoj stranici.

⁶⁸ S. Bell, *5 Websites that Successfully Leverage Dynamic Content Experiences to Increase Sales*, [web stranica], 2015., <https://www.cdnetworks.com/en/news/5-websites-that-successfully-leverage-dynamic-content-experiences-to-increase-sales/4226> (pristupljeno: 15.09.2017.)

⁶⁹ S. Bell, *5 Websites that Successfully Leverage Dynamic Content Experiences to Increase Sales*, [web stranica], 2015., <https://www.cdnetworks.com/en/news/5-websites-that-successfully-leverage-dynamic-content-experiences-to-increase-sales/4226> (pristupljeno: 15.09.2017.)



Slika 5 Odabir artikla i gumb koji pokreće aplikaciju Fits.me (označen u plavom kvadratu)⁷⁰

Ovu aplikaciju prepoznala je tvrtka *Tamar Collection*. U prvom koraku aplikacija nudi odabir artikla koji vas zanima kojeg korisnik može isprobati klikom na odgovarajuće dugme – slika 5. Potom se otvara prozor za upisivanje vaših mjera (visina, struk, kukovi i sl. – slika 6), nakon toga odabire se veličina koja se najbolje podudara sa vašim mjerama, te se prikazuje mogući izgled u tom određenom komadu odjeće, kao što vidimo na slici 7.⁷¹



Slika 6 Sučelje aplikacije za unos osobnih dimenzija⁷²

⁷⁰ S. Bell, op. cit.

⁷¹ S. Bell, *5 Websites that Successfully Leverage Dynamic Content Experiences to Increase Sales*, [web stranica], 2015., „<https://www.cdnetworks.com/en/news/5-websites-that-successfully-leverage-dynamic-content-experiences-to-increase-sales/4226>“ (pristupljeno: 15.09.2017.)

⁷² op. cit.



Slika 7 Slika mogućeg izgleda u odabranom artiklu prema unesenim dimenzijama⁷³

⁷³ S. Bell, ibid.

5 Web inteligencija zasnovana na rudarenju podataka

U prethodnom poglavlju govorili smo o semantičkom webu i kako je on zapravo web u kojem informacije imaju smisla i značenja, te ih računalo može samo obrađivati. Informacije su obrađeni, organizirani, strukturirani ili prezentirani podaci, dok su podaci sirove, neorganizirane činjenice koje treba obraditi.

U današnje vrijeme, podataka je zaista puno i nekada je teško doći do podataka koji će obradom imati smisla za poslovanje. U tome pomaže rudarenje podataka, odnosno tehnike rudarenja podataka, čime ćemo se zapravo baviti u ovom poglavlju. Metode rudarenja podataka služe za kreiranje podatkovnih modela, a koja metoda će se primijeniti ovisi o podacima koji nam trebaju i o tipu podatka koji tražimo. U praksi se istovremeno primjenjuje i više metoda za rudarenje podataka.

No s obzirom da je tema rada web inteligencija, bazirat ćemo se na rudarenje web-a.

Rudarenje weba (eng. *Web mining*) kao termin, skovan je 1996. godine. Odnosi se na primjenu metoda rudarenja podataka kako bi se automatski otkrivali web dokumenti i servisi, izvlačile informacije sa web izvora i otkrili opći uzorci na web-u. Tokom godina, istraživanje o rudarenju weba proširilo se kako bi se pokrila upotreba rudarenja podataka i sličnih tehnika za otkrivanje resursa, uzoraka, i znanja sa web-a i web povezanih podataka (primjerice podatci o korištenju weba ili web server log datoteke). Rudarenje weba preklapa se sa područjima koja uključuju rudarenje podataka, rudarenje teksta, povrat informacija i povrat weba.⁷⁴

Rudarenje weba dijeli se u 3 kategorije⁷⁵: “*Rudarenje sadržaja weba, rudarenje strukture web-a i rudarenje korištenja web-a*”.

Rudarenje sadržaja weba odnosi se na otkrivanje korisnih informacija iz web sadržaja, uključujući tekst, slike, zvuk i video. Osim toga, rudarenje sadržaja weba uključuje i otkrivanje resursa sa weba, kategorizaciju dokumenata i klasteriranje, te izvlačenje informacija sa web stranica.

Rudarenje strukture weba istražuje potencijalne modele koji su osnova vezne strukture weba. Obično uključuje i analizu *in-linkova* i *out-linkova*, te se koristi za rangiranje rezultata tražilice i drugih web aplikacije.

⁷⁴ H. Chen i M. Chau, 'Web Mining: Machine Learning for Web Applications', *Annual Review of Information Science and Technology*, vol. 38, no.1, 2004., p. 290

⁷⁵ A. K. Pujari, *Data Mining Techniques*, Hyderabad, Universities Press (India) Private Limited, 2001., p. 232

Rudarenje korištenja weba fokusira se na korištenje tehnika rudarenja podataka u analizi log datoteka traženja i drugih aktivnosti kako bi se pronašli zanimljivi uzorci. Jedna od osnovnih primjena rudarenja korištenja weba je u pravljenju korisničkih profila.⁷⁶

Neke od ključnih metoda rudarenja podataka su metoda asocijacije, klasifikacije, klasteriranja, predviđanja, sekvencijalnih uzoraka, te stabla odlučivanja.⁷⁷

5.1 Metoda asocijacije

Metoda asocijacije temelji se na asocijacijskim pravilima. Asocijacijska pravila slična su klasifikacijskim pravilima, na koja ćemo se osvrnuti u sljedećem podnaslovu. Asocijacijska pravila pronalaze sva pravila u bazi podatka kako bi zadovoljili zahtjeve minimalne podrške (eng. *Support*) i minimalne pouzdanosti (eng. *Confidence*). Podrška je pokazatelj koliko se često skup stavki (eng. *itemset*) pojavljuje u skupu podataka (eng. *Dataset*).

Pouzdanost je pokazatelj koliko često pravilo bude istinito.

Za asocijacijska pravila svrha nije predefinirana, kao što je slučaj kod klasifikacijskog pravila.⁷⁸

Kod asocijacije radi se jednostavna korelacija između dva ili više podataka, koji su često istog tipa radi identificiranja uzoraka. Primjerice, kada pratimo navike potrošača, možemo identificirati da kupac uvijek kupuje sladoled kada kupuje jagode, i tada sljedeći put kada kupac kupuje jagode, to nam implicira da bi mogao kupiti i sladoled, te mu možemo ponuditi i neki popust.⁷⁹

5.2 Metoda klasifikacije

Klasifikacija je, uz asocijaciju, jedna od ključnih metoda rudarenja podataka pomoću koje se neorganizirani skupovi poznatih vrijednosti s imenima (eng. *Tuples*), ponašajući se kao uzorak za vježbu, analiziraju kako bi se napravio model prema danim podacima. Svaki *tuple* pripada predefiniranoj klasi, koja je određena jednim od atributa, taj atribut se zove klasificirajući atribut (eng. *Classifying attribute*). Kada se model izvede jednom, u budućnosti može služiti za

⁷⁶ H. Chen i M. Chau, *ibid.*, p. 302

⁷⁷ M. Brown, *Data mining techniques*, [web stranica] „<https://www.ibm.com/developerworks/library/ba-data-mining-techniques/>“ (pristupljeno: 08.09.2017.)

⁷⁸ B. Liu, W. Hsu i Y. Ma, 'Integrating Classification and Association Rule Mining', *KDD-98 Proceedings, American Association for Artificial Intelligence*, 1998., p.1

⁷⁹ M. Brown, *Data mining techniques*, [web stranica] „<https://www.ibm.com/developerworks/library/ba-data-mining-techniques/>“ (pristupljeno: 08.09.2017.)

kategoriziranje uzoraka podataka, kao i za bolje razumijevanje sadržaja baze podataka.

Klasifikacija se najčešće primjenjuje na odobravanje kredita, u marketingu i medicinskoj dijagnostici.⁸⁰

5.3 Metoda klasteriranja

Klasteriranje se odnosi na dijeljenje podataka u grupe sličnih objekata. Svaka grupa, odnosno klaster, sastoji se od objekata koji su slični jedni drugima, a različiti u odnosu na objekte u drugim grupama. Kada prikazujemo malu količinu podataka, s relativno malim brojem klastera, postignemo jednostavnost, ali po cijenu gubitka nekih detalja. Klasteriranje je način modeliranja podataka, koji vuče korijene iz matematike i statistike. Iz perspektive strojnog učenja, klasteri korespondiraju za skrivenim uzorcima, traženje klastera je bez nadzora (eng. *unsupervised*), te kao takvo rezultira reprezentacijom koncepta podataka. Klasteriranje je bez nadzorno učenje (eng. *unsupervised learning*) skrivenih podatkovnih koncepta. Klasteriranje primijenjeno na aplikacije za rudarenje podataka, susrelo se s tri dodatne komplikacije: prvu čine velike baze podataka, potom objekti sa puno atributa i na kraju postojanje atributa različitih tipova. Ove komplikacije predstavljaju pravi izazov klasičnim algoritmima za klasteriranje. Navedeni izazovi pridonijeli su razvoju moćnih naširoko primjenjivih metoda klasteriranja podataka u rudarenju podataka, koji se temelje na klasičnim tehnikama.⁸¹

Neke od metoda klasteriranja koje su se razvile iz klasičnih tehnika, te se upotrebljavaju i danas, su hijerarhijska metoda, metoda relokacije particija (eng. *Partitioning relocation methods*), metoda bazirane na gustoći (eng. *Density-based partitioning methods*), metode bazirane na rešetci (eng. *Grid-based methods*), metode bazirane na pojavi kategorijskih podataka (eng. *Methods based on co-occurrence of categorical data*), skalabilni algoritmi za klasteriranje, te algoritmi za višedimenzionalne podatke.⁸²

⁸⁰ M. Kamber et. al., 'Generalization and Decision Tree Induction: Efficient Classification in Data Mining', *Proceedings Seventh International Workshop on Research Issues in Data Engineering. High Performance Database Management for Large-Scale Applications*, Birmingham, IEEE, 1997., p. 111

⁸¹ P. Berkhin, 'A Survey of Clustering Data Mining Techniques', u J. Kogan (ed.), *Grouping Multidimensional Data*, Netherlands, Springer, 1998., p. 25

⁸² P. Berkhin, op. cit. p. 29

5.4 Metoda stabla odlučivanja

Stabla odlučivanja su strukture u obliku stabla koje predstavljaju skup odluka. Metoda stabla odlučivanja može generirati pravila za klasifikaciju skupa podataka. Specifične metode stabla odlučivanja uključuju klasifikacijska i regresijska stabla (eng. *Classification and Regression Trees – CART*), hi-kvadrat automatsku detekciju interakcije (eng. *Chi Square Automatic Interaction Detection – CHAID*). CART i CHAID su tehnike stabla odlučivanja korištene za klasifikaciju skupa podataka. One pružaju skup pravila koja se mogu primjenjivati na nove (neklasificirane) skupove podataka za predviđanje koji zapisi će imati dani ishod. CART obično zahtjeva manju podatkovnu pripremu za razliku od CHAID-a.⁸³

5.5 Primjena rudarenja weba u poslovanju

U prethodnom dijelu objasnili smo metode rudarenja podataka čija je upotreba ključna u rudarenju weba. U nastavku ćemo se osvrnuti na primjere u kojima se vidi primjena rudarenja weba u poslovanju.

5.5.1 Personalizirano korisničko iskustvo u B2C e-trgovini – Amazon

U začetku Amazona vizionar i direktor Jeff Bezos je rekao:

„In a traditional (brick-and-mortar) store, the main effort is in getting a customer to the store. Once a customer is in the store they are likely to make a purchase – since the cost of going to another store is high – and thus the marketing budget (focused on getting the customer to the store) is in general much higher than in-store customer experience budget (which keeps the customer in the store). In the case of an on-line store, getting in or out requires exactly one click, and thus the main focus must be on customer experience in the store“⁸⁴

Dakle, u početku kada nije bilo poslovanja cilj je bio samo privući kupca u trgovinu i znali bi da bi barem nešto kupio, no razvojem interneta i e-poslovanja treba se više potruditi oko kupca

⁸³ S. J. Lee i K. Siau, *Industrial Management & Data Systems 101/1*, MCB University Press, 2001., p. 44

⁸⁴ Jeff Bezos, *Amazon*, citirano u W. Chu i T. Y. Lin, *Foundations and Advances in Data Mining*, Springer, 2005, p. 288

jer je, kako kaže J. Bezos, samo jedan klik dovoljan da kupac ode.

Ova opservacija je ono što je vodilo Amazonov sveobuhvatan pristup personaliziranom iskustvu kupca, što zovu „personalizirana trgovina za svakog kupca“. Korištenje metoda rudarenja weba, kao što su asocijacijske metode između posjećenih stranica i analiziranje uzorka „klikanja“, koriste se u poboljšanju korisničkog iskustva prilikom „posjete trgovini“. Znanje dobiveno putem rudarenja weba je ključ inteligencije iza Amazonovih značajki poput instant preporuke (eng. *instant recommendations*), krugovi kupnje (eng. *purchase circles*), liste želja (eng. *wish lists*) i mnogih drugih.⁸⁵ Primjer sustava preporuka (eng. *recommender system*) tvrtke Amazon, klasičan je primjer upotrebe web rudarenja.

5.5.2 Metoda klasteriranja kao način segmentacije tržišta

Već smo ranije govorili kako je klasteriranje metoda dijeljenja više sličnih podataka u grupe, koji su različiti u odnosu na podatke neke druge grupe. Kada malo bolje pogledamo sam opis metode klasteriranja, možemo je odmah povezati sa segmentacijom tržišta u marketingu u okviru CRM-a. S obzirom na ovu poveznicu, između marketinga i metode klasteriranja, započela je primjena klasteriranja u poslovanju u svrhu segmentacije tržišta.

Tvrtka Claritas je pružatelj rješenja za segmentiranje potrošača kako bi omogućila marketingu ostvarivanje efektivnije, efikasnije kampanje bilo *online* ili *offline*. Claritasovi segmentacijski algoritmi i diferencirani izvori podataka pružaju jedinstveni uvid u poslovanje koje se tiče njihova tržišta i potrošača.⁸⁶ Claritas dijeli SAD na 66 segmenata koji su dostupni na njihovim web stranicama. Segmenti su kombinacija demografskih podataka iz U.S. Cenzusa sa bihevioralnim podacima prikupljenim iz više izvora, kako što su pretplate na časopise, registracije proizvoda, registracije automobila, i slično. Svaki segment u matrici je definiran po socijalnim grupama, sa zvučnim imenima poput „*Urban Uptown*“ ili „*City Centers*“, i po grupama životnog stadija sa imenima poput „*Young Achievers*“ ili „*Conservative Classics*“. Segmenti, zvani PRIZM kodovi, su dostupni na nekoliko razina geografske hijerarhije, uključujući individualna domaćinstva, ZIP+4 (najfinije granuliran poštanski broj u SAD-u, koji može biti kat u stanarskoj zgradi ili pak mala grupa kuća) i ZIP kod (kod nas poznat kao

⁸⁵ J. Srivastava, P. Desikan i V. Kumar, *Web Mining – Concepts, Applications and Research Directions*, University of Minnesota, p. 408-409

⁸⁶ *About Claritas – Tapping into the Power of Big Data*, [web stranica], 2017., <https://www.claritas.com/about-us.html> (pristupljeno: 15.09.2017.)

poštanski broj koji predstavlja cijeli grad ili selo, ili susjedstvo većih gradova).

Susjedstvo, odnosno prema prethodnoj definiciji ZIP kod, jednog od autora knjige „*Data Mining Techniques*“, prema Claritasu, pripada u sljedeće segmente: „*Money & Brains, American Dreams, Bohemian Mix, Young Digerati, Urban Achievers*“.

Dok susjedstvo, odnosno ZIP kod, drugog autora iste knjige, prema Claritasu pripada sljedećim segmentima: „*Bohemian Mix, American Dreams, Money & Brains, Young Digerati, The Cosmopolitans*“. Kao što vidimo, jedina je razlika što su kod jednog autora zastupljeni „*Urban Achievers*“, dok kod drugog autora su to „*The Cosmopolitans*“.

Segment „*Urban Achievers*“, prema Claritasu, opisan je na sljedeći način⁸⁷:

„Koncentrirani u lučkim gradovima države, Urban Achieversi su često prvo stajalište za nadolazeće imigrante iz Azije, Južne Amerike i Europe. Ovi mladi samci, parovi, i obitelji su tipično visoko educirani i etnički različiti: oko trećine je rođeno u inozemstvu, a još više njih govori jezik koji nije engleski.“

Segment „*The Cosmopolitans*“ definiran je kao⁸⁸:

„Educiran, malo iznad srednje skale (eng. Upper-midscale), i etnički različiti, The Cosmopolitans su urbani parovi u američkim gradovima koji su u ubrzanom razvoju. Koncentrirani u većim gradovima poput Las Vegasa, Miamijsa, Albuquerquea, ova domaćinstva uključuju starije, tzv. empty-nest⁸⁹ vlasnike kuća. Živahna socijalna scena okružuje njihove starije kuće i stanove, a stanari vole noćni život i uživaju u intenzivnim životnim stilom sa puno slobodnog vremena.“

Diskutabilno je koliko imena segmenata više otkrivaju nego što su nejasna. Ime nastoji istaknuti jedno ili dva obilježja klastera koji može biti definiran prema puno više faktora. Podaci susjedstva dvojice autora otkrivaju da populacija susjedstva nije stvarno slična kao što to ime

⁸⁷ M. J. A. Berry i G. S. Lindoff, *Data Mining Techniques –Third Edition*, Indianapolis, Wiley Publishing Inc., 2011., p. 367

⁸⁸ M. J. A. Berry i G. S. Lindoff, loc. cit., p. 367

⁸⁹ *Empty-nest* – parovi koji imaju odraslu djecu, koja su se odselila od kuće

predlaže. Još jedan način za karakteriziranje klastera je promatranje tipičnog pripadnika i zatim ispitivanje koje značajke tipičnog pripadnika se najviše razlikuju od ukupne populacije.⁹⁰

5.5.3 Asocijacijska pravila za analizu potrošačke košarice u web trgovinama

Mnoge tvrtke spremaju zapise o transakcijama svojih kupaca. Takvi skupovi podataka pojavljuju se u različitim formama, primjerice proizvodi koji su kupljeni skupa, usluge koje se prate tokom vremena u CRM (eng. *Customer Relationship Management*) sustavima, posjete i aktivnosti na web stranicama, te telefonske zapise podrške za korisnike. Ovi zapisi vrlo su vrijedni za odjel marketinga jer sadrže informacije o kupcima, njihovim uzorcima kupnje, načinima na koje bi se mogao optimizirati način određivanja cijena, te veze između kupnji i drugih informacija o korisniku. Asocijacijska pravila nastoje pronaći skupove informativnih uzoraka iz velikih skupova podataka.⁹¹

Transakcijski podaci prikupljeni automatskim putem iz *online* trgovina pružaju pouzdane informacije o kupcima i proizvodima koje kupuju. Najčešća upotreba asocijacijskih pravila je pronalaženje proizvoda koji su komplementarni jedni drugima, ali korišteni skupa. To pozitivno utječe na prodaju jer kupcu koji kupuje proizvod A, možemo prikazati i proizvod B, C ili D koji se obično kupuju s tim proizvodom. Za upotrebu asocijacijskih pravila u svrhu analize potrošačke košarice koriste se:

- negativna asocijacijska pravila koja otkrivaju veze između proizvoda koji se nalazi u košarici i onih koji se ne nalaze. Primjerice možemo pronaći kupčevo ponašanje u kupnji poput onog kada netko kupuje Samsung, ali ne i Apple, ili čita znanstvenu fantastiku, ali ne i horore. Negativna pravila korisna su jer se putem njih može dinamično modificirati web stranica da kupcu ne prikazuje nešto što ono ne želi kupiti
- ciklična asocijacijska pravila uzimaju u obzir ponavljanje kupnje tokom vremena i putem toga se može prilagoditi izgled web stranica za kupca
- međutransakcijska asocijacijska pravila ne gledaju ovisnost pojedinih transakcija, već ispituju veze među kupnjama koje su obavljene u određenom vremenskom

⁹⁰ M. J. A. Berry i G. S. Lindoff, *ibid.*, p. 368

⁹¹ C. Chapman i E. McDonnell Feit, *R for Marketing Research and Analytics*, Springer, 2015., p. 12

periodu. Ova pravila u obzir uzimaju kontekst kupnje (vrijeme, mjesto, kupac) i cilj im je pronaći slijed događaja

- odnosna asocijacijska pravila – osim informacija o kupljenim proizvodima, sadrže i informacije o potrošenoj količini novca na razne proizvode i usluge. Kao rezultat, moguće je predvidjeti vrijednost transakcije i kupnje određenih proizvoda koji se nude u web trgovini.⁹²

⁹² A. Głowska, 'Data mining – how to analyze shopping carts to increase sales', *Synerise Blog*, [web blog], 5 travnja 2017., „<https://synerise.com/data-mining-how-to-analyze-customers-market-baskets-to-increase-sales>“ (pristupljeno: 15.09.2017.)

6 Web agenti kao elementi web inteligencije

Inteligentni agenti su računalni entiteti koji su sposobni donositi odluke u ime svojih korisnika i mogu sami poboljšati svoj učinak u dinamičnom promjenjivom i nepredvidivom okruženju. Autor J. Liu, u svojoj knjizi *Autonomous Agents and Multiagent Systems*, pruža opsežan pregled povezanih istraživanja u području autonomnih agenata i multi-agentskih sustava, s naglaskom na teorijske i računalne temelje kao i detaljne diskusije o tehnikama za razvijanje različitih utjelovljenja sustava baziranih na agentima, kao što su autonomni roboti, kolektivna vizija i pokreti, autonomna animacija, te agenti za pretraživanje i segmentaciju.

Inteligentni web agenti su *software*-ski programi koji igraju dvije vrlo važne uloge: autonomni entiteti za istraživanje i korištenje servisa temeljenih na web-u i prototip entiteti za izlaganje i objašnjavanje web-generiranih pravilnosti.⁹³

Inteligentni agenti dijele se na autonomne agente i multi-agentske sustave.

Autonomni agenti su entiteti koji “žive” u dinamičnom, nepredvidivom okruženju u kojem nastoje zadovoljiti vremenski ovisne ciljeve ili motivacije. Agenti su prilagodljivi ako poboljšavaju svoju kompetenciju u bavljenju s postavljenim ciljevima u skladu sa iskustvom. Autonomni agenti konstituiraju novi pristup istraživanju umjetne inteligencije (eng. *artificial intelligence*) koji je inspiriran biologijom, konkretnije, studijom o ponašanju inteligentnih živih organizama.

Agent je autonoman ukoliko djeluje potpuno autonomno, primjerice ako odlučuje u odabiranju najboljeg načina tako da njegovi ciljevi budu obavljeni uspješno. Agent je prilagodljiv ako je u moćnosti poboljšavati se tijekom vremena, primjerice ako postaje uspješniji u ostvarivanju ciljeva temeljem stečenog iskustva.⁹⁴

Kada govorimo o multi-agentskim sustavima, govorimo o grani umjetne inteligencije koja ima u cilju pružanje principa za pravljenje kompleksnih sustava koji uključuju više agenata i mehanizama za koordinaciju ponašanja pojedinih autonomnih agenata.⁹⁵

⁹³ Y. Y. Yao, et. al., 'Web Intelligence (WI) Research Challenges and Trends in the New Information Age', u N. Zhong et. al., *Web Intelligence: Research and Development*, Lecture Notes in Computer Science, vol. 2198., Springer, 2001., p.11

⁹⁴ P. Maes, 'Modeling Adaptive Autonomous Agents, MIT-Media Laboratory', *Artificial Life*, MIT Press Journals, p. 1-3

6.1 Inteligentni digitalni agent Alice – podrška pacijentima

Agentifai je tvrtka koja se bavi razvojem inteligentnih agenata, a specijalizirali su se za industriju zdravstvene njege i osiguranja. Koristeći algoritme poput dubokih nervnih mreža (eng. *Deep Neural Networks*), procesiranja prirodnog jezika (eng. *Natural Language Processing*) i ostalih moćnih algoritama, Agentifai je stvorio Alice, digitalnog inteligentnog agenta koji je sposoban razumjeti i voditi prirodni razgovor s klijentima.

Alice je najnapredniji agent umjetne inteligencije za odnose s pacijentima. Alice pruža informacije, odgovara na pitanja i izvršava zahtjeve pacijenata poput zakazivanja termina, potpuno autonomno, bez potrebe za ljudskom intervencijom. Alice je uvijek dostupna preko Facebook *Messengera* (web aplikacije za razgovor), na način da radi kao osobni asistent tvrtke za svakog klijenta. Alice je moguće integrirati i u Skype, Twitter ili na web stranicu kao *widget* za komunikaciju sa klijentima (eng. *Web Chat Widget*).

Alice ima mogućnost odgovaranja na većinu tipičnih pitanja koja dođu do kontakt centra, pitati klijente da ocijene uslugu, odgovarati samostalno na pitanja, autonomno povećavati životnu vrijednost kupca slanjem obavijesti pravim klijentima, odgovarati na komentare na društvenim mrežama, dostupna je 24h na dan 7 dana tjedno, podržava razgovore na više jezika, te se ponaša kao osobni asistent koji daje prilagođene informacije i podsjeća na uzimanje lijekova.

Ukratko, Alice je inteligentni agent namijenjen tvrtkama koje se bave zdravstvenom njegom i vrlo su ozbiljni u odnosima s kupcima.⁹⁶

6.2 Online aukcijski sustav temeljen na web agentima

Aukcije pružaju opće rješenje problema diskretne alokacije resursa među sebičnim agentima u multi-agentskim sustavima. Postoji više tipova aukcija, neke od njih su aukcije s jednim dobrom, s više dobra, kombinatoričke aukcije te duple aukcije. Aukcije predstavljaju posebnu klasu pregovora s više primjena u provođenju transakcija e-poslovanja. Aukcije su korisne za trgovanje u sljedećim područjima: licencama za spektar (korištenje radio frekvencijskog spektra u svrhu emitiranja na određenom području), tržišta električne energije, prava na emisiju zagađivača (štetnih supstanci), položajima za uzlijetanje i slijetanje na zračnim lukama, prava

⁹⁶Agentifai, *Autonomous by default*, [web stranica], “<http://agentifai.com/autonomous-by-default/>” (pristupljeno: 15.09.2017.)

na eksploataciju prirodnih resursa (primjerice nafta), prodaje kolekcionarskih stvari, antikviteta, luksuznih i *second-hand* proizvoda, javnih ugovora za nabavu, međunarodne razmjene i mnogih drugih.

Online aukcije razvijale su se zajedno s web-om, no nedavno su se sva istraživanja fokusirala na razvoj više generičkih, fleksibilnih rješenja i rješenja za višekratnu uporabu, s namjerom primjene u B2C (eng. *Business-to-Consumer*) i u B2B (eng. *Business-to Business*) sektorima. U ovom kontekstu, predložena je primjena sustava temeljenih na agentima kao novi pristup koji ideju o aukcijskoj usluzi iz svijeta web-a, kojim upravljaju ljudi, prenosi u svijet *software*-skih agenata. Primjerice, u engleskoj aukciji prodavač (ili aukcionar koji ga predstavlja) objavljuje inicijalnu cijenu za određeno dobro, a sudionici aukcije nadmeću se povećavanjem svote novca u predefinirano vrijeme trajanja aukcije. Na kraju aukcije, agent s najvećom ponudom je pobjednik engleske aukcije. Kada je aukcija modelirana kao multi-agentski sustav, interakcija agenata kroz poruke je potrebna za nadmetanje i ažuriranje cijene. Centralni agent odgovoran je za cijelu komunikaciju sa i između sudionika aukcije. Ako postoji prevelik broj sudionika, centrali agent postaje preopterećen. Kako bi se to riješilo, centralni agent i dalje prima sve ponude od sudionika, ali su sudionici podijeljeni na grupe kojima upravlja pojedini *proxy* agent. Svaki sudionik šalje poruke svom *proxy* agentu koji tada prosljeđuje poruke centralnom agentu.⁹⁷

6.3 Inteligentni osobni asistent – Bixby

Veliki proizvođač elektroničkih uređaja, Samsung, nedavno je predstavio na svom najnovijem modelu pametnog telefona, Galaxy S8, osobnog asistenta nazvanog Bixby.

Bixby je zapravo inteligentni agent koji je integriran u svaki segment Samsungovog pametnog telefona, u poruke, pozive, galeriju i Internet. Klikom na gumb pametnog telefona, Bixby postaje spreman za rad. Bixby koristi prirodni jezik za komunikaciju i dostupan je non-stop. Osim što s Bixby-jem možete plaćati račune, upravljati alarmom, pronaći stare slike, otključati mobitel i tako dalje, s Bixby-jem možete i pričati kao s osobom. Primjerice, otišli ste negdje na odmor u neki nepoznati grad, i želite nešto pojesti, a ne znate kamo bi išli jer ste ipak u nepoznatom gradu. Upalite Bixby i kažete: „Bixby, pronađi mi najbliži restoran“, Bixby tada

⁹⁷ C. Badica et. al., ‘Distributed Agent-Based Online Auction System’, *Computing and Informatics*, vol. 33, 2014., p. 519-521

ovisno o vašoj lokaciji, na webu pronalazi najbliže restorane oko vas i daje vam trenutne rezultate u svega nekoliko sekundi. To je samo jedan od primjera, recimo zanima vas kakva je vremenska prognoza vani, dovoljno je upaliti Bixby, reći mu da vas zanima vremenska prognoza i Bixby u nekoliko sekundi izbacuje rezultate o vremenu koje je našao na web-u. Dakle, Bixby može za vas pronaći bilo što na web-u, što god vas zanima. Mogućnosti je stvarno beskonačno.⁹⁸

Na sličan način funkcionira Siri, inteligentni agent koji je također osobni asistent, a napravila ga je tvrtka Apple, te ga koristi na svojim uređajima (iPhone, iPod, iPad).

⁹⁸ *Samsung - Meet Bixby*, [web stranica], <http://bixby.samsung.com/meet-bixby> (pristupljeno: 15.09.2017.)

7 Web inteligencija u upravljanju znanjem

Sir Francis Bacon je rekao: „*Ipsa scientia potestas est*“, što bi u prijevodu značilo da je znanje moć. U začetcima organizacije poslovanja ljudski faktor se zanemarivao sve do pojave neoklasične teorije. Nakon toga shvatilo se koliko je ljudski faktor važan za poslovanje, odnosno da ljudi jesu tvrtka, jer bez ljudi tvrtka ne bi poslovala.

Znanje zaposlenika tvrtke, od top menadžmenta do radnika, je ono što tvrtku čini kompetentnom. Znanja i vještine su nešto što nitko ne može oduzeti, a svaka tvrtka bi trebala cijeliti svoje zaposlenike upravo svog njihovih znanja i vještina. Upravo zbog toga važno je stalno nadograđivati znanje, te stjecati nova znanja i vještine, a zaključno time treba ostati u korak sa vremenom, odnosno s tehnologijom vremena. Web inteligenciju moguće je primijeniti i u segmentu upravljanja znanjem (eng. *Knowledge management*) u tvrtki. U nastavku će se dati pregled na koji način se to postiže implementacijom web ekspertnih sustava i web sustava za stjecanje novih poslovnih znanja.

7.1 Web ekspertni sustavi

Spajanje internetskih tehnologija i polja ekspertnih sustava (eng. *Expert Systems – ES*) stvorilo je nove načine dijeljenja i raspodjele znanja. Internetske tehnologije stvorile su nove mogućnosti za poboljšanje tradicionalnih sustava za donošenje odluka i ekspertnih sustava. Internetske tehnologije mogu promijeniti način na koji je ekspertni sustav razvijen i raspodijeljen. Kroz web ekspertne sustave (eng. *Web-based Expert Systems*) znanje o bilo kojoj temi može se direktno dostaviti korisniku bilo gdje i u bilo koje vrijeme putem web-a. Glavna funkcija ekspertnog sustava je oponašanje ekspertize i raspodjela ekspertnog znanja korisnicima koji nisu eksperti. Koristi ovakvog sustava mnogo su se poboljšale pojavom interneta.

Ekspertni sustav je sustav koji koristi ljudsko znanje uzeto s računala u rješavanju problema koji obično zahtijevaju ljudsku ekspertizu.⁹⁹

U svom članku, J. Durkin je objavio da su mnoge tvrtke iskoristile ekspertne sustave kako bi

⁹⁹ Y. Duan, 'Web-based Expert Systems', u M. Khosrow-Pour, *Encyclopedia of Information Science and Technology*, 2nd edn., New York, Information Resources Management Association, 2009., p. 4105

poboljšale produktivnost i profite kroz bolje donošenje poslovnih odluka.¹⁰⁰

U početcima ekspertni sustavi su bile samostalne aplikacije instalirane na glavnom centralnom računalu tvrtke (eng. *mainframe*), radnim stanicama za umjetnu inteligenciju ili računalnim platformama, a kasnije razvojem internetskih tehnologija poput LAN-a i web-a Ekspertni sustavi postali su dostupni svima.¹⁰¹

7.1.1 Primjer Web ekspertnog sustava - WITS

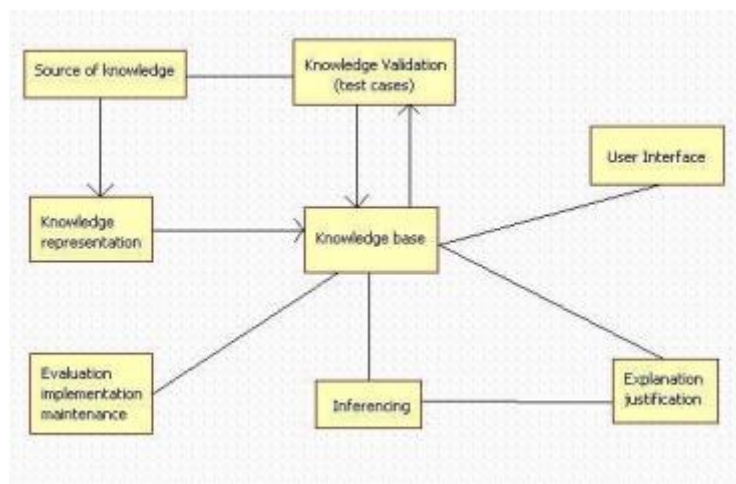
Primjer web ekspertnog sustava je WITS (eng. *Web-based Intelligent Training and Support System*). WITS je razvijen u svrhu pružanja treninga i inteligentne podrške malim i srednjim poduzećima u korištenju informacijsko-komunikacijskih tehnologija. Istraživanje je inspirirano dokazima iz literature da je nedostatak primjerenih vještina i znanja jedno od glavnih barijera za mala i srednja poduzeća u uspješnoj implementaciji i pokretanju e-trgovine i općenito e-poslovanja. Kao rezultat ove neefikasnosti, postoji hitna potreba za boljom edukacijom i podrškom za odlučivanje za menadžere malih i srednjih poduzeća koji žele prihvatiti tehnologiju, ali se boje zaostajanja za konkurencijom. WITS ima 3 podsustava, koji su dizajnirani s ciljem da olakšaju menadžerima malih i srednjih poduzeća proces donošenja odluka u implementaciji e-trgovine i e-poslovanja. Prema slici 8. arhitektura WITS-a sastoji se od potvrđivanja znanja (eng. *Knowledge validation*), reprezentacije znanja (eng. *Knowledge representation*), zaključivanja (eng. *Inferencing*) te objašnjenja i opravdavanja (eng. *Explanation and justification*).

Potvrđivanje znanja potvrđuje i verificira znanje s obzirom na testne slučajeve dok kvaliteta znanja nije prihvatljiva, potom reprezentacija znanja priprema mapu znanja i kodira dano znanje u bazu znanja. Komponenta zaključivanja omogućava računalu da donosi zaključke temeljene na znanju i specifičnosti problema. Objašnjenje i opravdavanje je program koji omogućava sustavu da odgovori na pitanja o specifičnom dijelu informacije ili kako je određeni zaključak donesen. Direktne strelice, na slici 8., označavaju jednostranu vezu, ostale označavaju dvostranu vezu.¹⁰²

¹⁰⁰ J. Durkin, 'Expert Systems: A View of the Field', *IEEE Expert: Intelligent Systems and Their Applications*, vol. 11, no. 2, 1996., p. 62

¹⁰¹ Op. cit. Y. Duan, p. 4105

¹⁰² N. Askarov, K. Kozhakhmet i L. Atymtayeva, 'Analysis of Web-based Applications for Expert System', *Computer Modelling and New Technologies*, vol. 15, no. 4, 2011., 41-42



Slika 8 Arhitektura WITS web ekspertnog sustava¹⁰³

7.2 Web sustavi za stjecanje novih znanja

Tradicionalni kontekst učenja doživio je drastične promjene. Poučavanje i učenje više nije ograničeno samo na tradicionalne učionice. Online učenje ili često u literaturi e-učenje (eng. *e learning*), odnosno učenje preko interneta je postao fenomen u zadnjem desetljeću. Škole i korporacije investiraju značajnu količinu novca i vremena u razvoj online alternativa tradicionalnog tipa edukacije i sustava treninga. Kako bi zaposlenici bili u toku (eng. *up-to-date*) sa najnovijim znanjima i tehnologijama, ulaže se u takve sustave kako bi tvrtke bile konkurentne u vrlo brzom poslovnom svijetu. Efektivne i efikasne metode treninga su velika potražnja tvrtki kako bi osigurale da su zaposlenici i njihovi partneri opremljeni najnovijim informacijama i najnaprednijim vještinama. Mnoge tvrtke implementirale su rješenja za e-učenje za njihov korporativni trening; Dell, CISCO i HP su samo neki od njih.¹⁰⁴

Što se tiče e-učenja, ono može biti asinkrono i sinkrono. Asinkrono e-učenje podrazumijeva sustav koji je već napravljen te je dostupan zaposlenicima bilo kada i bilo gdje. Sinkrono e-

¹⁰³ N. Askarov, K. Kozhakhmet i L. Atymtayeva, 'System Architecture of WITS', u N. Askarov, K. Kozhakhmet i L. Atymtayeva, 'Analysis of Web-based Applications for Exper System', *Computer Modelling and New Technologies*, vol. 15, no. 4, 2011., p. 42

¹⁰⁴ Y. Wang, H. Wang i D. Y. Shee, 'Measuring e-learning systems success in an organizational context: Scale development and validation', *Computers in Human Behavior*, vol. 23., 2007., p. 1793

učenje je rijetka pojava, jer ono ide uživo, a rijetkost je da svi zaposlenici budu pred ekranima u isto vrijeme.¹⁰⁵

7.2.1 Primjer Web sustava za stjecanje novih znanja - Dow Chemical-ov trening za intervju

Dow Chemical's je tvrtka koja je razvila asinkroni trening za mala i srednja poduzeća. To je tečaj za zaposlenike koji planiraju ići na intervju za posao. Kako bi potaknuli polaznike, koriste mnogo interaktivnih pristupa (poput *hyperlinkova* i gumba) kako bi polaznik morao komunicirati s tečajem. Interakcija i razumijevanje potiču se različitim vrstama testiranja, uključujući višestruke odabire i tzv. *'drag and drop'*¹⁰⁶. Primjerice, jedna vježba od korisnika traži da spoji aktivnosti, poput 'usporedite kvalifikacije kandidata za zahtjevima posla', sa različitim ulogama ispitivača, poput 'procjene podataka'. Kako bi odgovorili na pitanje polaznici povlače aktivnost s desne strane monitora kako bi je povezali sa ulogom na lijevoj strani monitora. Sustav daje trenutačnu povratnu informaciju je li odgovor na pitanje točan. Cijeli tečaj traje otprilike 2 sata, ali polaznici mogu stati, spremiti svoj napredak i vratiti se kasnije kako bi ga završili. Polaznici mogu pristupiti pravom ispitu nakon završenog tečaja, te ako ga polože položili su i tečaj.

Za ovaj i ostale sustave za e-učenje, Dow Chemical je osvojio ASTD-ovu nagradu izvrsnosti u nagradama za praksu za elektroničke tehnologije za učenje (eng. *ASTD's Excellence in Practice Award for Electronic Learning Technologies*)¹⁰⁷

7.2.2 Adobe Captivate Prime

Adobe Captivate Prime je sustav za upravljanje učenjem (eng. *Learning Management System*) sljedeće generacije koji nudi personalizirano iskustvo učenja kroz više uređaja. Pomoću njega zaposlenici mogu steći sve potrebne vještine za posao.

Adobe Captivate Prime nudi intuitivno korisničko sučelje, jednostavno objavljivanje sadržaja koje je potrebno za trening zaposlenika, jednostavno postavljanje, praćenje svih aktivnosti

¹⁰⁵ E. T. Welsh, et. al., 'E-learning: emerging uses, empirical results and future directions', *International Journal of Training and Development*, vol 7., no. 4, 2003., p. 246

¹⁰⁶ Drag and drop – interakcija sa računalom na način da povučete objekt iz jedne kućice u drugu kućicu kojoj pripada objekt (primjerice u jednoj kućici se nalazi više godina rođenja, a korisnik mora odabrati pravu i povući je u za to predviđeno područje – kućicu)

¹⁰⁷ E. T. Welsh, et. al, op. cit., p. 247

treninga zaposlenika *online* i *offline* koristeći jednu platformu.

Pomoću njega možete upravljati korisnicima, odnosno upravljati internim i eksternim korisnicima, premještati ih između eksternih grupa i slično.

Ima ugrađeni *software* za reprodukciju bilo kakvog sadržaja za učenje, videa, PDF-a, PPT-a, DOCX dokumenata, xAPI i mnogih drugih, i to sve unutar jednog *software*-a za reprodukciju. Korisnicima je omogućeno učenje i offline tako da ne moraju uvijek biti spojeni na internet, te imaju dostupnu aplikaciju za pametne telefone odakle mogu pristupiti sadržaju za učenje *online* i *offline*.

Adobe Captivate Prime je priznat u poslovanju i nalazi se među top 10 najboljih sustava za *online* učenje namijenjenog poslovnim subjektima. Ovaj sustav su prepoznale tvrtke poput *Publicis Media*-e, *PRISMHR*-a, *Bancsource*-a te *Camelot Software*-a.¹⁰⁸

¹⁰⁸ *Adobe Captivate Prime*, [web stranica], <http://www.adobe.com/products/captivateprime.html> (pristupljeno: 15.09.2017.)

8 Web inteligencija i društvene mreže

Društvene mreže postale sastavni dio naše svakodnevnice, a među njima ističemo Facebook, Twitter i Google+ kao primjere najpoznatijih društvenih mreža, premda ih ima na desetke, čak i stotine ne toliko poznatih. Društvene mreže idealan su način za tvrtke da oglašavaju svoje proizvode skoro besplatno i to široj društvenoj zajednici. Naravno, postoje i oglasi za koje tvrtke plaćaju određenu svotu novca kako bi se oglašavali, ali samim time što tvrtka ima svoju stranicu na Facebook-u, Twitteru ili Google+-u, tvrtka može predstaviti sebe i svoje proizvode široj društvenoj zajednici. Web inteligencija je primjenjiva i na društvenim mrežama metodama sentiment analize i rudarenja društvenih mreža. Navedenim metodama tvrtke mogu ostvariti koristi, a vidjeti ćemo i na koji način u narednim cjelinama.

8.1 Sentiment analiza

Od davnih vremena, uvijek se kao ljudi pitamo što netko drugi misli o nama, no isto tako se i tvrtke pitaju što potrošači misle o njihovim proizvodima i uslugama. Rudarenje mišljenja, prvi puta se kao pojam pojavljuje 2003. godine u radu sa WWW konferencije¹⁰⁹. Idealan alat za rudarenje podataka bi trebao obrađivati skup pretraživanih rezultata za dani predmet, generirajući tako listu proizvodnih atributa (kvaliteta, značajke itd.) i sakupljati mišljenja za svaki predmet (loš, osrednji, dobar). No, pojam je isto tako interpretiran u širem značenju i uključuje različite tipove analize teksta. Pojam sentiment analiza paralelan je s pojmom rudarenje mišljenja u nekim aspektima, no kada se primjeni šira interpretacija oba pojma predstavljaju isto polje istraživanja. Pojam sentiment je korišten u referiranju na automatsku analizu teksta i praćenja predvidivih prosudbi.¹¹⁰

Sentiment analiza ili rudarenje mišljenja je računalno istraživanje ljudskih mišljenja, stavova i emocija o nekom entitetu. Entiteti mogu predstavljati pojedince, događaje ili neke teme. Teme su uglavnom pokrivene recenzijama. Stav pojedinca može biti pozitivan, negativan ili neutralan.¹¹¹

¹⁰⁹ K. Dave, S. Lawrence i D. M. Pennock, 'Mining the peanut gallery: Opinion extraction and semantic classification of product reviews', *Proceedings of WWW*, 2003., p. 519-528

¹¹⁰ B. Pang i L. Lee, 'Opinion Mining and Sentiment Analysis', *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, 2008, p. 9-10

¹¹¹ W. Medhat, A. Hassan i H. Korashy, 'Sentiment analysis algorithms and applications: A survey', *Ain Shams Engineering Journal*, vol. 5, 2014., p. 1093

Ako uzmemo u obzir definiciju sentiment analize, vidimo korist za tvrtku. Primjerice, tvrtka Asus, koja je proizvođač elektroničkih uređaja (laptopi, pametni telefoni, tableti itd.), vidi da joj su joj prihodi od prodaje pametnih telefona izrazito niski, postavlja pitanje „Zašto kupci nisu zadovoljni našim pametnim telefonima?“. Kako bi to istražili, mogu koristiti neki alat za sentiment analizu koji bi istražio mišljenja o određenom pametnom telefonu na webu, društvenim mrežama i slično. Analizom mišljenja sa raznih izvora ASUS bi mogao dobiti odgovore koji je problem s njihovim pametnim telefonima. Neki od mogućih odgovora su da je dizajn loš, da je loša kamera, da je operativni sustav kompliciran i slično.

8.1.1 Upotreba sentiment analize u praksi

Društvene mreže postale su mjesta gdje ljudi govore o tvrtkama i njihovim proizvodima. Tvrtke to znaju i pažljivo slušaju zahvaljujući *software*-ima koji su dovoljno pametni da otkriju emocije iza riječi. Klasičan primjer je Starbucks, gdje svake sekunde nastane oko 10 *tweet*-ova o Starbucksu. Tvrtka tada pomoću sentiment analize analizira mišljenja i emocije iza riječi. Bez obzira na veliki broj *tweet*-ova, Starbucks konzistentno uspijeva pronaći *tweet*-ove gdje se kupci žale na proizvod te nastoje odgovoriti na njih.¹¹²

Starbucks koristeći sentiment analizu poboljšava svoj CRM sa kupcima, što znači da im je ipak stalo do svojih kupaca. Nije dakle čudno da je Starbucks vodeći lanac kafića u SAD-u.

8.2 Rudarenje podataka na društvenim mrežama

Svakodnevno se na društvenim mrežama generira ogromna količina podataka. Za tvrtke, potrošače i pružatelje usluga važno je pronaći način kako upravljati s masom generiranih podataka od strane korisnika. Generirani podaci na društvenim mrežama razlikuju se od tradicionalnih podataka za klasično rudarenje podataka. Podaci koje generiraju društvene mreže su ogromni, glasni, raspodijeljeni, nestrukturirani i dinamični. Sve to predstavlja izazov za rudarenje podataka, te je potrebno pronaći nove tehnike i algoritme. Primjerice, prema podacima iz 2017. Facebook mjesečno ima 1,3 milijarde aktivnih korisnika, dok Twitter ima

¹¹² J. Bort, 'How Starbucks And Other Companies Use Complex Math Algorithms To Read Your Feelings Online', *Business Insider*, [web stranica], 2012., <http://www.businessinsider.com/twitter-facebook-monitoring-2012-11> (pristupljeno: 15.09.2017.)

328 milijuna aktivnih korisnika mjesečno.¹¹³ Podaci na društvenim mrežama su svakako distribuirani jer nema centralnog autoriteta koji bi održavao te podatke. Distribuiranost predstavlja izazov za istraživače, a i tvrtke, da razumiju tok podataka na društvenoj mreži. Isto tako, podaci su često nestrukturirani što predstavlja veliki izazov.¹¹⁴

Iako je rudarenje podataka na društvenim mrežama kompleksno, iz ogromne količine podataka tvrtke zasigurno mogu izvući informacije koje su im potrebne da bi svoje proizvode i usluge približile potencijalnim kupcima, i naravno, već postojećim korisnicima.

8.2.1 Rapleaf Inc. – rudarenje web-a

Rapleaf Inc. je američka tvrtka koja se bavila marketinškim podacima i razvojem *software*-a. Tvrtka postoji i dan danas, no posluje kao podružnica tvrtke TowerData.¹¹⁵ Tada samostalna tvrtka, Rapleaf Inc. je rudarila podatke od više od 389 milijuna klijenata. Radili su to na način da su rudarili forume, društvene mreže, blogove i slične stranice gdje su informacije javno dostupne. Klijenti s kojima oni posluju uključuju automobilske tvrtke, zrakoplovne kompanije, hotele, banke, prodavače, neprofitne organizacije i političare. Ako mogu naučiti više o svojim klijentima, onda mogu personalizirati iskustvo za pojedinog korisnika. Njihovi korisnici već očekuju visoki nivo usluge, i sve ovisi o proizvodu, usluzi i iskustvu koje oni preferiraju.¹¹⁶

¹¹³ *Social Media Statistics*, [web stranica], 2017., <https://chrisniderdesign.com/blog/resources/social-media-statistics/> (pristupljeno: 15.09.2017.)

¹¹⁴ P. Gundecha i H. Liu, 'Mining Social Media: A Brief Introduction', u *INFORMS Tutorials in Operations Research*, 2014., p. 4

¹¹⁵ *Tower Data – Under the hood*, [web stranica], <http://intelligence.towerdata.com/under-the-hood> (pristupljeno: 15.09.2017.)

¹¹⁶ L. Betancourt, *How Companies Are Using Your Social Media Data*, [web stranica], 2010., „<http://mashable.com/2010/03/02/data-mining-social-media/#XWIGuDKccEqd>“ (pristupljeno: 15.09.2017.)

Rasprava

Razvoj interneta bio je revolucija koja je potpuno promijenila svijet. Sam razvoj interneta sa sobom je povukao i razvoj mnogih drugih internetskih tehnologija, poput *world wide web-a*. Pojava *www-a* od velikog je značaja, jer bez te tehnologije ne bi postojalo e-poslovanje u onom obliku kakvog poznajemo danas. Pojava web-a omogućila je daljnji razvoj tehnologija koje se na njemu temelje, a samim time omogućila se i primjena tih tehnologija. Tehnologije poput web servisa, semantičkog weba, dinamičnih web stranica, rudarenja web-a, web agenata, te web ekspertnih sustava samo su neke od tehnologija koje su se razvile iz web-a. Upravo ove tehnologije predstavljaju web inteligenciju, koja je relativno novo područje informacijskih tehnologija. Web inteligencija je područje koje spaja napredne informacijske tehnologije i umjetnu inteligenciju na web-u i internetu. Web inteligencija ima vrlo široko područje primjene u poslovanju od kojih smo, za potrebe rada, izdvojili primjenu u okviru semantičkog web-a, unaprjeđenje web sjedišta tvrtki, rudarenje web-a, primjenu web agenata u poslovanju, te upotrebu u svrhu stjecanja novih znanja i vještina zaposlenika.

Semantički web dao je osnovnu arhitekturu na osnovu koje se stvorila mogućnost razvoja inteligentnih web sustava. Arhitektura semantičkog web-a temelji se na slojevima, a svaki sloj obavlja svoju funkciju u funkcioniranju semantičkog weba.

Već smo rekli kako web inteligencija ima široko područje primjene u poslovanju. Jedno od tih područja primjene je unaprjeđenje web sjedišta tvrtke. Gotovo sve tvrtke imaju vlastito web sjedište. Putem njih tvrtke mogu na neki način komunicirati sa svojim klijentima i partnerima, bili oni postojeći ili potencijalni. Implementacijom web inteligencije na web sjedište nudi se mogućnost stvaranja personalizirane stranice za svakog individualnog korisnika. Personaliziranom web stranicom tvrtka se može više „približiti“ korisniku, odnosno prikazivati mu sadržaj koji on želi vidjeti, i samim time se on osjeća vrijednijim, premda je sav taj proces automatiziran.

Kako bi tvrtke saznale što njihovi kupci žele i što ih zanima, koriste metode rudarenja web-a. One im omogućuju, putem raznih algoritama koji su programirani za specifične namjene, da dođu do podataka koji će im pomoći da bolje razumiju potrebe svojih postojećih i budućih klijenata. Putem rudarenja web-a tvrtke su u mogućnosti dobiti uvid u to kako poboljšati svoje proizvode i/ili usluge kako bi njima mogli privući što više klijenata. Rudarenje web-a posebno koristi odjel marketinga, što je i logično s obzirom da je marketing odjel tvrtke koji je „najbliže“

tržištu. Za odjel marketinga posebno su zanimljive društvene mreže poput Twittera i Facebooka na koje je također moguće primijeniti metode rudarenja weba. Za društvene mreže posebno je osmišljen algoritam rudarenja mišljenja odnosno sentiment analize. Ovaj algoritam kompleksnim matematičkim metodama putem web-a analizira mišljenja ljudi i otkriva emocije koje stoje iza njega. Putem ovog algoritma odjel poput marketinga ima mogućnost otkrivanja što ljudi misle o proizvodima tvrtke i jesu li njime zadovoljni, nezadovoljni ili podijeljenog mišljenja. Po mojem mišljenju, tvrtke puno mogu otkriti o svojim proizvodima putem ovog algoritma, jer im on omogućava traženje negativnih mišljenja, odnosno izjava ljudi u kojima se osjeća nezadovoljstvo proizvodom. Ona su posebno značajna jer tvrtke putem njih mogu ispraviti nedostatke svojih proizvoda ili usluga za koje su saznali putem nezadovoljnih izjava njihovih potrošača.

Tvrtke koriste web inteligenciju upravo zato da bi si olakšale poslovanje što je više moguće. Osim što web inteligencija nudi tehnologiju za izradu dinamičkih web stranica, omogućava izvlačenje podataka iz raznih web izvora, ona nudi i mogućnost stvaranja inteligentnih web agenata. Inteligentni web agenti su autonomni entiteti koji imaju mogućnost samostalnog obavljanja nekih zadataka za koje su namijenjeni. Tako postoje inteligentni agenti koji mogu komunicirati s klijentima tvrtke na prirodnom jeziku te mogu služiti kao korisnička podrška. Možemo samo zamisliti koliko se troškovi snižavaju korištenjem više inteligentnih agenata koji rade umjesto zaposlenika u pozivnom centru, a da ne govorimo o prednostima što te zaposlenike možemo iskoristiti i educirati kako bi radili u nekim drugim odjelima tvrtke.

Kada smo se već dotakli edukacije, web inteligenciju je moguće primijeniti i u tom području poslovanja. Postoje web sustavi posebno dizajnirani za edukaciju zaposlenika i učenje novih vještina. Putem tih sustava tvrtke su u mogućnosti osmisliti program za trening zaposlenika, pratiti njihov napredak i vidjeti tko se ističe. Prednost ovih sustava je što su zaposlenici u mogućnosti učiti kod kuće i na poslu, a programi za edukaciju dostupni su i putem pametnih telefona. Sav proces edukacije zaposlenika odvija se kroz zanimljivo korisničko sučelje, te čak nudi natjecateljsku komponentu kako bi se zaposlenici mogli međusobno natjecati u učenju i time zaraditi „priznanja“ na sustavu. Ovakav način edukacije zaposlenika zasigurno nije monoton, i može pridonijeti da zaposlenici uče brže, bolje i efikasnije.

No, osim web sustava za edukaciju, postoje i web ekspertni sustavi koji su posebno dizajnirani da koriste ljudsku ekspertizu u rješavanju nekih određenih problema. Postoje web ekspertni sustavi posebno osmišljeni za menadžere kako bi im pomogli u lakšem donošenju poslovnih

odluka. Znamo da je menadžerski posao stresan, a upravo ovi sustavi trebali bi smanjiti stres menadžera olakšavajući mu proces donošenja odluka prvenstveno pružanjem svih potrebnih informacija.

Zaključak

Web inteligencija ima vrlo široko područje primjene i teško je dati točnu definiciju što ona predstavlja. Snažnim razvojem interneta i *world wide web*-a nastala je i potreba za web inteligencijom. Web inteligencija je područje u znanosti koje spaja umjetnu inteligenciju i informacijske tehnologije. Za razvoj web inteligencije najprije je bilo potrebno postaviti temelje koji su postavljeni razvojem semantičkog web-a. Iako je relativno mlado područje istraživanja u informacijskim znanostima, razvilo se mnogo tehnologija web inteligencije. Rudarenje web-a pokazalo se kao jednim od najkorisnijih elemenata web inteligencije u području marketinga, dok su se ekspertni sustavi pokazali vrlo važnima u donošenju poslovnih odluka. Također ne smijemo zanemariti inteligentne web agente koji služe za pružanje korisničke podrške i time utječu na smanjenje troškova. Važno je napomenuti da web inteligencija ima inovativnu i efikasnu upotrebu u web sustavima za stjecanje novih znanja i vještina koji pomažu zaposlenicima steći nova znanja i vještine. Kroz primjere u praksi pokazali smo kolike prednosti sa sobom donosi upotreba web inteligencije. Nakon provedenog istraživanja možemo savjetovati tvrtkama koje žele steći konkurentsku prednost, koje žele biti bolje u poslovanju da svakako razmisle o implementaciji web inteligencije ukoliko je već ne koriste.

Literatura

- [1] *About Claritas – Tapping into the Power of Big Data*, [web stranica], 2017., <https://www.claritas.com/about-us.html> (pristupljeno: 15.09.2017.)
- [2] *Adobe Captivate Prime*, [web stranica], <http://www.adobe.com/products/captivateprime.html> (pristupljeno: 15.09.2017.)
- [3] Agentifai, *Autonomous by default*, [web stranica], <http://agentifai.com/autonomous-by-default/> (pristupljeno: 15.09.2017.)
- [4] *A JSON-based Serialization for Linked Data (JSON-LD)*, [web stranica], 2014., <https://www.w3.org/2001/sw/wiki/JSON-LD> (pristupljeno: 17.08.2017.)
- [5] Askarov, N., Kozhakhmet, K., i Atymtayeva, L., 'Analysis of Web-based Applications for Expert System', *Computer Modelling and New Technologies*, vol. 15, no. 4, 2011., {41-42}
- [6] Badica, C. et. al., 'Distributed Agent-Based Online Auction System', *Computing and Informatics*, vol. 33, 2014., {519-521}
- [7] Bas Uribe, J., L., *Case Study: Real Time Suggestion of Related Ideas in Financial Industry*, [web stranica], May 2007., <https://www.w3.org/2001/sw/sweo/public/UseCases/Bankinter/> (pristupljeno: 17.08.2017.)
- [8] Bell, S., *5 Websites that Successfully Leverage Dynamic Content Experiences to Increase Sales*, [web stranica], 2015., <https://www.cdnetworks.com/en/news/5-websites-that-successfully-leverage-dynamic-content-experiences-to-increase-sales/4226> (pristupljeno: 15.09.2017.)
- [9] Berkhin, P., 'A Survey of Clustering Data Mining Techniques', u J. Kogan (ed.), *Grouping Multidimensional Data*, Netherlands, Springer, 1998., {25,29}
- [10] Berry, M., J., A., i Lindoff, G., S., *Data Mining Techniques –Third Edition*, Indianapolis, Wiley Publishing Inc., 2011., {367}

- [11] Betancourt, L., *How Companies Are Using Your Social Media Data*, [web stranica], 2010., <http://mashable.com/2010/03/02/data-mining-social-media/#XWIguDKccEqd> (pristupljeno: 15.09.2017.)
- [12] Bezos, J., *Amazon*, citirano u W. Chu i T. Y. Lin, *Foundations and Advances in Data Mining*, Springer, 2005, {288}
- [13] Booth, D. et. al., 'Web Services Architecture', *W3C Working Group Note*, 2004., dostupno na: <https://www.w3.org/TR/ws-arch/wsa.pdf> (pristupljeno: 09.07.2017.) {7}
- [14] Bort, J., 'How Starbucks And Other Companies Use Complex Math Algorithms To Read Your Feelings Online', *Business Insider*, [web stranica], 2012., <http://www.businessinsider.com/twitter-facebook-monitoring-2012-11> (pristupljeno: 15.09.2017.)
- [15] Brown, M., *Data mining techniques*, [web stranica] <https://www.ibm.com/developerworks/library/ba-data-mining-techniques/> (pristupljeno: 08.09.2017.)
- [16] Bryant, M., *20 years ago, the World Wide Web opened to the public*, The Next Web (INSIDER), [web stranica], 2011., https://thenextweb.com/insider/2011/08/06/20-years-ago-today-the-world-wide-web-opened-to-the-public/#.tnw_5o4nxfsx (pristupljeno: 05.07.2017.)
- [17] Chapman, C., i McDonnell, Feit, E., *R for Marketing Research and Analytics*, Springer, 2015., {12}
- [18] Chen, H., i Chau, M., 'Web Mining: Machine Learning for Web Applications', *Annual Review of Information Science and Technology*, vol. 38, no. 1, 2004., {290}
- [19] Cheung, K. et. al., 'Semantic Web for Health Care and Life Sciences: a review of the state of the art', *Briefings in Bioinformatics*, vol. 10, no. 2, 2009., dostupno na: <https://academic.oup.com/bib/article/10/2/111/184475/Semantic-Web-for-Health-Care-and-Life-Sciences-a> (pristupljeno: 15.09.2017.)

- [20] Dave, K., Lawrence, S., i Pennock, D., M., 'Mining the peanut gallery: Opinion extraction and semantic classification of product reviews', *Proceedings of WWW*, 2003., {519-528}
- [21] Dingli, A., i Cassar, S., 'An Intelligent Framework for Website Usability', *Advances in Human-Computer Interaction*, vol. 2014., 2014., {1-2}
- [22] Duan, Y., 'Web-based Expert Systems', u M. Khosrow-Pour, *Encyclopedia of Information Science and Technology*, 2nd edn., New York, Information Resources Management Association, 2009., {4105}
- [23] Durkin, J., 'Expert Systems: A View of the Field', *IEEE Expert: Intelligent Systems and Their Applications*, vol. 11, no. 2, 1996., {62}
- [24] Fall, K., R., i Stevens, W., R., *TCP/IP Illustrated*, Addison-Wesley Professional, 2011., {196}
- [25] Fernandez, M., J., *Case Study: CRUZAR – An application of semantic matchmaking for eTourism in the city of Zaragoza*, [web stranica], August 2008., <https://www.w3.org/2001/sw/sweo/public/UseCases/Zaragoza-2/> (pristupljeno: 25.08.2017.)
- [26] Ferris, C., i Farrell, J., 'What Are Web Services?', *Communications of the ACM*, vol. 46, no. 6, 2003., {31}
- [27] Frauenfelder, M., *A Smarter Web*, [web stranica], 2001., <https://www.technologyreview.com/s/401241/a-smarter-web/> (pristupljeno: 01.08.2017.)
- [28] Gehringer, E., F., *Web services: SOAP, WSDL, UDDI*, CSC/ECE 517 Lecture Notes, North Carolina State University, [white paper] 2005., dostupno na: <https://people.engr.ncsu.edu/efg/517/f05/common/lectures/lectures/lec24.pdf> (pristupljeno: 09.07.2017.)
- [29] *Gleaning Resource Description from Dialects of Languages (GRDDL)*, [web stranica], 2007., <https://www.w3.org/2001/sw/wiki/GRDDL> (pristupljeno: 17.08.2017.)

- [30] Głowacka, A., 'Data mining – how to analyze shopping carts to increase sales', *Synerise Blog*, [web blog], 5 travnja 2017., <https://synerise.com/data-mining-how-to-analyze-customers-market-baskets-to-increase-sales/#> (pristupljeno: 15.09.2017.)
- [31] Goralski, W., *The Illustrated Network*, Massachusetts, Morgan Kaufmann Publishers, 2009., {569-570}
- [32] Greenly, W., *Case study: Contextual Search for Volkswagen and the Automotive Industry*, [web stranica], October 2011., <https://www.w3.org/2001/sw/sweo/public/UseCases/Volkswagen/> (pristupljeno: 17.08.2017.)
- [33] Gruber, T., 'Ontology', u L. Liu i M. Tamer Ozsu (eds.), *Encyclopedia of Database Systems*, 2009., dostupno na: <http://tomgruber.org/writing/ontology-definition-2007.htm> (pristupljeno: 15.09.2017.)
- [34] Gundecha, P., i Liu, H., 'Mining Social Media: A Brief Introduction', u *INFORMS Tutorials in Operations Research*, 2014., {4}
- [35] Hardin, S., 'Tim Berners-Lee: The semantic Web-Web of machine processable data', *Bulletin of the American Society for Information Science and Technology*, vol. 31, no. 3, 2005., {13}
- [36] Holly, R., *Geek 101: Beginners guide to IMAP vs. PoP*, [web stranica], 2013., <https://www.geek.com/chips/geek-101-pop-vs-imap-1536343/> (pristupljeno: 19.07.2017.)
- [37] *Internet 101: What is Internet?*, GCF LearnFree.org, Goodwill Community Foundation Inc., [web stranica], 2013. <https://www.gcflearnfree.org/internetbasics/what-is-the-internet/1/> (pristupljeno: 05.07.2017.)
- [38] *IPv4 and IPv6*, [web stranica], http://biotech.law.lsu.edu/blog/ipv4_ipv6.pdf (pristupljeno: 05.07.2017.)
- [39] Jiang, B., 'Is Inmon's Data Warehouse Definition Still Accurate?', *BeyeNETWORK*, 2012., dostupno na: <http://www.b-eye-network.com/view/16066> (pristupljeno: 14.07.2017.)

- [40] Kamber, M. et. al., 'Generalization and Decision Tree Induction: Efficient Classification in Data Mining', *Proceedings Seventh International Workshop on Research Issues in Data Engineering. High Performance Database Management for Large-Scale Applications*, Birmingham, IEEE, 1997., {111}
- [41] Katragadda, R., Tirumala, S., S., i Nandigam, D., 'ETL tools for Data Warehousing: An empirical study of Open Source Talend Studio versus Microsoft SSIS', *Computing Conference Papers*, 2015., p. 1-2 <http://unitec.researchbank.ac.nz/handle/10652/3366> (pristupljeno: 09.07.2017.)
- [42] Kleinrock, L., 'History of the Internet and Its Flexible Future', *IEEE Wireless Communications*, vol. 15, no. 1, 2008., {10}
- [43] Koper, R., 'Use of the semantic Web to Solve Some Basic Problem sin Education: Increase Flexible, Distrubuted Lifelong Learning, Decrease Teacher's Workload', *Journal of Interactive Media in Education*, vol. 6, 2004., dostupno na: <https://www-jime.open.ac.uk/articles/10.5334/2004-6-koper/> (pristupljeno: 15.09.2017.)
- [44] Lassila, O., i Swick, R., R., *Resource Description Framework (RDF) Model and Syntax Specification*, W3C Proposed Recommendation, [web stranica], 1991., <https://www.w3.org/TR/PR-rdf-syntax/> (pristupljeno: 15.08.2017.)
- [45] Lee, S., J., i Siau, K., *Industrial Management & Data Systems 101/1*, MCB University Press, 2001., {44}
- [46] Leiner, B., M. et. al., *Brief History of the Internet*, [web stranica], 1997., <http://www.internetsociety.org/internet/what-internet/history-internet/brief-history-internet> (pristupljeno: 05.07.2017.)
- [47] Liu, B., Hsu, W., i Ma, Y., 'Integrating Classification and Association Rule Mining', *KDD-98 Proceedings, American Association for Artificial Intelligence*, 1998., {1}
- [48] Maes, P., 'Modeling Adaptive Autonomous Agents, MIT-Media Laboratory', *Artificial Life*, MIT Press Journals, {1-3}

- [49] McCreesh, J., i Daniel, E., *Definition of Interoperability*, [web stranica], <http://interoperability-definition.info/en/> (pristupljeno: 15.09.2017.)
- [50] McGuinness, D. L., i van Harmelen, F., *OWL Web Ontology Language Overview*, [web stranica], 2004., <https://www.w3.org/TR/owl-features/> (pristupljeno: 03.08.2017.)
- [51] Medhat, W., Hassan, A., i Korashy, H., 'Sentiment analysis algorithms and applications: A survey', *Ain Shams Engineering Journal*, vol. 5, 2014., {1093}
- [52] Michel, J., P., *Web Service APIs and Libraries*, Chicago, American Library Association, 2013., {xi}
- [53] Pang, B., i Lee, L., 'Opinion Mining and Sentiment Analysis', *Foundations and Trends in Information Retrieval*, vol. 2, no. 1-2, 2008, {9-10}
- [54] Polzar, B., *Dynamic Web Pages*, University of Wisconsin, 2007., {3}
- [55] Postel, J., B., *Simple Mail Transfer Protocol*, University of Southern California, Information Sciences Institute, 1982. {2-3}
- [56] *Protocol for Web Description Resources (POWDER)*, [web stranica], 2009., <https://www.w3.org/2001/sw/wiki/POWDER> (pristupljeno: 17.08.2017.)
- [57] *PROV*, [web stranica], 2013., <https://www.w3.org/2001/sw/wiki/PROV> (pristupljeno: 17.08.2017.)
- [58] Pujari, A., K., *Data Mining Techniques*, Hyderabad, Universities Press (India) Private Limited, 2001., {232}
- [59] *RDF in Attributes (RDFa)*, [web stranica], 2013., <https://www.w3.org/2001/sw/wiki/RDFa> (pristupljeno: 17.08.2017.)
- [60] *RDF Vocabulary Description Language 1.0: RDF Schema (RDFS)*, [web stranica], 2004., <https://www.w3.org/2001/sw/wiki/RDFS> (pristupljeno: 17.08.2017.)

- [61] *Relational Databases to RDF (RDB2RDF)*, [web stranica], 2012., <https://www.w3.org/2001/sw/wiki/RDB2RDF> (pristupljeno: 18.07.2017.)
- [62] *Rule Interchange Format (RIF)*, [web stranica], 2010., <https://www.w3.org/2001/sw/wiki/RIF> (pristupljeno: 17.08.2017.)
- [63] Ruiz, J., F., *Case Study: An Intelligent Search Engine for Online Services for Public Administrations*, [web stranica], June 2007., <https://www.w3.org/2001/sw/sweo/public/UseCases/Zaragoza/> (pristupljeno: 17.08.2017.)
- [64] *Samsung - Meet Bixby*, [web stranica], <http://bixby.samsung.com/meet-bixby> (pristupljeno: 15.09.2017.)
- [65] *Simple Knowledge Organization System (SKOS)*, [web stranica], 2009., <https://www.w3.org/2001/sw/wiki/SKOS> (pristupljeno: 17.08.2017.)
- [66] *SPARQL Query Language for RDF*, [web stranica], 2013., <https://www.w3.org/2001/sw/wiki/SPARQL> (pristupljeno: 17.08.2017.)
- [67] *Social Media Statistics*, [web stranica], 2017., <https://chrissniderdesign.com/blog/resources/social-media-statistics/> (pristupljeno: 15.09.2017.)
- [68] 'The Semantic Web – A Web of Metadata', *Protagonist Blog*, [web blog], 2012., <https://protogenist.wordpress.com/tag/proof-trust-layer/> (pristupljeno: 03.08.2017.)
- [69] *Tower Data – Under the hood*, [web stranica], <http://intelligence.towerdata.com/under-the-hood> (pristupljeno: 15.09.2017.)
- [70] *W3C Semantic Web – Main Page*, [web stranica], 2014., https://www.w3.org/2001/sw/wiki/Main_Page (pristupljeno: 15.08.2017.)
- [71] Wagner, D., i Scheiner, B., *Analysis of the SSL 3.0 protocol*, University of California, 1997., {1}

- [72] Wang, Y., Wang, H., i Shee, D., Y., 'Measuring e-learning systems success in an organizational context: Scale development and validation', *Computers in Human Behavior*, vol. 23., 2007., {1793}
- [73] *Web Ontology Language (OWL)*, [web stranica], 2012., <https://www.w3.org/2001/sw/wiki/OWL> (pristupljeno: 17.08.2017.)
- [74] Welsh, E., T. et. al., 'E-learning: emeing uses, empirical results and future directions', *International Journal of Training and Development*, vol 7., no. 4, 2003., {246}
- [75] Wenzel, K. et. al., *Semantic Web Based Dynamic Energy Analysis and Forecasts in Manufacturing Engineering*, Fraunhofer Institute for Machine Tools and Forming Technology IWU, dostupno na: https://www.greencarbody.de/csdata/download/1/de/2011_02_05_cirp_semantic_web_207.pdf (pristupljeno: 15.09.2017.)
- [76] *Why Tourism?*, [web stranica], 2017., <http://www2.unwto.org/content/why-tourism> (pristupljeno: 20.09.2017.)
- [77] Wilamowski, B., M., i Irwin, J., D., *Industrial Communication Systems*, CRC Press, 2011., {59}
- [78] Williams, P., *A Short History of Data Warehouseing*, [web stranica], 2012., <http://www.dataversity.net/a-short-history-of-data-warehousing/> (pristupljeno: 14.07.2017.)
- [79] Woolcock, K., *Serialization in Java (Binary and XML)*, Iowa State Univeristy, 2014., dostupno na: <http://web.cs.iastate.edu/~smkautz/cs430s14/tutorials/current/Serialization%20in%20Java%20Tutorial.pdf> (pristupljeno: 09.07.2017.) {3}
- [80] 'XML (eXtensible Markup Language)', *Encyclopedia of Information Systems*, 2003., dostupno na: <http://www.sciencedirect.com/science/article/pii/B0122272404001994> (pristupljeno: 09.07.2017.)

- [81] Yao, Y., Y. et. al., 'Web Intelligence (WI) Research Challenges and Trends in the New Information Age', u N. Zhong et. al., *Web Intelligence: Research and Development*, Lecture Notes in Computer Science, vol. 2198., Springer, 2001., {11}
- [82] Yao, Y., Y. et. al., *Web Intelligence (WI) Research Challenges and Trends in the New Information Age*, Springer, 2001., {1-2}
- [83] Yi, K., 'Event driven dynamic web pages', Disertation thesis, Rochester Institute of Technology, 2006., {7-12}
- [84] Zhong, N., Liu, J., i Yao, Y., 'Web Intelligence', *Computer Software and Applications Conference, 2000. COMPSAC 2000. The 24th Annual International*, 2000., {5}

Popis slika

Slika 1 Prikaz svih protokola u TCP/IP modelu	7
Slika 2 Prikaz arhitekture centralnog Web API-ja	9
Slika 3 Bottom-Up i Top-Down pristup skladištenju podataka	13
Slika 4 Arhitektura semantičkog weba.....	16
Slika 5 Odabir artikla i gumb koji pokreće aplikaciju Fits.me (označen u plavom kvadratu).	27
Slika 6 Sučelje aplikacije za unos osobnih dimenzija.....	27
Slika 7 Slika mogućeg izgleda u odabranom artiklu prema unesenim dimenzijama.....	28
Slika 8 Arhitektura WITS web ekspertnog sustava	43

Web Intelligence

Summary

This thesis is about the field of web intelligence and its application on business. Methods and ways of how web intelligence is applied in different areas of business, such as marketing, CRM, company management and customer support, are covered within this thesis. We have put a special emphasis on real-life examples so we can present successful applications of web intelligence and try to show its significance in business. We have considered ways on how companies can benefit by applying web intelligence in decision making process and employee education. We have put an emphasis on importance of web mining using different methods and algorithms, such as sentiment analysis, so companies can collect as much data as they can about their customer and try to understand their behaviour during shopping. The results we came to, show that applying web intelligence in business contributes in better customer understanding, better possibility of adapting to customers and developing better customer relations.

Keywords: Web intelligence, semantic web, web mining, web agents, web expert systems, sentiment analysis