



Sveučilište u Rijeci
University of Rijeka
<http://www.uniri.hr>

Polytechnica: Journal of Technology Education, Volume 3, Number 2 (2019)
Politehnika: Časopis za tehnički odgoj i obrazovanje, Volumen 3, Broj 2 (2019)



Politehnika
Polytechnica
<http://www.politehnika.uniri.hr>
cte@uniri.hr

DOI: <https://doi.org/10.36978/cte.3.2.2>

Izvorni znanstveni članak
Original scientific paper
UDK 004.42:004.62

Karakteristike softvera za vizualizaciju podataka

Vedran Juričić

Filozofski fakultet

Sveučilište u Zagrebu

I. Lučića 3, 10000 Zagreb

vjuricic@ffzg.hr

Sažetak

U radu je definirana vizualizacija podataka, navedena je njezina važnost prvenstveno u znanosti i modernim poslovnim procesima, ali i u ostalim područjima, poput medicine i bankarstva. Analiza, pa tako i vizualizacija, velikih količina podataka danas se obavlja pomoću raznih alata, koji se razlikuju u cijeni, platformi, kvaliteti korisničke podrške, mogućnostima i funkcionalnosti. U radu se analiziraju i uspoređuju karakteristike velikog broja alata za vizualizaciju podataka s ciljem da se pronađu njihove najzastupljenije mogućnosti, odnosno zajednička funkcionalnost većine današnjih alata. Osim navedene analize, u radu se analizira i njihova raspoloživost na platformama i operacijskim sustavima, opcije korisničke podrške, cijena itd.

Ključne riječi: vizualizacija podataka; alati za vizualizaciju podataka; analiza alata; usporedba alata.

1 Uvod

Zbog brzog razvoja informacijsko-komunikacijske tehnologije i njezine sve veće primjene u različitim područjima, raste i količina podataka koji se pohranjuju na medijima ili koji se putem medija izmjenjuju između njezinih korisnika. Time se povećava i kompleksnost njihovog arhiviranja, uređivanja, obrade i vrednovanja, zbog čega su razvijene brojne tehnike prikupljanja, ekstrahiranja, pohranjivanja, analize, ali i vizualizacije podataka. Vizualizacija je vrlo važna u

modernim sustavima jer omogućuje brže i uspješnije odlučivanje, bolju analizu, poboljšanu suradnju i dijeljenje informacija, uštedu vremena, novaca i ostalih resursa (Wang i dr., 2015; Sucharitha i dr., 2014).

U vizualizaciji se vrlo često koriste termini: podaci, informacija i znanje, kako bi se naznačile različite razine apstrakcije i razumijevanja, pri čemu znanje ima najvišu razinu razumijevanja. Podaci, informacije i znanje tako mogu biti i ulaz i izlaz iz procesa vizualizacije (Chen i dr., 2009). Za navedena tri termina ne postoje jedinstvene definicije jer su one vezane za promatrano

područje, poput ekonomije, novinarstva, organizacije znanja itd. U računarstvu se podatak odnosi na bitove spremljene na mediju ili komunicirane putem njega, pa je u tom području i reprezentacija znanja samo jedna od mogućih vrsta podataka.

Velika količina podataka, njihova raznolikost i područja otežavaju postupke njihove analize i vizualizacije. Jedan od najčešćih problema je vizualni šum, koji nastaje kad su podaci previše povezani i kad ih je vrlo teško razdvojiti. Sljedeći problem je gubitak informacija, koji nastaje kad se nastoje poboljšati performanse, odnosno ubrzati prikaz ili analiza. Također, jedan od problema je i ograničenje samih korisnika, koje uključuje nemogućnost percepcije određenih slika i reagiranja na brze promjene prikaza (Ali i dr., 2016 ; Yur'evich Gorodov i dr., 2013).

Razvijeni su brojni alati koji na različite načine rješavaju dio problema koji se javljaju prilikom vizualizacije. Alati se, osim u pristupu navedenim problemima, razlikuju u cijeni, podržanim operacijskim sustavima, vrsti edukacije i korisničke podrške itd.

U ovom radu se promatra velik broj alata za vizualizaciju podataka, koji su opisani s više od 1600 različitih obilježja. Analiziraju se navedena obilježja, kako bi se dobilo trenutno stanje današnjih alata u različitim segmentima, poput podržanih operacijskih sustava i korisničke podrške, korištenja sigurnosnih mehanizama i mrežnih resursa, kolaboracije s drugim korisnicima itd. Analizom se nastoji utvrditi koje su najzastupljenije mogućnosti danas dostupnih alata, odnosno koja je zajednička funkcionalnost većine današnjih alata.

2 Vizualizacija podataka

U literaturi se mogu pronaći razne definicije vizualizacije podataka. Prema Liu i dr. (2002), vizualizacija podataka je metoda analize i obrade podataka s ciljem prezentiranja setova podataka pomoću slika i grafova, pri čemu se koriste tehnike poput interakcije, višedimenzionalne analize i rudarenja podataka. Few i Edge (2007) vizualizaciju podataka definiraju kao korištenje slika u reprezentiranju informacije i znanja, a

Sadiku i dr. (2016) kao prezentiranje podataka u grafičkom i slikovnom obliku, kako bi korisnik lakše razumio informaciju.

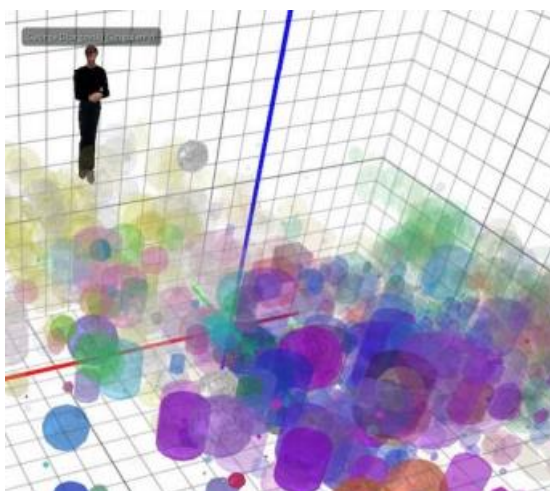
Zbog toga vizualizacija svakako nije novitet, a najstariji primjeri vizualizacije sežu u 2 stoljeće pr. Kr. i uključuju dijagrame, tablice s pozicijama zvijezda i navigacijske karte. S razvojem znanosti, posebno matematike, a kasnije i računarstva, razvile su se današnje, napredne tehnike vizualizacije, pri čemu je najveći rast zabilježen u zadnjih 30 godina (Friendly, 2006).

Robertson i dr. (1991) dijele vizualizaciju na znanstvenu vizualizaciju i vizualizaciju informacija. Znanstvena vizualizacija razlikuje tehnike na temelju dimenzionalnosti tipova podataka, odnosno broja nezavisnih varijabli, i vrste podataka, odnosno prema tome radi li se sa skalarima, vektorima, tenzorima i sl. i prikladna je za vizualizaciju konkretnih, fizičkih podataka. Vizualizacija informacija odnosi se na korištenje vizualnih reprezentacija za apstraktne, odnosno nefizičke podatke (Nagel, 2006).

Obzirom na različite vrste podataka, različita problemska područja i potrebne analize, korisnici i istraživači nailaze na problem odabira prikladnih tehnika vizualizacije. Razvijene su brojne taksonomije koje pokušavaju uspostaviti hijerarhiju i organizaciju znanstvenih vizualizacija i vizualizacija informacija. Tory i Moller (2004) razvili su taksonomiju temeljenu na modelu, pri čemu su za klasifikaciju odredili dva kriterija: kriterij diskretnosti ili kontinuiranosti podataka, i kriterij količine atributa koje je potrebno prikazati. Shneiderman (2000) predlaže taksonomiju na temelju tipova podataka. Postoji čitav niz taksonomija za tehnike vizualizacije temporalnih podataka, odnosno podataka koji predstavljaju stanje u vremenu, poput taksonomija autora Daassi i dr. (2005) i Chi (2000), koji je predložio taksonomiju temeljenu na referentnom modelu stanja podataka.

Standardne metode vizualizacije podataka odnose se na korištenje tablica, histograma, grafova, vremenskih linija, Vennovih dijagrama i dijagrama entiteta i veza. Iako nisu toliko poznate, koriste se i paralelne koordinate, stablaste strukture, semantičke mreže i cirkularni mrežni dijagrami. (Wang, 2015; Robertson, 1991). Za

određena područja razvijene su specijalizirane metode, poput metoda za vizualizaciju bežičnih mreža (Parbat i dr., 2010), gena (Michaels, Carr, 1998) i društvenih mreža (Ahn, 2013). Novije metode koriste virtualnu stvarnost (Donalek i dr., 2014), koja pokazuje vrlo dobre rezultate u vizualizaciji apstraktnih višedimenzionalnih podataka. Na slici 1. prikazan je istraživač koji u virtualnom svijetu radi analizu osmodimenzionalnih podataka. Osim prostornih koordinata koriste se oblici, veličine, boje, transparentnost, orijentacije, pulsiranja, teksture itd.



Slika 1. Korištenje virtualne stvarnosti u vizualizaciji podataka (Donalek et al. 2014)

3 Istraživanje

Podaci o alatima za vizualizaciju prikupljeni su sa mrežne stranice Capterra (Capterra.com, 2019). Capterra je pokrenuta 1999. i zamišljena je kao usluga koja korisnicima pomaže pronaći softver koji najviše odgovara njihovim potrebama. S druge strane, proizvođačima softvera omogućuje povećavanje njihove vidljivosti, ali i poboljšavanje funkcionalnosti, budući da korisnici mogu komentirati i ostavljati recenzije. Stranicu mjesečno posjeti više od milijun korisnika, a softver mogu pretražiti prema raznim kriterijima, poput cijene, broja korisnika, karakteristika i platforme. Za svaki softver korisnici mogu pogledati detaljne podatke, kao i recenzije drugih

korisnika, kojih trenutno ima više od milijun (Capterra.com, 2019). Softver je podijeljen u više od 700 kategorija, što omogućuje lakše pretraživanje i snalaženje jer su kategorije vrlo detaljno raščlanjene. Na primjer, stranica sadrži više od 20 kategorija koje u nazivu imaju riječ podatak, poput ekstrahiranja podataka, analize podataka, prevencije gubitka podataka i kvalitete podataka. U ovom radu analiziraju se alati koji pripadaju kategoriji vizualizacija podataka.

Capterra nema aplikacijsko programsko sučelje (engl. *Application Programming Interface, API*) pomoću kojeg se mogu dohvatiti podaci u obliku prikladnom za računalnu obradu. Zbog toga je za prikupljanje podataka napravljena aplikacija koja za svaki alat dohvaća HTML kôd njegove mrežne stranice sa sljedećim podacima:

- naziv, država i datum osnivanja tvrtke, odnosno vlasnika alata
- cijena i dostupnost besplatne verzije
- podržane platforme
- podržana korisnička podrška i vrste edukacije korisnika
- broj recenzija i ocjene za korisničku podršku, jednostavnost korištenja i ukupno zadovoljstvo softverom
- podržane značajke i njihove kategorije

Aplikacija parsira navedeni HTML kôd i dohvaćene podatke sprema u relacijsku bazu podataka, u obliku prikladnom za provođenje analiza.

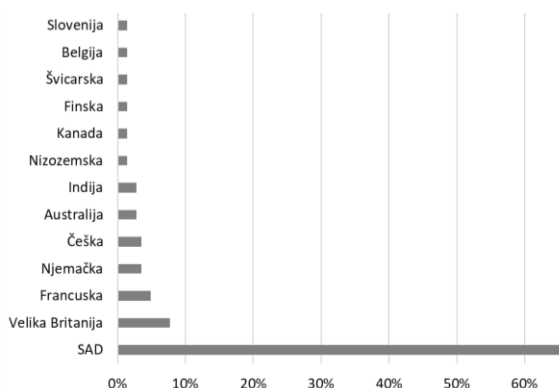
U kategoriji vizualizacija podataka nalazi se 208 alata, ali je inicijalnom analizom utvrđeno da za 17 alata podaci nisu potpuni, tako da oni nisu uzimani u obzir u daljnjem istraživanju. Dakle, analize provedene u radu se temelje na podacima o 191 alatu.

3.1 Analiza tvrtki

Za svaki alat prikupljeni su podaci o tvrtkama u čijem su vlasništvu. Osim naziva tvrtke i mrežnog sjedišta, prikupljeni su i podaci o državama u kojima se nalaze, kao i godina njihovog osnivanja.

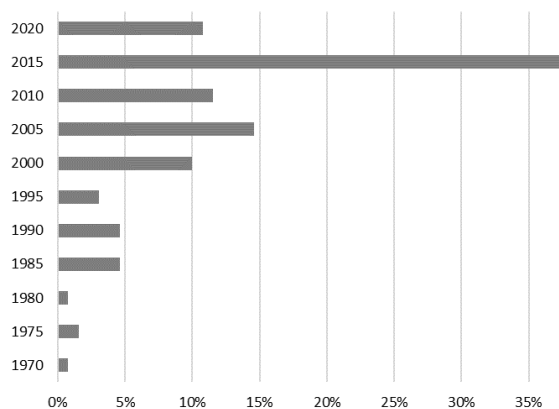
Na slici 2. prikazan je graf koji prikazuje raspodjelu alata prema državama. Vertikalna os prikazuje države u kojima se nalaze tvrtke -

vlasnici alata, a horizontalna os prikazuje njihov postotak. Prikazane su samo one države u kojima se nalaze barem dvije takve tvrtke. Vidljivo je da više od 65% alata dolazi iz tvrtki iz SAD-a. Prva država nakon nje je Velika Britanija za koju je evidentirano nešto manje od 8%. Iz Australije, Azije i Južne Amerike dolazi svega 9% alata, a iz Europe 34%. U Hrvatskoj se nalazi samo jedan alat, Kyubit Business Intelligence (Kyubit.com 2019).



Slika 2. Raspodjela prema državama

Slika 3 prikazuje raspodjelu prema godini osnivanja. Razdoblje je prikazano u grupama od 5 godina, što je prikazano na vertikalnoj osi, a horizontalna os prikazuje postotak tvrtki koje su osnovane u tom razdoblju. Gornja vrijednost, 2020, obuhvaća tvrtke osnovane od 2016. do današnjeg dana.



Slika 3. Raspodjela prema godini osnivanja

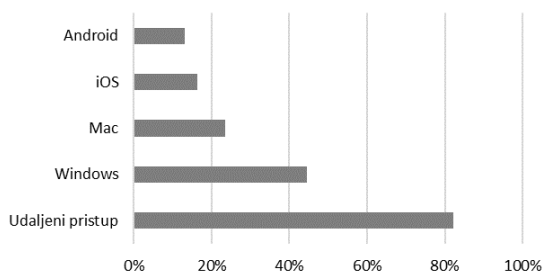
Iz slike je vidljivo da je u razdoblju od 2011. do 2015. godine otvoreno najviše tvrtki koje su razvile alate za vizualizaciju. U navedenom razdoblju se nalazi više od trećine dostupnih alata, preciznije 38%. Otprilike isti postotak tvrtki se nalazi u razdobljima od 2000. do 2014., odnosno 36%. Tvrtke koje su osnovane do 1999. razvile su 15% dostupnih alata.

3.2 Analiza osnovnih podataka

Među podacima koji su se prikupljali je i cijena, ali usporedbu alata na temelju nje je teško napraviti, budući da postoje razni modeli plaćanja: plaćanje po korisniku, plaćanje na mjesečnoj i godišnjoj bazi, te jednokratno plaćanje, ali i kombinacije navedenih varijanti. Dva najčešća modela su plaćanje na mjesečnoj bazi i jednokratno plaćanje. Cijene se za oba modela kreću od nekoliko američkih dolara do 10.000 dolara.

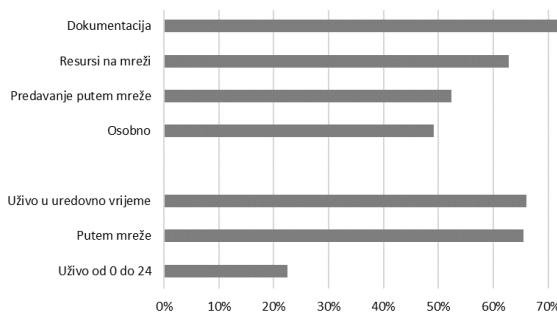
38% alata dolazi u nekoj od besplatnih varijanti, odnosno mogu se koristiti ograničeno vrijeme sa svim funkcionalnostima (engl. *demo*) ili se mogu koristiti neograničeno vrijeme s dijelom funkcionalnosti (engl. *trial*).

Kao što je vidljivo iz slike 4., najviše alata je dostupno putem mreže (82%), što uključuje standardne web aplikacije (engl. *web application*), aplikacije u oblaku (engl. *cloud application*) i softver kao uslugu (engl. *Software as a Service*, SaaS). Od klasičnih aplikacija (engl. *desktop application*) ima najviše aplikacija na Windows operacijskom sustavu (45%), dok su u mobilnim operacijskim sustavima podjednako zastupljeni Android s 13% i iOS s 16%.



Slika 4. Podržane platforme

Analiza dostupnih vrsta edukacija pokazuje da su najzastupljeniji korisnički priručnici, odnosno dokumentacija koja dolazi uz sam alat. Nakon toga slijede resursi poput mrežnih stranica, foruma i najčešćih pitanja. Za otprilike 50% alata dostupna su predavanja putem mreže (engl. *webinar*) i klasične radionice, kao što je vidljivo iz slike 5.



Slika 5. Dostupne vrste edukacije i korisničke pomoći

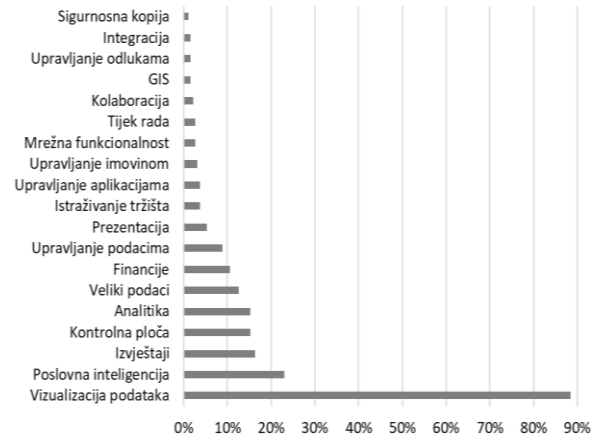
Iz slike 5 je također vidljivo da je korisnička podrška dostupna uglavnom uživo, za vrijeme standardnog radnog vremena ili putem mreže, dok samo 23% alata ima neprekidnu podršku uživo.

3.2 Analiza značajki

S mrežnih stranica alata prikupljeno je 1163 različitih značajki koje su bile razvrstane u 117 kategorija. Za alate je prikupljeno ukupno 4657 značajki, odnosno u prosjeku 24 značajke po alatu. Budući da su se neke od značajki pojavljivale više puta u raznim kategorijama, sve značajke su ponovno kategorizirane, kako bi situacija odgovarala stvarnom stanju. Na primjer, značajka OLAP (Online Analytical Processing) je kod nekih alata bila pridružena kategoriji Vizualizacija podataka, a kod nekih kategoriji Izvještaji. Nakon prekategoriizacije su preostale 53 kategorije.

Slika 6 prikazuje kategorije i postotak alata koji imaju pridruženu barem jednu značajku u navedenoj kategoriji. Očekivano, najzastupljenija kategorija je Vizualizacija podataka (84% alata), koja sadrži 6 značajki:

1. Upravljanje sadržajem (engl. *Content management*)
2. Filtrirani pogledi (engl. *Filtered views*)
3. OLAP
4. Relacijski prikaz (engl. *Relational display*)
5. Modeli simulacije (engl. *Simulation models*)
6. Vizualno pronalaženje (engl. *Visual discovery*)



Slika 6. Zastupljenost kategorija značajki

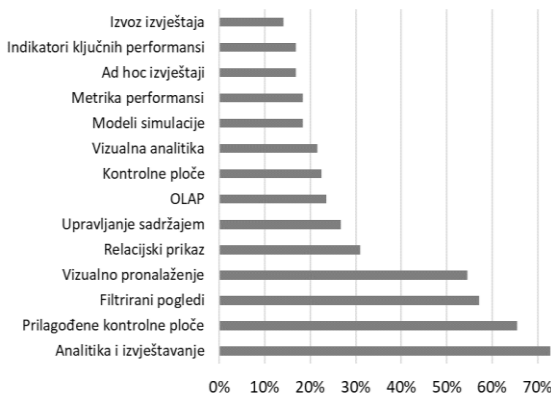
Sljedeća najzastupljenija kategorija je Poslovna inteligencija (24% alata) u kojoj dominiraju sljedeće značajke:

1. Ad hoc izvještaji (engl. *Ad hoc reports*)
2. Vrednovanje (engl. *Benchmarking*)
3. Indikatori ključnih performansi (engl. *Key performance indicators*)
4. Strateško planiranje (engl. *Strategic planning*)
5. Indikatori trenda i problema (engl. *Trend & problem indicators*)
6. Tablice rezultata (engl. *Scorecards*)

Ostale kategorije koje su zastupljene u više od 10% alata obuhvaćaju značajke vezane uz izradu izvještaja, rad s kontrolnom pločom, analitiku, podršku za velike podatke i podršku za financije.

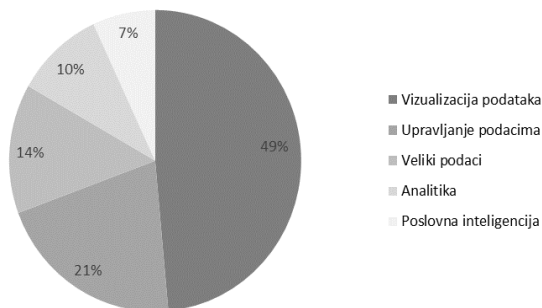
Graf na slici 7 prikazuje najzastupljenije značajke alata, bez obzira kojoj kategoriji pripadaju, ali zbog prikaza nisu navedene one koje su prisutne u manje od 10% alata. Iz slike je vidljivo da je najzastupljenija značajka Analitika i izvještavanje (engl. *Analytics & Reporting*), koju uključuje više od 70% alata. Prilagođene

kontrolne ploče, Filtrirani pogledi i vizualno pronalaženje su značajke prisutne u više od 50% alata.



Slika 7. Najzastupljenije značajke

Slika 8 prikazuje rezultate analize alata za vizualizaciju koji su prema mišljenju korisnika najbolji, odnosno kod kojih je ukupna prosječna ocjena veća ili jednaka 4.9. Graf prikazuje postotke zastupljenosti pojedinih kategorija u takvim alatima.



Slika 8. Kategorije najviše ocijenjenih alata

Iz slike je vidljivo da je kod takvih alata utvrđena nešto drugačija raspodjela od one prikazane na slici 6. Ako se u obzir uzmi svi alati, kategorija Upravljanje podacima je na osmom mjestu prema zastupljenosti, dok je kod boljih alata na drugom mjestu. Također, više ocjene su dobili oni alati kod kojih je zastupljenija kategorija Veliki podaci.

4 Zaključak

U radu su napravljene analize alata i njihovih karakteristika iz kojih je vidljiv presjek mogućnosti danas dostupnih alata za vizualizaciju, te su prikazane njihove najzastupljenije značajke i kategorije značajki. Iz navedenih analiza je vidljivo da alati uglavnom nude osnovnu funkcionalnost vezanu uz izvještaje, filtriranje i korisničko sučelje, a ne omogućuju rad s velikim podacima, što je u današnje vrijeme vrlo velik nedostatak, niti ne ispunjavaju zahtjeve naprednijih korisnika, poput prediktivne analitike, otkrivanja uzoraka i automatskog upozoravanja u slučaju prethodno određenih stanja. Također, vidljivo je da je potreban dodatan fokus na moderne metode vizualizacije, koje uključuju umjetnu inteligenciju, virtualnu ili proširenu stvarnost i analizu u stvarnom vremenu.

Sljedeći korak u istraživanju je prikupljanje podataka o broju korisnika navedenih alata, budući da mrežna stranica koja je korištena za istraživanje ne sadrži navedeni podatak. Time bi se mogla napraviti analiza, odnosno korelacija, između broja korisnika i pojedinih karakteristika alata, te na taj način utvrditi u čemu se razlikuju najpopularniji alati od ostalih alata.

Također, obzirom da se vizualizacija podataka nadovezuje na analizu i obradu podataka, provest će se istraživanje koje uspoređuje karakteristike alata iz navedene tri kategorije.

Literatura

Ahn, J. W., Plaisant, C., & Shneiderman, B. (2013). A task taxonomy for network evolution analysis. *IEEE transactions on visualization and computer graphics*, 20(3), 365-376.

Ali, S. M., Gupta, N., Nayak, G. K., & Lenka, R. K. (2016, December). Big data visualization: Tools and challenges. In *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)* (pp. 656-660). IEEE.

Capterra (2019). Retrived from <http://www.capterra.com>

- Chen, M., Ebert, D., Hagen, H., Laramée, R. S., Van Liere, R., Ma, K. L., ... & Silver, D. (2008). Data, information, and knowledge in visualization. *IEEE Computer Graphics and Applications*, 29(1), 12-19.
- Chi, E. H. H. (2000, October). A taxonomy of visualization techniques using the data state reference model. In *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings* (pp. 69-75). IEEE.
- Daassi, C., Nigay, L., & Fauvet, M. C. (2005). A taxonomy of temporal data visualization techniques. *Information-Interaction-Intelligence*, 5(2), 41-63.
- Donalek, C., Djorgovski, S. G., Cioc, A., Wang, A., Zhang, J., Lawler, E., ... & Davidoff, S. (2014, October). Immersive and collaborative data visualization using virtual reality platforms. In *2014 IEEE International Conference on Big Data (Big Data)* (pp. 609-614). IEEE.
- Few, S., & Edge, P. (2007). Data visualization: past, present, and future. *IBM Cognos Innovation Center*.
- Friendly, M. (2008). A brief history of data visualization. In *Handbook of data visualization* (pp. 15-56). Springer, Berlin, Heidelberg.
- Gorodov, E. Y. E., & Gubarev, V. V. E. (2013). Analytical review of data visualization methods in application to big data. *Journal of Electrical and Computer Engineering*, 2013, 22.
- Kyubit Business Intelligence (2019). Retrieved from <https://www.kyubit.com/>
- Liu, K., Zhou, X., & Zhou, D. (2002). Data Visualization Research and Development [J]. *Computer Engineering*, 8.
- Michaels, G. S., Carr, D. B., Askenazi, M., Fuhrman, S., Wen, X., & Somogyi, R. (1998). Cluster analysis and data visualization of large-scale gene expression data. In *Pacific symposium on biocomputing* (Vol. 3, pp. 42-53).
- Nagel, H. R. (2006). Scientific visualization versus information visualization. In *Workshop on State-Of-The-Art in Scientific and Parallel Computing*.
- Parbat, B., Dwivedi, A. K., & Vyas, O. P. (2010). Data visualization tools for WSNs: a glimpse. *International Journal of Computer Applications*, 2(1), 14-20.
- Robertson, G. G., Mackinlay, J. D., & Card, S. K. (1991, March). Cone trees: Animated 3d visualizations of hierarchical information. In *CHI* (Vol. 91, pp. 189-194).
- Sadiku, M., Shadare, A. E., Musa, S. M., & Akujuobi, C. M. (2016). Data Visualization. *International Journal of Engineering Research And Advanced Technology (IJERAT)*, 2(12), 11-16.
- Shneiderman, B. (1996, September). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE symposium on visual languages* (pp. 336-343). IEEE.
- Sucharitha, V., Subash, S. R., & Prakash, P. (2014). Visualization of big data: its tools and challenges. *International Journal of Applied Engineering Research*, 9(18), 5277-5290.
- Tory, M., & Moller, T. (2004, October). Rethinking visualization: A high-level taxonomy. In *IEEE Symposium on Information Visualization* (pp. 151-158). IEEE.
- Wang, L., Wang, G., & Alexander, C. A. (2015). Big data and visualization: methods, challenges and technology progress. *Digital Technologies*, 1(1), 33-38.

Characteristics of data visualization software

Abstract

This paper defines data visualization, its importance primarily in science and modern business processes, but also in other field such as

medicine and banking. Analysis and visualization of large amounts of data is done today with a variety of tools, which vary in price, platform, quality of customer support, capabilities and functionality. The paper analyzes and compares the characteristics of a large number of data visualization tools in order to find their most common capabilities, that is, the common functionality of most of today's tools. In addition to this analysis, the paper also analyzes their availability on platforms and operating systems, customer support, pricing etc.

Keywords: *data visualization; data visualization tools; tool analysis; tool comparison.*