

INFORMATION DENSITY AS A SUBJECTIVE AND OBJECTIVE CATEGORY

GUSTOĆA INFORMACIJA KAO SUBJEKTIVNA I OBJEKTIVNA KATEGORIJA

ĐOKIĆ, Kristian

Abstract: *It can be noted in the scientific literature that the term information density is used in different contexts and with different meanings. This paper proposes to divide these meanings into two categories, subjective and objective. An overview of the papers of authors from different fields of science and their definitions/meanings of information density is given. Finally, a general definition is given that covers all of the above.*

Key words: *information density, information, entropy*

Sažetak: *U znanstvenoj literaturi se može primjetiti da se pojam gustoće informacija koristi u različitim kontekstima i sa različitim značenjima. U radu je predložena podjela tih definicija u dvije kategorije, subjektivnu i objektivnu. Dan je pregled radova autora sa različitih područja znanosti i njihovih definicija/značenje gustoće informacija. Na kraju je dana jedna općenita definicija koja obuhvaća sve navedene.*

Ključne riječi: *gustoća informacija, informacija, entropija*



Authors' data: Kristian, Đokić, dr.sc., Veleučilište u Požegi, Vukovarska 17, Požega, kdjokic@vup.hr

1. Uvod

Pojam „gustoće informacija“ u literaturi nije jednoznačno definiran, a analizom se može utvrditi da bi jedna od podjela mogla biti na subjektivne i objektivne pristupe mjerenju i samoj definiciji. U ovom radu će biti predstavljene različite definicije gustoće informacija sa različitih područja znanosti. Krenuti će se od definicija informacije u drugom poglavlju, a potom na pojašnjenje pojma gustoće informacija u trećem poglavlju. Pri tome će se kao što je rečeno koristiti radovi autora koji su pojam spominjali u različitim kontekstima. Konačno, zaključak je u četvrtom poglavlju.

2. Definicija informacije

Pojam informacije definiran je u *Webster* engleskom rječniku na više načina, a između ostalog i kao „znanje dobiveno istraživanjem, proučavanjem ili uputom“ [1]. U *Cambridge Dictionary of Philosophy* informacija je definirana kao „objektivni entitet koji može biti generiran ili prenesen porukom (riječju, rečenicom) ili drugim načinom kognicije (interpretacije). Informacija može biti kodirana i prenesena, ali informacija postoji nezavisno od kodiranja i prijenosa.“ [2].

Informaciju povezujemo sa smanjivanjem neizvjesnosti. Neizvjesnost događaja se mjeri vjerojatnošću njegovog događanja i obrnuto je proporcionalna toj vrijednosti. Autor fundamentalnog rada kojim je postavio temelje teorije informacija je C. E. Shannon sa radom „*A Mathematical Theory of Communication*“ iz 1948. godine. Autor je u radu definirao neizvjesnost, odnosno entropiju diskretne slučajne varijable i to kao u formuli 1:

$$H = - \sum_{i=1}^n p_i \log p_i \quad (1)$$

pri čemu slovo H označava entropiju događaja čije su vjerojatnosti p_1, \dots, p_n . [3]. Na taj način entropijom se mjeri gubitak informacija o stanju sustava.

Shannon nekoliko desetljeća nakon objave navedenog rada navodi da su mnogi autori na različite načine definirali pojam informacije, te je svjestan da se teško može očekivati da bi jedna univerzalna definicija mogla zadovoljiti sve pristupe tom pojmu. Isto tako vjeruje da će neke od definicija i koncepata drugih autora biti priznate i biti temelj za dalja istraživanja i priznanje znanstvene zajednice [4].

Nužnost postojanja različitih pristupa prijenosu informacija uočio je Weaver nedugo nakon objave spomenutog rada Shanona iz 1948. godine. Weaver je predložio da se osim pristupa koji je koristio Shanon, pojam prijenosa informacija sagleda na još dva načina. Treba naglasiti da se autor u svom radu usredotočio na pojam komunikacije, ali s ciljem prijenosa informacija. Osim tehničkog aspekta komunikacije, odnosno prijenosa informacija, Weaver je uočio da se problem treba sagledati i sa aspekta semantike, kao i sa aspekta utjecaja. Pri tome pod semantičkim problemima prijenosa informacija prije svega prepoznaje problem tumačenja značenja primljenih informacija primatelja u usporedbi sa značenjem koje su te informacije imale kod pošiljatelja. Autor navodi govornu komunikaciju u kojoj se teži semantičke teškoće

smanjiti na prihvatljivu razinu, a jedna od osnovnih pretpostavki za to je da su pojmovi u jeziku prethodno jasno definirani. Treći problem se tiče utjecaja ili učinkovitosti, odnosno očekuje se da primljena informacija kod primatelja izazove određeno ponašanje. Autor navodi da sama ta izjava ima negativne implikacije, mada je u širem smislu razumljivo da primanje informacija izaziva, odnosno ne izaziva promjenu u ponašanju primatelja. Isti autor je razjasnio da pojam informacije u Shannonovom modelu ne govori što je preneseno, nego koliko može biti preneseno informacija po komunikacijskom kanalu, pri čemu naglašava da se sam pojam informacije ne koristi u smislu značenja kao što je uobičajeno. Ključna podjela je poimanje informacije s tehničkog, odnosno matematičkog aspekta, odnosno poimanje informacije s aspekta semantike [5].

3. Definicija gustoće informacija

S obzirom na prethodno navedene definicije informacije, za očekivati je da ne postoji jednoznačna definicija gustoće informacija. Navedeni kriterij podjele na matematičko tehnički, odnosno semantički aspekt informacija ima utjecaj i na poimanje gustoće informacija. Slična podjela na objektivnu i subjektivnu gustoću informacija koristit će se dalje u radu, pri čemu je ključan kriterij podjele prisutnost, odnosno odsutnost subjektivne procjene.

3.1. Objektivna gustoća informacija

Pod objektivnom gustoćom informacija bit će navedene definicije i tumačenja pojma kada u procesu mjerenja nije uključen ljudski faktor i subjektivnost.

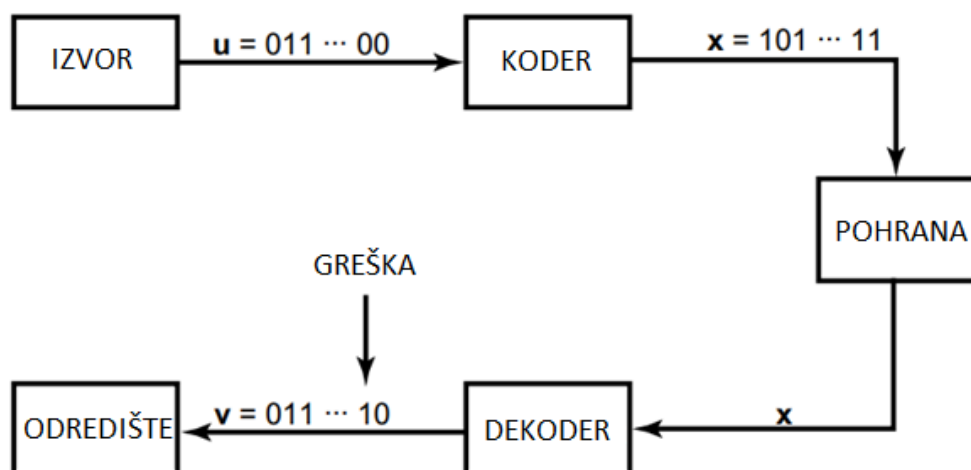
Gustoća informacija često se povezuje s medijima za pohranu podataka jer se na njima pohranjuju podaci različite gustoće najčešće s obzirom na iskorištenu površinu. Pri pohrani se često koristi sažimanje podataka s ciljem efikasnijeg korištenja samog medija, ali sažimanje se koristi i pri prijenosu podataka s ciljem efikasnijeg korištenja prijenosnog kanala.

Slika 1 prikazuje općenit proces sažimanja koji uključuje sažimanje bez i sa gubitkom. Podatci iz izvora se dijele u blokove i sažimaju se nekim algoritmom, te pohranjuju na nekom mediju. U idealnom slučaju kada nema gubitka podataka iz sažetih podataka se dobivaju izvorni. U slučaju da se radi o sažimanju s gubitkom, može doći do razlike između izvornih podataka i podataka dobivenih iz sažetih podataka [6].

Razina sažimanja može se izračunati iz izraza 2:

$$R = n / k, \quad (2)$$

pri čemu je n prosječan broj bitova koji su potrebni za pohranu k izvornih bitova. Što je R manji, sažimanje je efikasnije, a entropija veća [6].



Slika 1 -Shematski prikaz sažimanja podataka [6]

Ipak, potrebno je razlučiti terminološku razliku između sažimanja podataka i gustoće informacija, pri čemu je sažimanje vezano uz način i efikasnost kodiranja podataka, dok je entropija mjera količine informacija koju sadrže podatci. Očigledno je da između entropije i gustoće informacija postoji veza, ali entropija je egzaktan pojam u informacijskim znanostima dok pojam gustoće informacija autori koriste na različite načine što se može vidjeti u nastavku.

Pojam gustoće informacija spominje se u pravilu konstantne gustoće informacija (eng. *Principle of Constant Information Density*) koje se primjenjuje u kartografiji kod inteligentnog zumiranja. U kartografiji je to pravilo postavio Topfer pri čemu je isto izrazio i formulom broj 3:

$$n_f = n_a * \sqrt{m_a/m_f} \quad (3)$$

pri čemu je m_a trenutno mjerilo karte, m_f sljedeće mjerilo karte, n_a broj objekata trenutnog mjerila i n_f broj objekata sljedećeg mjerila. Ako se to pravilo primjeni na vidljivo područje karte umjesto na mjerilo karte, broj objekata po prikazanoj površini treba biti konstantan. Gustoćom informacija autori smatraju broj objekata prikazanih na vidljivom području [7][8].

Jednu od sličnih definicija također na području prikaza objekata predložio je Jahnke pri čemu ju je bazirao na broju svojstava po jedinici površine. Njegova definicija vezana je uz područje trodimenzionalnih modela u kartografiji, pri čemu se vidi da u realizaciji autor pod svojstvima smatra detalje tekstura trodimenzionalnih objekata [9].

U svom radu Perrin spominje gustoću informacija, ali je ne pokušava kvantitativno izraziti. S obzirom da se rad bavi analizom komunikacije uz pomoć višestrukih slika, autor navodi da je gledatelj kojem se istovremeno prikazuju višestruke slike ili video sadržaji izložen većoj gustoći informacija od gledatelja kojem se prikazuje samo jedna slika ili video sadržaj. U radu se analiziraju prednosti i nedostaci metode, a sama definicija je vrlo slična prethodno opisanoj mada se koristi u potpuno drugačijem kontekstu. Naime, u radu se analizira primjena višestrukih istovremeno prikazanih slika u obrazovanju [10].

Pojam gustoće informacija koriste Chang i suradnici u radu u kojem testiraju utjecaj različitih metoda prezentacije informacija (istovremeno i uzastopno) na sposobnost

traženja i pamćenja. Pri tome su koristili matricu od četiri stupca i četiri retka u kojoj su prikazivali četiri, osam i dvanaest točaka. Autori su gustoću informacija izražavali u postotku, kao 25%, 50% i 75%, te je očigledno da autori smatraju da veći broj objekata, konkretno točaka na ekranu linearno utječe na povećanje gustoće informacija [11].

Pojam gustoće informacija nalazimo u radovima koji se bave korisničkim sučeljem. Gabillon i suradnici proučavali su korisnička sučelja općenito, te smatraju da se gustoća informacija odnosi na radno opterećenje korisnika sa kognitivnog aspekta i aspekta percepcije i pri tom se u obzir treba uzeti cjelovit niz informacija prezentiranih korisniku. Model koji su predložili sastoji se od jedanaest komponenti čije vrijednosti se izražavaju brojčano pa se mjerenje može automatizirati te se na taj način izbjegava subjektivan utjecaj dizajnera pri izradi sučelja [12].

Ziefle također spominje pojam gustoće informacija u svom radu u kojem analizira utjecaj gustoće informacija i veličine znakova na brzinu navigacije na mobilnim telefonima. Ovaj rad se također bavi korisničkim sučeljem i pri tome autorica koristi gustoću informacija kao zavisnu varijablu u istraživanju. Taj pojam autorici označava broj mogućnosti u izborniku koji u slučaju analiziranog generičkog sučelja može poprimiti pet vrijednosti, odnosno broj pojmova u izborniku može biti između jedan i pet [13].

Hammouda se također bavio analizom sučelja u svom radu, te je metriku vezanu uz gustoću informacija sučelja podijelio u dvije kategorije s obzirom na vrstu sučelja kao što se vidi u tablici broj 1. Autor navodi da je glavni problem gustoće vizualnih informacija neravnomjerna distribucija podataka, te nudi drugačiji pogled na problem iz perspektive ekologije [14].

		Vrsta sučelja	
		Grafičko	Tekstualno
Metrika		vrhovi/jedinici površine	znakova/jedinici površine
		objekti/jedinici površine	riječi/jedinici površine
		dijelovi pravaca/jedinici površine	
		popunjene točke/jedinici površine	

Tablica 1– Metrika gustoće informacija [14]

3.2. *Subjektivna gustoća informacija*

Pod subjektivnom gustoćom informacija su navedene samo dvije definicije koje u procesu mjerenja uključuju ljudski faktor i subjektivnost u nekoj mjeri. Mada neki od autora egzaktno definiraju mjeru gustoće informacija, ako pri mjerenju sudjeluje osoba moguće su razlike u mjerenju zbog različitog tumačenja pojmova.

Gustoću informacija Chu i Li definiraju na drugačiji način od prethodno navedenih. U svom radu u kojem uspoređuju gustoću informacija mobilnih web trgovina autori definiraju istu kvantitativno koristeći dužinu web stranice i ukupan broj objekata na web stranici. Dužinu stranice izražavaju kao relativan broj u odnosu na predloženu konstantnu razlučivost ekrana dok objektima koji se broje na web stranici smatraju poveznice, slike, video isječke, naslove, odlomke i natpise. Uz pomoć navedenih

kvantitativnih mjera računaju omjer istih i dobivaju novu mjeru koju nazivaju „gustoćom prikaza“. Subjektivnost koja nastaje pri analizi i brojenju objekata je razlog zašto je ova metoda svrstana u subjektivne [15].

Gustoću informacija možda je najobuhvatnije definirala A. Lang u sklopu Modela ograničenog kapaciteta motivirane obrade posredovanih poruka (eng. *Limited Capacity Model of Motivated Mediated Message Processing*) kao količinu novih uvedenih informacija po svakoj poruci. Mjerom uvedenih informacija (eng. *information introduced*) mjeri se količina informacija uvedena svakim pomakom kamere i to na skali od sedam različitih varijabli, odnosno dimenzija. Te dimenzije su: promjena žarišnog objekta, novi žarišni objekt, povezanost, udaljenost, perspektiva, emocija i izmjena oblika.

Svaka od dimenzija može poprimiti vrijednosti 0 ili 1. Analizu i mjerenje vrše dvije osobe koje su prethodno obučene i čija točnost je veća od 97%. Gustoća informacija za pojedinu poruku izračunava se kao zbroj prethodno navedenih dimenzija koje su zabilježene u poruci, odnosno video isječku, podijeljene sa dužinom poruke (brojem sekundi). Bez obzira na navedeni kriterij točnosti i prethodno testiranje osoba koje vrše mjerenje, očigledno je da i ovdje može doći do određenog odstupanja uzrokovanog subjektivnošću opažača [16].

4. Zaključak

Bez obzira na različite pristupe te navedenu podjelu na subjektivnu i objektivnu gustoću informacija, može se zaključiti da je gustoća informacija količina informacija na različitim razinama apstrakcije mjerena u vremenskoj, odnosno prostornoj domeni. Definicija je dovoljno općenita da obuhvati i interpretaciju gustoće informacija u slučaju subjektivne procjene.

Ovaj pregled radova sigurno nije obuhvatio sve autore koji su se bavili gustoćom informacija, ali je dao pregled koji ukazuje da se radi o općenitom pojmu koji nije jasno definiran i pri korištenju istog treba biti iznimno pažljiv i obrazložiti kontekst na koji se pojam odnosi.

Dalji rad na tom području trebao bi uključivati još više različitih interpretacija pojma u različitim područjima i poljima znanosti, ali predložena podjela bi vrlo vjerojatno ostala ista. Ključno svojstvo koje podjelu čini sveobuhvatnom za nove interpretacije je visoka razina apstrakcije.

5. Literatura

- [1] Webster, Definition of information, <https://www.merriam-webster.com/dictionary/information> (posjećeno: 07. siječnja, 2017.)
- [2] Floridi, L., (2005), Is Semantic Information Meaningful Data?, *Philosophy and Phenomenological Research*, Vol. LXX, No. 2, March 2005
- [3] Shannon, C.E. (1948), A Mathematical Theory of Communication, *The Bell System Technical Journal*, Vol XXVII, No.2, July, 1948.
- [4] Shannon, C.E. (1993), *Collected Papers*, Editors: Sloane, N.J.A., Wyner, A.D., Wiley-IEEE Press, 1993.

- [5] Weaver, W., (1949), “The Mathematics of Communication”, *Scientific American*, 181(1): 11–15.
- [6] Whitaker, J.C., (2005), *The Electrical Engineering Handbook* Second edition, Taylor & Francis Group, Boca Raton, FL, SAD
- [7] Topfer, F., Pillewitzer, W. (1964), Das Auswahlgesetz, ein Mittel zur kartographischen Generalisierung. *Kartographische Nachrichten* 14, 117-121 (1964).
- [8] Frank, A.U., Timpf, S., (1994), Multiple representations for cartographic objects in a multi-scale tree—An intelligent graphical zoom, *Computers & Graphics*, Volume 18, Issue 6, November–December 1994, Pages 823-829, DOI:10.1016/0097-8493(94)90008-6
- [9] Jahnke, M., Krisp, J.M., Kumke, H., (2011), How Many 3D City Models Are There? – A Typological Try, *The Cartographic Journal* Vol. 48 No. 2 pp. 124–130 International Cartographic Conference, Paris 2011-Special Issue May 2011.
- [10] Perrin, D.G., (1969), A Theory of Multiple-Image Communication, *AV Communication Review*, Vol. 17, No. 4 (Winter, 1969), pp. 368-382, Springer
- [11] Chang T-W, Kinshuk, Chen, N-S, Yu P-T (2012) The effects of presentation method and information density on visual search ability and working memory load, *Computers & Education* 58 (2012) 721–731,
- [12] Gabillon, Y., Lepreux, S., Oliveira, K.M., (2013), Towards Ergonomic User Interface Composition: A Study about Information Density Criterion, *Human-Computer Interaction, Part I, HCII 2013, LNCS 8004*, pp. 211–220, 2013., Springer-Verlag Berlin Heidelberg 2013
- [13] Ziefle, M., (2010), Information presentation in small screen devices: The trade-off between visual density and menu foresight, *Applied Ergonomics* 41 (2010) 719-730, Elsevier, doi:10.1016/j.apergo.2010.03.001
- [14] Hammouda, K., (2000), An Ecological Perspective of Visual Information Density in Interface Design, Department of Systems Design Engineering, University of Waterloo, Ontario, Canada, <http://pami.uwaterloo.ca/pub/hammouda/sde744-paper.pdf> (pristup: 22.12.2016.)
- [15] Chu, J., Li, M., (2012), A Culture-based Study on Information Density of Mobile E-commerce Websites, 2012 Fifth International Symposium on Computational Intelligence and Design, 28-29. October 2012, Volume 2, Hanzhou, China, 2012. DOI: 10.1109/ISCID.2012.218
- [16] Fox, J.R., Park, B., Lang, A., (2007), When Available Resources Become Negative Resources: The Effects of Cognitive Overload on Memory Sensitivity and Criterion Bias, *Communication Research*, 2007 34: 277, DOI: 10.1177/0093650207300429



Photo 021. Kruh / Bread