

O PODATKU, INFORMACIJI IN ZNANJU

Tomaž Mohorič
Fakulteta za računalništvo in informatiko, Univerza v Ljubljani
tomazm@fri.uni-lj.si

Povzetek

Izraza podatek in informacija se pogosto uporabljata kot sinonima, čeprav obstaja med njima precejšnja pomenska razlika. Podobno se zadnje čase dogaja tudi z izrazom znanje, ki se ponekod uporablja v pomenu podatek. Namen prispevka je razčistiti pomenske razlike med omenjenimi izrazi. V prvem delu je poglobljeno predstavljen način človekovega dojemanja, razumevanja in predstavljanja sveta, v drugem delu pa so podane nekatere standardne definicije podatka, informacije in znanja.

Abstract

The terms data and information are often used as a synonyms in spite of quite a big semantic difference between them. Similar destiny shares the term knowledge, which is often used in place of term data. The intention of this article is to advocate that each term mentioned has its own meaning. First, the human way of making perceptions, conceptions, and representations is thoroughly described, followed by some standard definitions of data, information, and knowledge.



Uvod

Izraza podatek in informacija se v vsakodnevem življenju pogosto uporabljata kot sinonima. Tudi na računalniškem področju med njima ni vedno potegnjene stroge ločnice. Do prekrivanja pomenov prihaja očitno zaradi enačenja procesov v človekovi glavi s procesi v računalniku. Tako govorimo o računalniški obdelavi *podatkov*, pa tudi o računalniški obdelavi ali procesiranju *informacij*. Marsikje lahko tudi preberemo, da se, na primer, vsebina tabele z numeričnimi podatki pri preslikavi v graf spremeni v informacijo.

Podobno je z znanjem. Kaj imajo ljudje v svojih glavah, kadar vedo nekaj? Ali lahko znanje izrazimo z besedami? Kako lahko nekdo zna stvari, ki jih je lažje narediti kot pa opisati – na primer, kako se zaveže vezalka na čevlju? Če znanja ni moč izraziti z besedami, kako ga lahko komu posredujemo z besedami? V kakšni zvezi sta si znanje in svet? Kakšna so razmerja med zunanjim svetom, znanjem v glavi in jezikom, uporabljenim za izražanje znanja o svetu? Ali lahko znanje upravljamo? Očitno ga lahko, saj v okviru informacijskih sistemov obstaja disciplina, ki se ukvarja z njegovim upravljanjem. Pa je res znanje tisto, kar upravljamo? Kakšne misli se nam pode po glavi tačas, ko upravljamo znanje? Kot vse kaže, se da zastaviti veliko vprašanj, morebiti pa se da na kakšno tudi odgovoriti.

1. ČLOVEK IN NJEGOVO DOJEMANJE SVETA

Kako človek dojema svet, katerega del je tudi sam? Odgovor na to vprašanje ni preprost in tudi ne vemo,

ali bomo kdaj zvedeli za pravi odgovor. Pri iskanju odgovora si bomo pomagali z vrsto predpostavk, ki jih srečamo v [3] in [5]. Prva izmed njih je naslednja:

- svet obstaja neodvisno od nas oziroma naših spoznavnih sposobnosti.

Posamezniku je težko verjeti, da je obstajal svet pred njim in da bo obstajal tudi po njem. Zelo rad verjame, da svet brez njega ni tisti pravi svet. Pravimo tudi, da živimo v svojem svetu, ki ga seveda tudi vidimo po svoje. Za potrebe našega prispevka se moramo privaditi na misel, da svet od posameznika ni prav veliko odvisen.

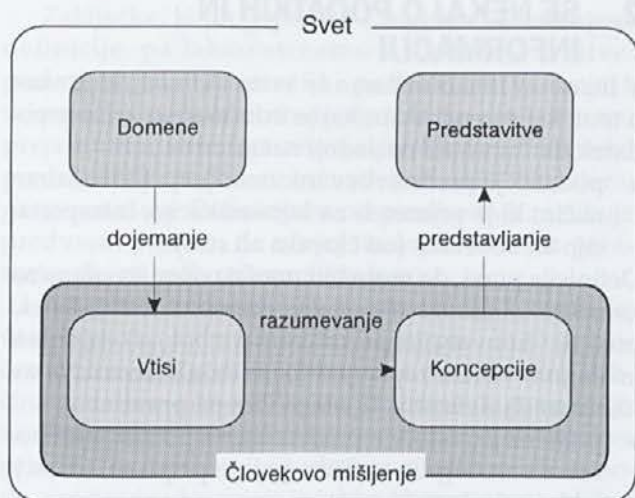
Človek pa v svetu živi, ga spoznava in se odziva na spremembe v njem:

- človek je sposoben s pomočjo svojih čutil opazovati in dojemati vtise o posameznih delih ali aspektih sveta, imenovanih tudi domene. Dojeto domeno razume človek kot množico prepoznavnih stvari, ki pa se lahko prekrivajo, vsebujejo druga drugo ali so druga z drugo v kakršnikoli povezavi (S1).

Posledica dojemanja sveta so vtisi, ki jih človek oblikuje v svojih mislih. Vtisi so lahko urejeni ali neurejeni, bežni, površni in tudi taki, ki na človeka naredijo močan vtis:

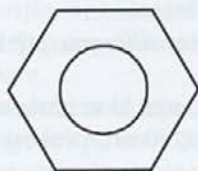
- na osnovi dojetih vtisov iz sveta oblikuje človek v svoji glavi koncepcije (ideje, modele), pri čemer mu pomagajo mentalni procesi, kot so prepoznavanje, karakterizacija, abstrakcija, refleksija.

Omenjeni procesi mu omogočajo opazovani svet spoznati, doumeti in razumeti. Pri tem lahko zaide tudi v težave. Dojeti vtisi ga lahko postavijo pred dilemo,



S1. Dojemanje, razumevanje in predstavljanje delov sveta

katera konceptija je tista prava. Ali naj, na primer, vtis razume kot matico ali pa kot vijak (S2).



S2. Katera konceptija je prava - ali je matica ali vijak?

V človekovi glavi se tako zbere množica različnih konceptij. Vsaka od njih prispeva k bolj poglobljenemu poznavanju sveta:

- množica človekovih stabilnih in med seboj skladnih konceptij sestavlja njegovo znanje.

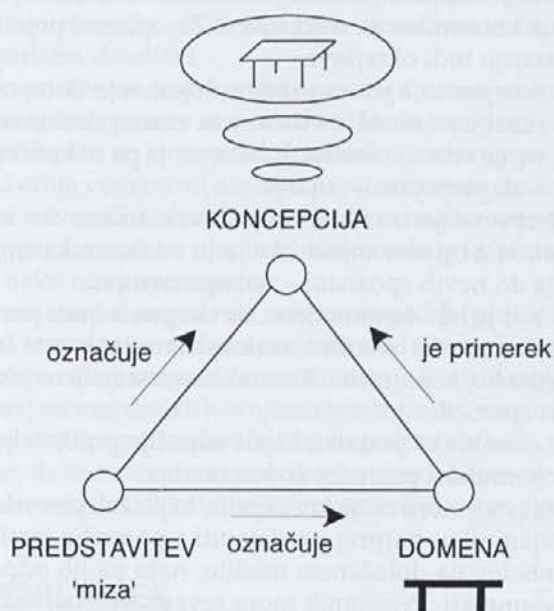
Pojav stabilnih konceptij v človekovi glavi je lahko posledica opazovanja in razumevanja sveta in / ali rezultat komunikacije s sočlovekom, ki se nanaša na ta isti svet. Vsaka taka konceptija je del človekovega 'znanja o svetu'.

Za posredovanje svojih konceptij sočloveku uporablja človek predstavitev.

- Konceptije se predstavijo s pomočjo nekega jezika na nekem pomnilniškem mediju (papir, računalniški pomnilnik, ...).

Konceptija predstavlja abstraktno idejo, kaj neka stvar je, kaj pomeni, kakšne vrste je, s kakšnimi lastnostmi jo lahko opišemo. Konceptiji lahko pripadajo objekti domene – primerki konceptije, ki posedujejo prav take lastnosti, kot jih konceptija predstavlja. Če želimo govoriti (pisati) o konceptijah in objekti domene, jih je potrebno označiti, za kar se uporabljajo predstavitve –

simboli. Konceptije, predstavitve in domene lahko predstavimo s pomenskim trikotnikom (S3).



S3. Pomenski trikotnik

Slika S3 prikazuje pomenski trikotnik (znan tudi kot Ogdenov predstavitveni trikotnik), katerega vrhnje oglišče je konceptija, ki je predstavljena s predstavitvijo 'miza' (podatek), primerki konceptije pa so elementi domene, ki obsega množico miz.

V posameznih primerih lahko pomenskemu trikotniku kakšno oglišče tudi manjka. Obstajajo lahko konceptije (npr. sedmeroglavi zmaj), katerih primerki ne obstajajo (domena je prazna množica). Pojavijo se lahko objekti, ki jih ne znamo klasificirati – ne poznamo njihovih konceptij. In končno si lahko izmislimo predstavitve, ki ničesar ne označujejo, vsaj ne da bi mi vedeli.

Predstavitve znanja o svetu obstaja izven človekove glave. To, kar človek ve in zna, je večidel sposoben tudi predstaviti soljudem. V ta namen izrazi svoje znanje s pomočjo jezika, besede in stavke pa zapiše s pomočjo simbolov:

- *podatek* (podatki) je poljubna množica predstavitev znanja, izraženih s pomočjo jezika in zapisanih s pomočjo simbolov.

Vsaka predstavitev postane tako del sveta, ki človeka obkroža. Predstavitev je lahko individualna, taka, ki jo razume le avtor predstavitve. Če pa naj nekaj pomeni tudi soljudem, se morajo le-ti z njo strinjati:

- s stališča skupnosti (skupine ljudi) je pomembno, da je posamezna domena predstavljena na način, s katerim vsi soglašajo – ne glede na 'resnično' realnost.

Taka 'realnost' skupine ljudi je poimenovana kot 'medsubjektivna realnost'. To je torej realnost, za katero se je skupina ljudi dogovorila in znotraj katere obstaja komunikacija med njimi. Za opisane pojme pa obstajajo tudi okrajšave:

- *interpretacija* je, na primer, dojemanje domene, ki rezultira v množici vtisov, tem vtisom sledi razumevanje vtisov, rezultat razumevanja pa so konceptije – ali preprosto – znanje.

Interpretacijo izvaja seveda človek, imenovan interpret, ki z opazovanjem okolja in s komunikacijo prihaja do novih spoznanj – novega znanja.

Kot je bilo že omenjeno, se skupnost ljudi praviloma dogovori o 'resnični' realnosti znotraj katere lahko med seboj komunicira. Komunikacija temelji na izmenjavi sporočil:

- *sporočilo* so podatki, ki jih odpošlje pošiljatelj prejemniku s posredovanjem medija.

Pošiljatelj mora svoje konceptije, ki jih želi posredovati prejemniku, najprej predstaviti s pomočjo jezika in simbolov na določenem mediju, nato pa jih odpošlje prejemniku. Prejemnik mora seveda poznati tako jezik, kot tudi simbole, s katerimi je sporočilo zapisano.

In kaj je informacija:

- *informacija* je inkrement znanja pridobljen s sprejemom sporočila. To je razlika med konceptijami, pridobljenimi z interpretacijo sprejetega sporočila, in znanjem prejemnika pred sprejemom sporočila.

Koliko znanja pridobi prejemnik z interpretacijo sporočila, je odvisno od marsičesa. Denimo, da je sporočilo zapisano s kitajskimi pismenkami. Povprečen Slovenec jih bo dojel kot simpatične sličice, razumel pa jih seveda ne bo. Olajšajmo si problem – pri zapisovanju sporočila je bila uporabljena latinica. Črke sedaj poznamo, vendar, če je besedilo zapisano v španščini, bo marsikomu vse skupaj še vedno 'španska vas'. Sporočila še vedno ne bomo znali interpretirati. Šele če razumemo jezik in poznamo simbole, s katerimi je sporočilo zapisano, bomo znali razbrati, kaj nam sporočilo sporoča.

Sporočilo torej znamo prebrati in ga razumemo. Za koliko se je povečalo naše znanje? Odvisno od tega, koliko smo že vedeli. Če se 'novica' nanaša na 'lanski sneg', ne bomo nič pametnejši. Če pa sporočilo sporoča pomembno, doslej neznano dejstvo, bo to pomenilo kar precejšen inkrement k našemu znanju. Prejeta informacija lahko torej poveča naše znanje, njena prava vrednost pa se pokaže tedaj, ko jo znamo učinkovito izkoristiti za sprejem odločitve, ki nam osebno ali pa širši skupnosti prinese koristi.

2. ŠE NEKAJ O PODATKIH IN INFORMACIJI

V literaturi lahko srečamo še vrsto definicij, ki govore o tem, kaj so podatki in kaj je informacija. V [1] je podatek definiran na naslednji način¹:

- podatek je predstavitev informacije na formaliziran način, ki je primeren za komunikacijo, interpretacijo ali obdelavo (od človeka ali stroja).

Definicija pravi, da mora biti predstavitev izvedena na *formaliziran način*, kar pomeni, da mora obstajati neki predpis – konvencija, po katerem simbole ali vrednosti analognih veličin zapisujemo oziroma beremo. Nekoliko starejša definicija iz istega vira pa pravi:

- podatek je poljubna predstavitev s pomočjo simbolov ali analognih veličin, ki ji je pripisan, ali se ji lahko pripiše neki pomen.

Iz definicije lahko izluščimo, da je podatek lahko *diskreten*, če se pri predstavitvi uporabljajo simboli (npr. 25° C), ali pa *analogen*, če se za predstavitev uporablja kakšna fizikalna veličina (npr. dolžina živosačrpnega stolpca).

Obema definicijama je skupno to, da se podatku lahko pripiše neki pomen na osnovi nekega predpisa znotraj nekega konteksta.

Kaj pa pomeni izraz *informacija*? Po ISO je pomen naslednji:

- informacija je znanje, ki se nanaša na objekte, kot so dejstva, dogodki, stvari, procesi ali ideje, skupaj s koncepti, ki imajo v okviru nekega konteksta določen pomen.

G.C. Everest [2] je zapisal definicijo nekoliko drugače, ki pa se po pomenu vsekakor ujema s predhodnico:

- informacija so ovrednoteni podatki v posamezni specifični situaciji.

Zadnji dve definiciji implicitno predpostavljata, da obstaja interpret, ki informacijo izlušči iz podatkov. Kot pa smo že omenili, informacija ni nekaj absolutnega in nekaj takega, kar bi bilo vsebovano v podatkih, pač pa je njena količina močno odvisna od obstoječega interpretovega znanja.

Avtor ene izmed zelo znanih definicij informacije je Börje Langefors [4]. Definicija je znana tudi pod imenom *informacijska enačba*. Langefors pravi naslednje:

- Informacija je *ново spoznanje*, ki ga človek (interpret) doda svojemu poznavanju sveta. Odnos med informacijo, podatki, časom in interpreterjevim znanjem predstavlja informacijska enačba: $I = i(D, S, t)$.

Simboli v informacijski enačbi pomenijo naslednje: I – količina informacije, i – informacijska funkcija, D – podatki, S – prejemnikovo znanje, t – čas, ki je na voljo prejemniku za interpretacijo podatkov.

1 V komentarju k slovarju je zapisano, da povzema definicije iz standarda ISO/IEC2382, *Information technology – Vocabulary*.

Zaključke, ki jih je povzel Langefors iz omenjene definicije, pa lahko strnemo v naslednje trditve: podatki niso informacija, podatki ne vsebujejo informacije, podatki posredujejo informacijo le tistemu prejemniku, katerega znanje je skladno z izbrano predstavitvijo podatkov in modelom sveta, na katerega se nanašajo. Definicija se od predhodnih razlikuje predvsem po tem, da v njej nastopa dodaten parameter *čas* ali bolje rečeno *pravočasnost*.

Kot je bilo že omenjeno, je informacija koristna le, če se lahko na njeni osnovi tudi ukrepa – in to pravočasno ukrepa. Zgodi se lahko, da je posredovana količina podatkov tako velika, da jih v času, ki je na voljo za ukrepanje, ni mogoče interpretirati. Prej ali slej sicer pridemo do konca in iz podatkov izluščimo informacijo, vendar za ukrepanje je lahko že prepozno – ali kot pravi pesnik 'ura zamujena ne vrne se nobena'. Do podobnega izida lahko pridemo tudi v primeru, ko je prava informacija sicer na voljo, ni pa človeka, ki bi pravočasno ukrepal na njeni osnovi.

Langeforsovo definicijo je zanimivo primerjati z definicijo informacije, ki se uporablja v informacijski teoriji. Informacija je v tem primeru definirana kot:

- znanje, ki zmanjša negotovost, povezano s pojavom določenega dogodka iz končne množice možnih dogodkov.

Informacija, ki jo pridobimo s tem, ko zvemo, da se je pripetil določen dogodek x , se izračuna po formuli:

$$I = -\log_2 p(x) \text{ [bit]}$$

Običajno se kot enota informacije uporablja *bit*, če je logaritem dvojiški, *nit*, če je logaritem naravni in *dit*, če je logaritem desetiški.

Če poiščemo korespondenco med omenjenima definicijama, potem so podatki v slednjem primeru sporočilo, da se je pripetil dogodek x_i , znanje pa je poznavanje verjetnosti nastopa posameznih dogodkov $p(x_i)$, $i = 1..n$. Količina informacije, ki jo izračunamo po omenjeni formuli, se ujema s trditvijo, da je informacija le neko novo spoznanje.

Če se namreč pripeti zelo verjeten dogodek ($p(x_i) = 1$), potem je to enakovredno sprejetim podatkom, ki nam sporočajo nekaj, kar smo že vedeli, zato je $I = 0$ in nimamo torej ničesar dodati k svojemu prejšnjemu znanju. In narobe, če se pripeti malo verjeten dogodek ($p(x_i) = 0$), na primer, da zadenemo na tomboli terno, pridobimo v tem primeru zelo veliko količino informacije (S3).

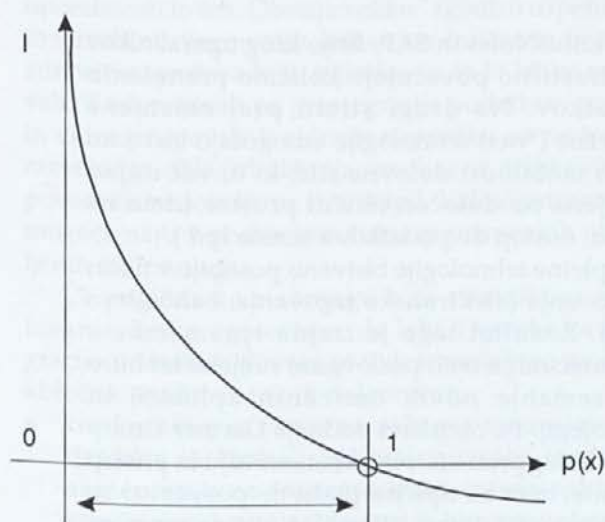
3. ZAKLJUČEK

Oba, človek in računalnik sta si podobna v tem, da sprejemata podatke iz okolja, pri tem, kaj z njimi storita, pa sta si različna. Človek podatkom s pomočjo svojega že obstoječega znanja pripiše pomen, s tem svoje znanje dopolni in na njegovi osnovi ustrezno ukrepa. Nasprotno pa računalnik podatkom ne pripisuje pomena, pač pa jih s pomočjo v njem že shranjenih podatkov in programov le preoblikuje iz ene oblike v drugo.

Na kratko lahko rečemo, da so podatki vse, kar lahko zapišemo na tak ali drugačen pomnilniški medij, informacija in znanje pa sodita v sfero človekovega notranjega miselnega sveta.

4. Viri

- [1] Dictionary of Computer Science – The standardized vocabulary, ISO/IEC2382, 1997, ISBN 2-12-486922-1
- [2] EVEREST, Gordon C.: Database Management, 1986, (str. 7 - 11), McGraw-Hill, ISBN 0-07-Y66456-0
- [3] FALKENBERG, Eckhard D. in ostali: A Framework of Information System Concepts, (str. 16 - 72), IFIP 1998, ISBN 3-901882-01-4
- [4] LANGEFORS, Börje: Infological Models and Information Users Views, (str. 17 - 31), Information Systems 5(1), 1980
- [5] SOWA, John F.: Conceptual Structures, 1984, (str. 1 - 26), Addison-Wesley, ISBN 0-201-14472-7



S3. Informacija kot funkcija verjetnosti nastopa posameznega dogodka

Doc. dr. Tomaž Mohorič je zaposlen na Fakulteti za računalništvo in informatiko, kjer predava vrsto predmetov s področja podatkovnih baz in informacijskih sistemov. Raziskovalno se ukvarja s konceptualnim modeliranjem, z načrtovanjem relacijskih in objektno usmerjenih podatkovnih baz ter s sistemi za upravljanje podatkovnih baz.