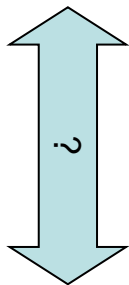


Automatsko prevođenje



Matematički postupci

Matematički postupci u automatskom prevođenju

Goran Igaly

Sveučilište u Zagrebu

PMF – Matematički odsjek

prosinac, 2010.

Sadržaj

- Što je prijevod?
- Relacija prevođenja
- Slovo, riječ, jezik
- Nekomutativne operacije s riječima
- Višeznačne funkcije
- Faktorizacije
- Funkcije udaljenosti i sličnosti
- Gramatika i simboličko računanje
- Višedimenzionalnost i projekcija

Matematički postupci

Automatsko prevodenje

Automatsko prevođenje

- automatski
- prevođenje
 - tekst (riječ)
 - jezik
 - konačni jezik
 - relacija prevođenja
 - prijevod

Automatski

- koji je kao automat, sam bez voljnog poticaja [**automatska** obrana; **automatski** pokret]
- koji radi, djeluje pomoću podešenog mehanizma, tako da skraćuje neke radnje [**automatsko** oružje; **automatska** obrada podataka], opr. ručno

Automat

- *tehn.* uređaj koji samostalno obavlja koristan rad, skraćuje broj radnji ili radni proces bez izravnog sudjelovanja čovjeka u svakoj radnji, ali prema njegovoj zamisli [**automat** za kavu (čaj, čokoladu, osvježavajuće piće)]
- *pren. pejor.* onaj koji radi kao **automat**, bez kreativnosti i inicijative u onome što obavlja

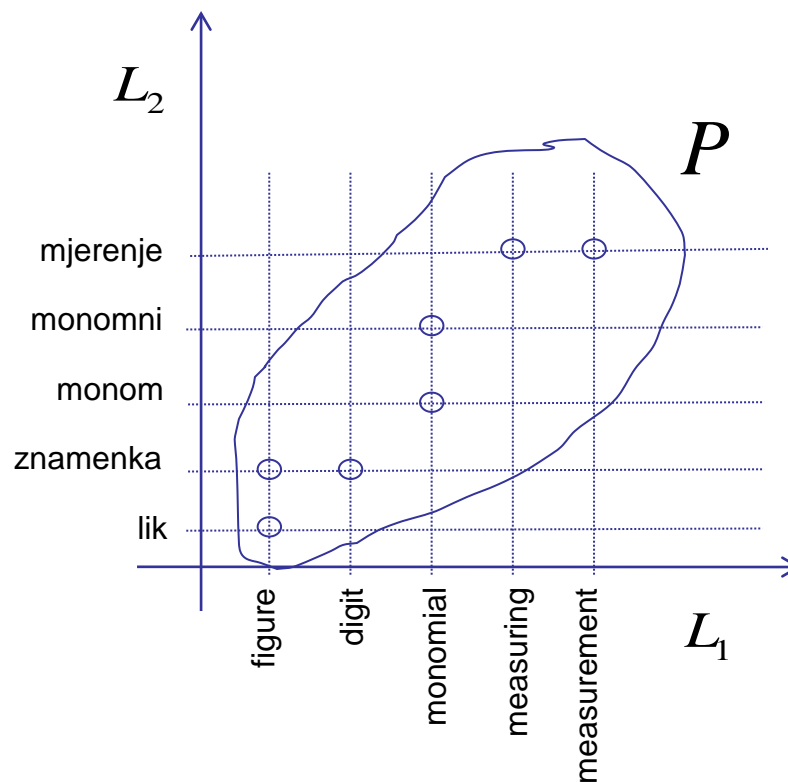
Prevođenje (prevoditi)

- tekst s **jednog jezika** (L_1) izgovoriti ili napisati u **drugom jeziku** (L_2)

- prijevod $(e, h) \in L_1 \times L_2$

- relacija **prevođenja**

$$P \subseteq L_1 \times L_2$$



Jezik, riječ, slovo

- **Jezik** je skup **riječi** tj. skup konačnih nizova znakova (slova) nad nekom abecedom. Skup znakova nad kojim je jezik definiran naziva se **abeceda**.
- **Abeceda** je bilo koji (konačni) skup. Elementi abecede nazivaju se znakovi (**slova**).
- **Riječ** nad abecedom je svaki konačni niz znakova. Skup svih riječi nad abecedom A se označava oznakom A^* . Za svaku abecedu postoji **prazna riječ** - jedna riječ duljine 0, koju označavamo s ϵ .
- Operacijom **konkatenacije** spajamo dvije riječi u jednu tako da na kraj prve riječi nadopišemo drugu. Prazna riječ je neutralni element za operaciju konkatenacije. Operacija konkatenacije je asocijativna ali nije komutativna.

$$A = \{a, b, c\}$$

$$A^* = \{\varepsilon, a, b, c, aa, ab, ac, ba, bb, bc, \dots\}$$

$$L_1 = \{b, ab, bb, cb\} \quad L_2 = \{\varepsilon, b, bb, bbb, bbbb, \dots\}$$

$$aba \cdot cca = abacca$$

$$cca \cdot aba = ccaaba$$

$$aba \cdot \varepsilon = \varepsilon \cdot aba = aba$$

$$ab \cdot (cab \cdot bccba) = ab \cdot cabbccba = abcabbccba$$

$$(ab \cdot cab) \cdot bccba = abcab \cdot bccba = abcabbccba$$

$$aba \cdot cca \neq cca \cdot aba$$

Konačni jezici

- Konačni jezici sadrže konačan skup riječi.

- **Svaki prirodni jezik je konačan.**

$$L = \{e_1, e_2, \dots, e_m\}$$

- Relacija prevođenja za prirodne jezike je konačna (ali velika!).

$$P = \{(e_1, h_1), (e_2, h_2), \dots, (e_n, h_n)\} \subseteq L_1 \times L_2$$

Značenje

- Svakoj riječi (rečenici) nekog jezika pridružen je skup značenja. Ovaj skup može imati jedan ili više elemenata
- Riječ može imati više značenja (luk, kolut, trag)
- Više riječi može imati isto značenje (krevet - postelja, ručak - objed, ljekarna - apoteka)

Značenje - luk

- *bot.* rod višegodišnjih zeljastih biljaka s lukovicom (*Allium*) iz porodice ljljana (*Liliaceae*), karakterističnog oštrog mirisa, s nekoliko važnih kultiviranih vrsta (npr. poriluk *A. porum*)
- **1.** savijen i elastičan dio napregnut tetivom
 - a. jednostavno oružje [*strijelac lukom*]
 - b. sportski rekvizit [*strijeljanje iz luka*]
 - c. igračka za odapinjanje strijela
- **2.** *mat.* dio kružnice ili druge krivulje omeđen dvjema točkama
- **3.** *arhit.* gornji, nadsvođeni završetak otvora u zidu [*polukružni luk; šiljasti luk*]

Značenje - more (hrvatski)

- **1.** velika površina, prostor slane vode koji okružuje kontinente, dio oceana koji više ili manje zalazi u kopno [*Sredozemno more*]
- **2.** *pren.* velika količina, mnoštvo, velik broj čega, veliko prostranstvo [*more problema*]
- **3.** morska voda [*napiti se mora; ugrijati more*]

Značenje - more (engleski)

- (n) **More**, [Thomas More](#), [Sir Thomas More](#) (English statesman who opposed Henry VIII's divorce from Catherine of Aragon and was imprisoned and beheaded; recalled for his concept of Utopia, the ideal state)
- (adj) [many](#) (a quantifier that can be used with count nouns and is often preceded by `as' or `too' or `so' or `that'; amounting to a large but indefinite number) "*many temptations*"; "*the temptations are many*"; "*a good many*"; "*a great many*"; "*many directions*"; "*take as many apples as you like*"; "*too many clouds to see*"; "*never saw so many people*"
- (adj) **more**, [more than](#) ((comparative of `much' used with mass nouns) a quantifier meaning greater in size or amount or extent or degree) "*more land*"; "*more support*"; "*more rain fell*"; "*more than a gallon*"
- (adj) **more** ((comparative of `many' used with count nouns) quantifier meaning greater in number) "*a hall with more seats*"; "*we have no more bananas*"; "*more than one*"
- (adj) [much](#) ((quantifier used with mass nouns) great in quantity or degree or extent) "*not much rain*"; "*much affection*"; "*much grain is in storage*"
- (adv) **more**, [to a greater extent](#) (used to form the comparative of some adjectives and adverbs) "*more interesting*"; "*more beautiful*"; "*more quickly*"

Značenje - more (engleski)

- (adv) **more** (comparative of much; to a greater degree or extent) *"he works more now"; "they eat more than they should"*
- (adv) much (to a great degree or extent) *"she's much better now"*
- (adv) much (very) *"he was much annoyed"*
- (adv) a lot, lots, a good deal, a great deal, much, very much (to a very great degree or extent) *"I feel a lot better"; "we enjoyed ourselves very much"; "she was very much interested"; "this would help a great deal"*
- (adv) much, practically ((degree adverb used before a noun phrase) for all practical purposes but not completely) *"much the same thing happened every time"; "practically everything in Hinduism is the manifestation of a god"*
- (adv) much, a great deal, often (frequently or in great quantities) *"I don't drink much"; "I don't travel much"*

Prijevod kao višeznačno preslikavanje

- Cilj prijevoda je **sačuvati značenje**
- Svakoj riječi (rečenici) izvornog jezika pridružen je **skup riječi (rečenica)** ciljnog jezika

$$e \in L_1$$

$$P(e) = H \subseteq L_2$$

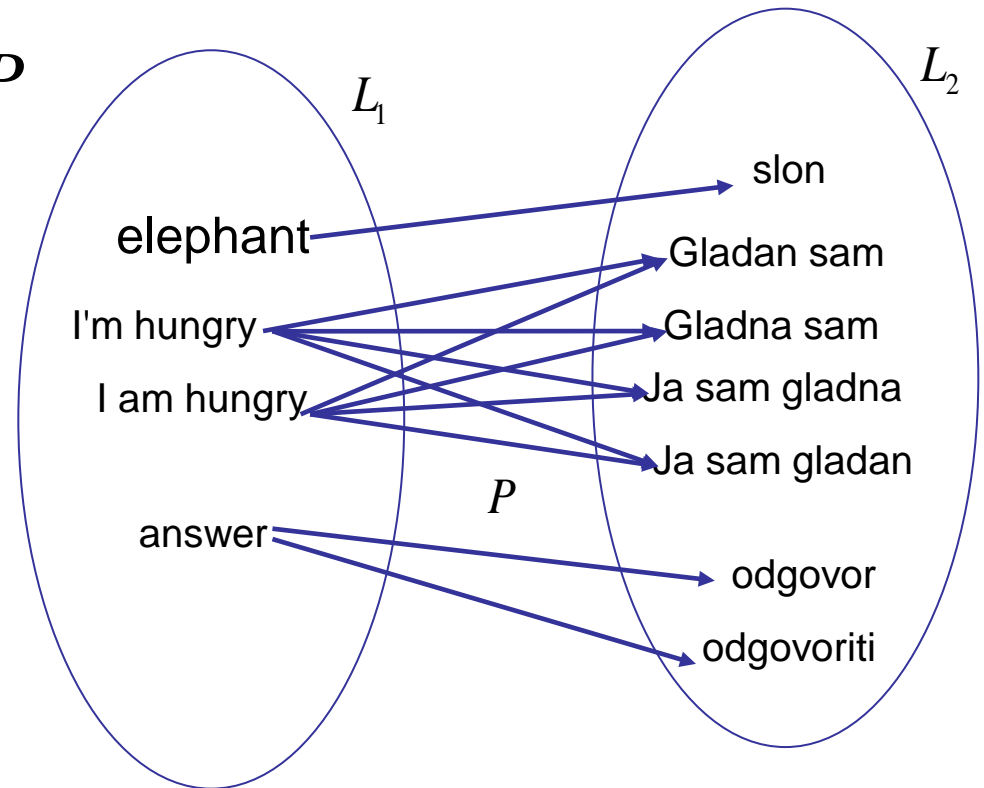
- e i $P(e)$ bi trebali imati isto značenje

Jednoznačne i višeznačne riječi

$P(\text{"answer"}) \supseteq \{\text{"odgovoriti"}, \text{"odgovor"}\}$

$(\text{"answer"}, \text{"odgovoriti"}) \in P$

$(\text{"answer"}, \text{"odgovor"}) \in P$



$P(\text{"elephant"}) = \{\text{"slon"}\}$

$(\text{"elephant"}, \text{"slon"}) \in P$

Primjer višeznačnog preslikavanja u matematici

- Jednadžbi pridružimo skup svih njezinih rješenja
 - beskonačan skup rješenja: $\sin(2x + 3) = \frac{4}{5}$
 - konačan skup rješenja: $x^9 = 13$
 - jedinstveno rješenje: $2x + 7 = 5$
 - prazan skup rješenja: $x + 7 = x + 5$
- Svako preslikavanje koje elementima domene pridružuje skupove možemo shvatiti kao višeznačno preslikavanje

Automatsko prevođenje

- Definicija za automatsko prevođenje obično vodi na definiciju za **strojno prevođenje**
- Prevest ćemo tu definiciju pomoću dva sustava za strojno prevođenje.
- Koji je prijevod bolji?

<p>Machine translation, sometimes referred to by the abbreviation MT, is a sub-field of computational linguistics that investigates the use of computer software to translate text or speech from one natural language to another.</p>	<p>Strojno prevođenje, ponekad se po skraćenica MT, je pod-polje računalna lingvistika koja istražuje upotrebu računala software za prevođenje teksta ili govora s jednog prirodnog jezika na drugi.</p>	<p>Strojno prevođenje, katkada koji se odnosi na koga to mimo skraćenica MT, je pod - polje od izračunavanje lingvistika taj istraživati iskoristiti od računalo softver prevoditi tekst ili govor from neki prirodni jezik to drugi.</p>
<p>At its basic level, MT performs simple substitution of words in one natural language for words in another.</p>	<p>Na svojoj osnovnoj razini, MT obavlja jednostavne zamjena riječi u jednoj riječi prirodnog jezika umjesto na drugom.</p>	<p>At svoj osnovni nivo MT provodi jednostavan zamjena riječi in neki prirodni jezik za riječ in drugi.</p>
<p>Using corpus techniques, more complex translations may be attempted, allowing for better handling of differences in linguistic typology, phrase recognition, and translation of idioms, as well as the isolation of anomalies.</p>	<p>Koristeći korpusu tehnika, složeniji prijevodi mogu se pokušati, koji dozvoljava za bolje rukovanje razlika u jezičnoj tipologiji, izraz priznanje, i prijevod idioma, kao i izolaciju anomalija.</p>	<p>Koristeći čestica tehnika, više kompliciran prevođenje može biti pokus, koji dozvoljava za bolji rukovanje od razlika in jezični typology, fraza priznanje, i prevođenje od idiomi, isto kao i odvajanje od anomalije</p>
<p>Current machine translation software often allows for customisation by domain or profession (such as weather reports) — improving output by limiting the scope of allowable substitutions.</p>	<p>Trenutni strojni prijevod softver često omogućuje prilagodbu prema domeni ili struci (kao što su vremenska izvješća) — unapređenje proizvodnje za ograničavanjem opsega dozvoljenih zamjena.</p>	<p>Tekući software za strojno prevođenje više puta dopustiti za customisation mimo domena ili stalno zanimanje (kao što je hidrometeorološki izvještaj — napredan iskopana ruda mimo ograničavajući cilj od koji se smije zamjena.</p>
<p>This technique is particularly effective in domains where formal or formulaic language is used.</p>	<p>Ova tehnika je posebno učinkovit u domenama kojima formalni šablonski ili jezik upotrebljava.</p>	<p>Ovaj tehnika je osobito djelotvoran in domena gdje svi formalan ili formula jezik je korišten.</p>

	lipanj 2010	prosinac 2010
Machine translation, sometimes referred to by the abbreviation MT, is a sub-field of computational linguistics that investigates the use of computer software to translate text or speech from one natural language to another.	Strojno prevođenje , ponekad se po skraćenica MT, je pod-polje računalna lingvistika koja istražuje upotrebu računala software za prevođenje teksta ili govora s jednog prirodnog jezika na drugi.	Strojni prijevod , ponekad na koje skraćenica MT, je pod-polje računalna lingvistika koja istražuje upotrebu računala software za prevođenje teksta ili govora s jednog prirodnog jezika na drugi.
At its basic level, MT performs simple substitution of words in one natural language for words in another.	Na svojoj osnovnoj razini, MT obavlja jednostavne zamjena riječi u jednoj riječi prirodnog jezika umjesto na drugom .	Na svojoj osnovnoj razini, MT obavlja jednostavna zamjena riječi u jednoj riječi prirodnog jezika u drugi .
Using corpus techniques, more complex translations may be attempted, allowing for better handling of differences in linguistic typology, phrase recognition, and translation of idioms, as well as the isolation of anomalies.	Koristeći korpusu tehnika , složeniji prijevodi mogu se pokušati , koji dozvoljava za bolje rukovanje razlika u jezičnoj tipologiji , izraz priznanje, i prijevod idioma, kao i izolaciju anomalija.	Koristeći tehnikе korpus , složeniji prijevodi mogu biti pokušano , dopuštajući za bolje rukovanje razlike u jezičnom tipologijom , izraz priznanje, i prijevod idioma, kao i izolaciju od anomalija.
Current machine translation software often allows for customisation by domain or profession (such as weather reports) — improving output by limiting the scope of allowable substitutions.	Trenutni strojni prijevod softver često omogućuje prilagodbu prema domeni ili struci (kao što su vremenska izvješća) — unapređenje proizvodnje za ograničavanjem opsega dozvoljenih zamjena .	Trenutni strojni prijevod softvera često dopušta za prilagodbu prema domeni ili struke (kao što su vremenske prognoze) — poboljšanje izlaz ograničavanjem opsega dopuštenog zamjene .
This technique is particularly effective in domains where formal or formulaic language is used.	Ova tehnika je posebno učinkovit u domenama kojima formalni šablonski ili jezik upotrebljava .	Ova tehnika je posebno učinkovit u područjima gdje formalni šablonski ili jezik koristi .

Automatsko prevodenje "riječ po riječ"

- At its basic level, MT performs simple substitution of words in one natural language for words in another.
- Na svojoj osnovnoj razini, MT obavlja **jednostavne** zamjena riječi u jednoj riječi prirodnog jezika **umjesto na drugom**.
([Google, lipanj 2010](#))
- Na svojoj osnovnoj razini, MT obavlja **jednostavna** zamjena riječi u jednoj riječi prirodnog jezika **u drugi**.
([Google, prosinac 2010](#))
- At svoj osnovni nivo MT provodi jednostavan zamjena riječi in neki prirodni jezik za riječ in drugi.

Na	Na	At	At
svojoj	svojoj	its	svoj
osnovnoj	osnovnoj	basic	osnovni
razini	razini	level	nivo
,	,	,	
MT	MT	MT	MT
obavlja	obavlja	performs	provodi
jednostavne	jednostavna	simple	jednostavan
zamjena	zamjena	substitution	zamjena
riječi	riječi	of words	riječi
u	u	in	in
jednoj riječi	jednoj riječi	one	neki
prirodnog	prirodnog	natural	prirodni
jezika	jezika	language	jezik
umjesto		for	za
		words	riječ
na	u	in	in
drugom	drugi	another	drugi
.	.	.	.

(Google, lipanj 2010) (Google, prosinac 2010)

Mogućnosti poboljšanja

- prijevod većih dijelova teksta (blokova)
- zamjena redoslijeda prevedenih blokova
- pronalaženje isprepletenih blokova u izvornom tekstu
- ispreplitanje blokova u prijevodu

Matematika:

- uvrštavanje u algebarske izraze
- uočavanje poznatih izraza (adicijske formule, kvadrat binoma, trigonometrijski identiteti)

Faktorizacija

- Veliki tekst se dijeli na manje cjeline i te cjeline se odvojeno prevode

Matematika:

- Faktorizacija brojnika, nazivnika

Prva razina faktorizacije: **podjela teksta na rečenice**

- 1) Machine translation, sometimes referred to by the abbreviation MT, is a sub-field of computational linguistics that investigates the use of computer software to translate text or speech from one natural language to another.
- 2) At its basic level, MT performs simple substitution of words in one natural language for words in another.
- 3) Using corpus techniques, more complex translations may be attempted, allowing for better handling of differences in linguistic typology, phrase recognition, and translation of idioms, as well as the isolation of anomalies.
- 4) Current machine translation software often allows for customisation by domain or profession (such as weather reports) — improving output by limiting the scope of allowable substitutions.

Nizanje prijevoda

- Svaka rečenica se prevodi posebno i konačni prijevod se dobiva nizanjem prijevoda rečenica.
- Ovaj koncept se često prenosi na riječi u rečenici i to je česti uzrok pogrešaka pri automatskom prevođenju
- *At its basic level, MT performs simple substitution of words in one natural language for words in another.*
- Na svojoj osnovnoj razini, MT obavlja jednostavne zamjena riječi u jednoj riječi prirodnog jezika umjesto na drugom. (G, 06/2010)
- Na svojoj osnovnoj razini, MT obavlja jednostavna zamjena riječi u jednoj riječi prirodnog jezika u drugi. (G, 12/2010)
- At svoj osnovni nivo MT provodi jednostavan zamjena riječi in neki prirodni jezik za riječ in drugi.

Jednostavna zamjena i nizanje

$$f(a + b) = f(a) + f(b) \quad P(abc) \mapsto P(a)P(b)P(c)$$

$$f(a \cdot b) = f(a) \cdot f(b)$$

- Općenito:

$$f(a \circ b) = f(a) \bullet f(b)$$

Matematika:

$$e^{x+y} = e^x \cdot e^y$$

$$\sin(x + y) \neq \sin(x) + \sin(y)$$

$$\ln(x \cdot y) = \ln(x) + \ln(y)$$

$$(f + g)' = f' + g'$$

$$a \cdot (x + y) = a \cdot x + a \cdot y$$

$$(f \cdot g)' = f' \cdot g + f \cdot g'$$

Prijevod bloka

- Current **machine translation software** often allows for customisation by domain or profession
- Trenutni **strojni prijevod softver** često omogućuje prilagodbu prema domeni ili struci (G, lipanj 2010)
- Trenutni **strojni prijevod softvera** često dopušta za prilagodbu prema domeni ili struke (G, prosinac 2010)
- Tekući **software za strojno prevođenje** više puta dopustiti za customisation mimo domena ili stalno zanimanje

Podjela rečenice na blokove

- Rečenica se dijeli na manje povezane cjeline, ali ne odmah na riječi.
- Primjer:
 - Rečenica
 - se dijeli
 - na manje povezane cjeline,
 - ali ne odmah
 - na riječi.

Automatsko prevođenje

- Prijevodi cjelina (rečenice, nizovi riječi, riječi) se traže u bazi
- Ako u bazi ne postoji prijevod kompletne cjeline, konstruira se prijevod na temelju postojećeg prijevoda slične cjeline ili se cjelina rastavlja na manje dijelove, svaki se posebno prijevodi i sastavlja se prijevod na temelju određenih pravila

Matematika:

Sličnost s postupkom integriranja - svodenje na tablični integral: Ako izraz kojega treba integrirati imamo u tablicama, očitamo integral iz tablica, ako ga nemamo, svodimo ga na tablični supstitucijom, rastavljanjem na više integrala, rastavom na parcijalne razlomke, ...

Faktorizacija (segmentacija) korištenjem zareza

- Using corpus techniques, more complex translations may be attempted, allowing for better handling of differences in linguistic typology, phrase recognition, and translation of idioms, as well as the isolation of anomalies.
 - Using corpus techniques,
 - more complex translations may be attempted,
 - allowing for better handling of differences in linguistic typology,
 - phrase recognition,
 - and translation of idioms,
 - as well as the isolation of anomalies.

Matematika:

Definiranje redoslijeda operacija pomoću zagrada

"Prosti faktori"

- Using **corpus techniques**, **more complex translations** may be attempted, allowing for better **handling of differences** in **linguistic typology**, **phrase recognition**, and **translation of idioms**, as well as the **isolation of anomalies**.
- Koristeći **korpusu tehnika**, **složeniji prijevodi** mogu se pokušati, koji dozvoljava za bolje **rukovanje razlika** u **jezičnoj tipologiji**, **izraz priznanje**, **G lipanj 2010** i **prijevod idioma**, kao i **izolaciju anomalija**.
- Koristeći **čestica tehnika**, **više kompliciran prevođenje** može biti pokus, koji dozvoljava za bolji **rukovanje od razlika** in **jezični typology**, **fraza priznanje**, i **prevođenje od idiomi**, isto kao i **odvajanje od anomalije**

- Koristeći **korpusu tehnika**, **složeniji prijevodi** mogu biti pokušano, omogućava bolje **rukovanje razlika** u **jezičnom tipologijom**, **izraz priznanje**, **G prosinac2010** i **prijevod idioma**, kao i **izolaciju od anomalija**.

<p>The documents referred to in this Article sent to the Commission shall be in one of the official languages of the institutions of the European Union or be accompanied by a certified translation in one of those languages.</p>	<p>Dokumenti iz ovoga članka poslan Komisiji G lipanj 2010 će se na jednom od službenih jezika institucija Europske unije ili pratiti ovjereni prijevod od strane u jednom od tih jezika.</p>	<p>Dokumenti navedeni u ovom članku koji se šalju Komisiji moraju biti napisani na jednom od službenih jezika institucija Europske unije ili moraju biti praćeni ovjerenim prevodom na jedan od tih jezika.</p>
<p>The documents referred to in this Article that are sent to the Commission shall be written in one of the official languages of the institutions of the European Union or shall be accompanied by a certified translation in one of those languages.</p>	<p>Dokumente navedenih u ovom članku koje se šalju Komisiji mora biti napisan na jednom od službenih jezika institucija Europske unije ili mora biti popraćen by ovjereni prijevod u jednom od tih jezika.</p>	<p>Dokumenti navedeni u ovom članku koji se šalju Komisiji moraju biti napisani na jednom od službenih jezika institucija Europske unije ili moraju biti praćeni ovjerenim prevodom na jedan od tih jezika.</p>

<p>The documents referred to in this Article sent to the Commission shall be in one of the official languages of the institutions of the European Union or be accompanied by a certified translation in one of those languages.</p>	<p>Dokumente iz ovog članka poslana Komisiji. G prosinac 2010 će biti na jednom od službenih jezika institucija Europske unije ili biti popraćen po ovjereni prijevod u jednom od tih jezika.</p>	<p>Dokumenti navedeni u ovom članku koji se šalju Komisiji moraju biti napisani na jednom od službenih jezika institucija Europske unije ili moraju biti praćeni ovjerenim prijevodom na jedan od tih jezika.</p>
<p>The documents referred to in this Article that are sent to the Commission shall be written in one of the official languages of the institutions of the European Union or shall be accompanied by a certified translation in one of those languages.</p>	<p>Dokumente iz ovog članka koji se šalju Komisiji mora biti napisana na jednom od službenih jezika institucija Europske unije ili moraju biti popraćeni po ovjereni prijevod u jednom od tih jezika.</p>	<p>Dokumenti navedeni u ovom članku koji se šalju Komisiji moraju biti napisani na jednom od službenih jezika institucija Europske unije ili moraju biti praćeni ovjerenim prijevodom na jedan od tih jezika.</p>

Faktorizacija u matematici

- Pogodno grupiranje članova

$$\begin{aligned}x^3 + 17x^2 + x + 17 &= \\&= x^2(x + 17) + (x + 17) = \\&= (x^2 + 1) \cdot (x + 17)\end{aligned}$$

Problem pravilne segmentacije

- Pogledaćemo rješenje zadatka kad dođemo doma
- Pogledat ćemo rješenje zadatka kad dođemo doma.

- pet plus iks kroz dva minus iks

$$5 + \frac{x}{2} - x$$

- pet plus iks kroz dva minus iks

$$\frac{5 + x}{2 - x}$$

$$5 + \frac{x}{2 - x}$$

$$\frac{5 + x}{2} - x$$

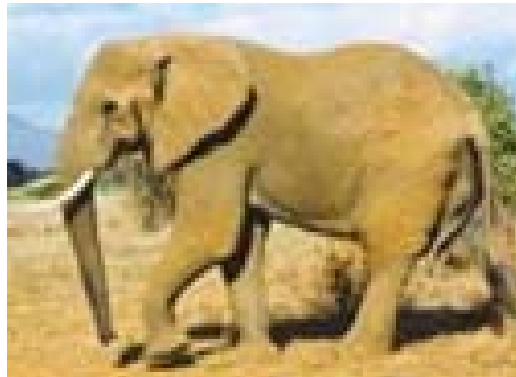
Sličnost

- Jesu li slični?
- Koliko su slični?
- Gdje su razlike?

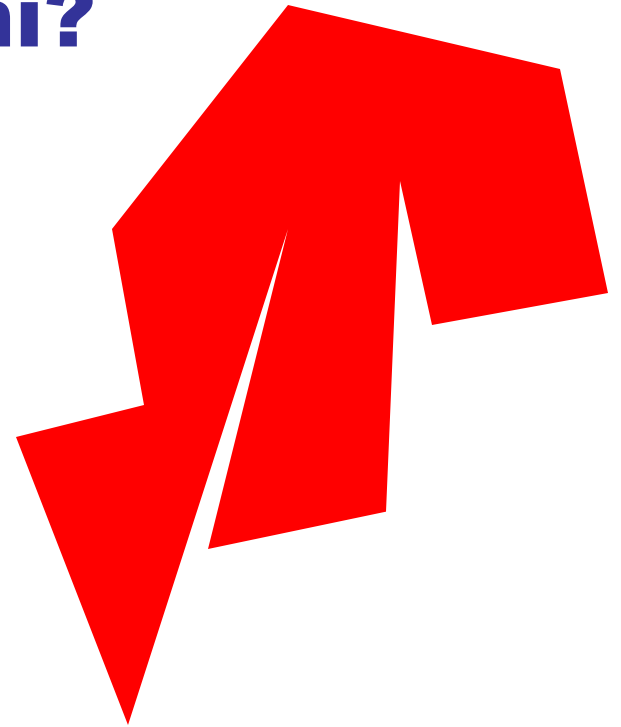
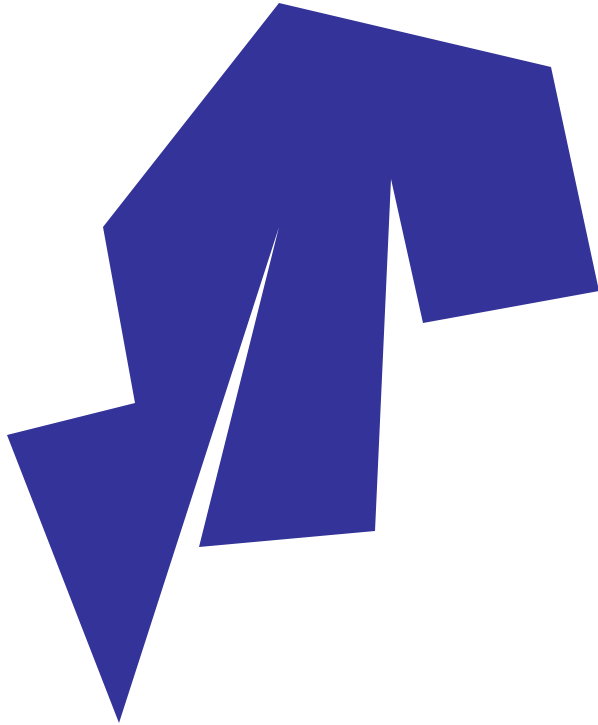


Poravnavanje nizova

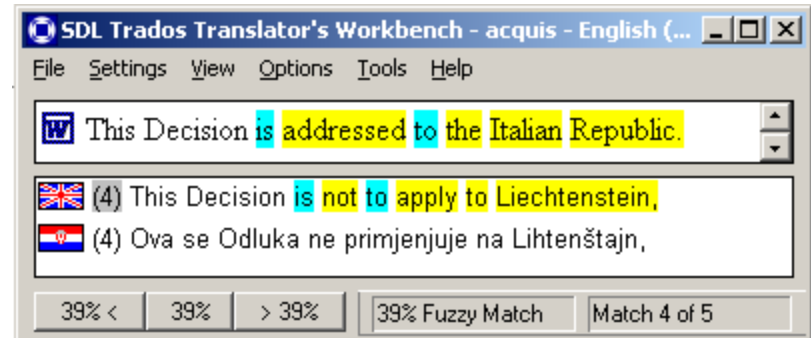
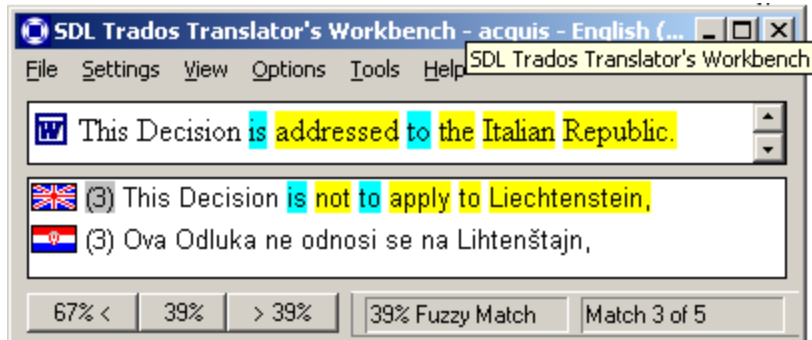
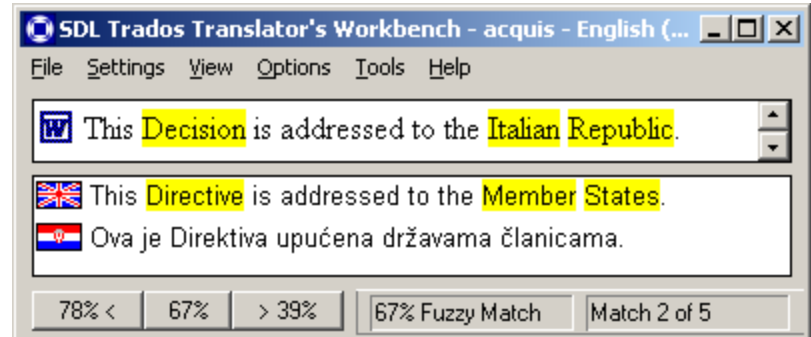
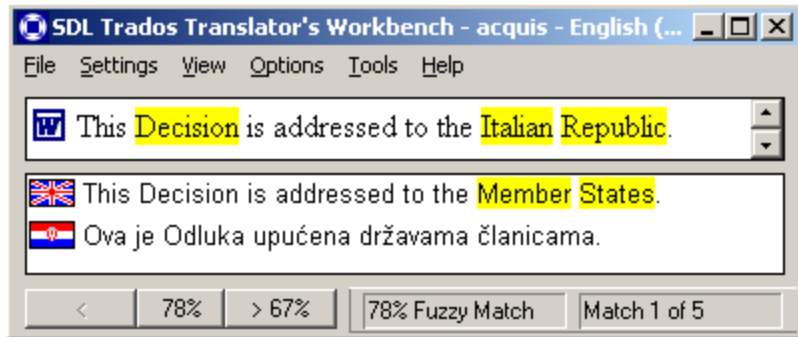
	250	260	270	280	290	300	310	320	330	340	
Hippopotamus/1-1162	GGGTGAGTCATCCGCTACATACACGCAAACGGCGCATCCATCTTCTTCATCTGCCTCTTTACTCAGGTAGGACGCGGCCTATACTATGGCTCCTT										
Rhino/1-1162	GGCTGAATTATCCGCTACCTACATGCAAACGGAGCATCCATATTTTTATCTGCCTATTCAATCCACATAGGACGCGGCCTCTATTACGGATCCTT										
Mammoth/1-1162	GGCTGAATTATTCGACAACCTACACTCAAACGGAGCATCCATCTTCTTCCTCTGCCTATAACACACACATTGGACGAAACATCTACTATGGGTCCCTT										
African Elephant/1-1162	GGCTGAATTATTCGACAACCTACACTCAAACGGAGCATCCATTTTTCTTCCTCTGCCTATAACACACACATTGGACGAAACATCTACTATGGGTCCCTT										
Asian Elephant/1-1162	GGCTGAATTATTCGACAACCTGCACTCAAACGGAGCATCTATCTTTTTCTTCCTCTGCCTATAACACACACATTGGACGAAACATCTACTATGGATCCTT										
Mastodon/1-1162	GGCTGAATTATTCGACAACCTACACTCAAACGGAGCATCCATCTTCTTCCTCTGCCTATAACACACACATTGGACGAAACATCTACTACGGATCCTT										
Doq/1-1162	GGCTGAATTATCCGCTATATGCAACGCAAATGGCGCTCCATATTTCTTATCTGCCTATTCTTACATGTAGGACGAGGCCTATATTACGGATCCTT										
Human/1-1162	GGCTGAATCATCCGCTACCTTCACGCCAATGGCGCCTCAATATTTCTTATCTGCCTCTTCTTACACATCGGGCGAGGCCTATATTACGGATCAT										



Jesu li slični?



Sličnost rečenica

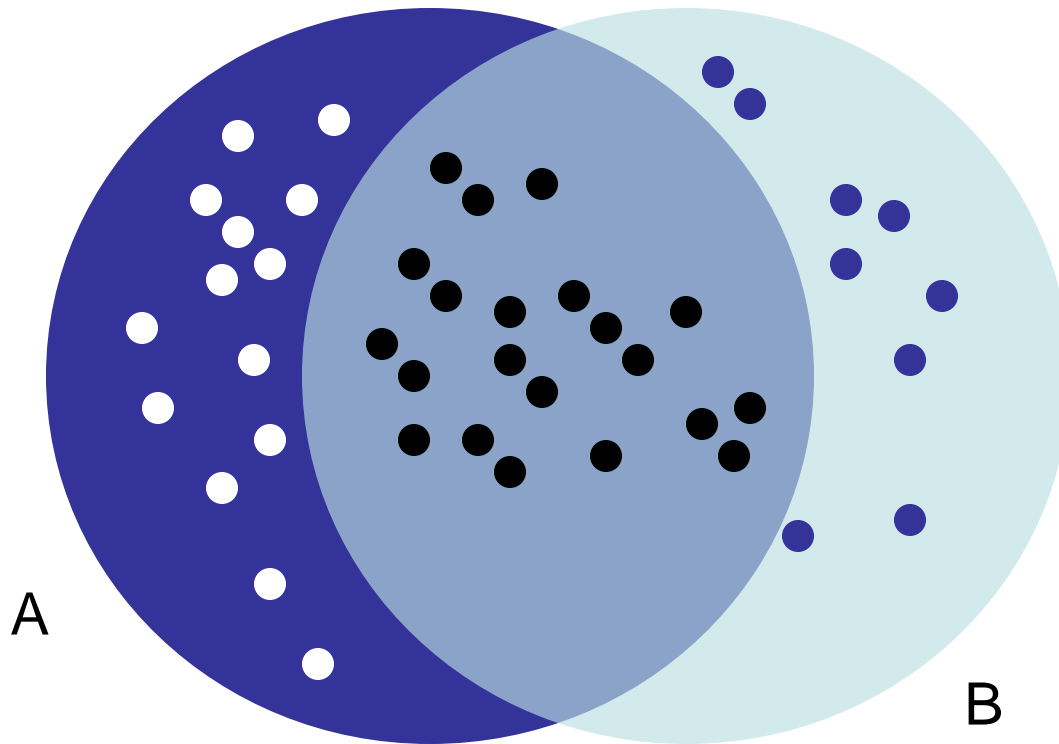


Postotak sličnosti



$$\frac{(x-3)(2x+5)(5x-4)(x+9)}{(x-3)(2x+5)(5x-4)(x+11)(x+9)}$$

Sličnost skupova



$$|A| = 35$$

$$|B| = 30$$

$$|A \cap B| = 21$$

$$|A \cup B| = 44$$

$$|A \setminus B| = 14$$

$$|B \setminus A| = 9$$

$$\frac{|A \cap B|}{|A \cup B|} = \frac{21}{44} = 48\%$$

Neprekidnost

- Očekujemo da sličnim rečenicama odgovaraju slični prijevodi:
- Bear eats **s** strawberries = Medvjed jede **e** jagode
- Bears **s** eat strawberries = Medvjedi **i** jedu **u** jagode
- **Dogs** eat = **Psi** jedu
- **Cats** eat = **Mačke** jedu

Matematika:

Neprekidna funkcija bliskim točkama pridružuje bliske funkcijske vrijednosti

Testiranje na kratkim rečenicama

- Dijete prvo uči jednostavne i kratke rečenice.
- Pogledajmo kako najčešće korišteni sustav danas prevodi kratke rečenice

Primjeri prijevoda kratkih rečenica

- Hladno mi je >> I'm cold
- Dosadno mi je >> I'm bored
- Vruće mi je >> Download me // Hot me (12/2010)
- vruće mi je >> hot me

- Download me >> Preuzmite mene // Preuzimanje mene (12/2010)
- Download me! >> Preuzmite me! // Preuzimanje me! (12/2010)
- Download me. >> Vruće mi je. // Preuzimanje mene. (12/2010)

- I am hot >> Ja sam hot // Ja sam vruće (12/2010)
- I am cold >> Ja sam hladna // Ja sam hladno (12/2010)

Primjeri prijevoda kratkih rečenica (2)

- Men eat = Muškarci jesti
- men eat = ljudi jedu // ljudi jesti (12/2010)
- MEN EAT = MUŠKI EAT // MEN EAT (12/2010)
- MEN EAT. = MEN jesti.

- Women eat = Žene jedu
- women eat = žene da jedu porod // žene da jedu (12/2010)

Primjeri prijevoda kratkih rečenica (3)

- Vruće mu je >> Popular him
- Vruće joj je >> Download her // She is hot (12/2010)
- Vruće mi je >> Download me // Hot me (12/2010)
- Vruće nam je >> Download our // Popular sorry (12/2010)

Još nekoliko primjera

06/2010

12/2010

Gladan sam. Žedan sam.	I'm hungry. I thirst.	I'm hungry. I am thirsty.
Pospan sam. Budan sam.	I'm sleepy. I am awake.	I'm sleepy. I'm awake.
Umoran sam. Odmoran sam.	I'm tired. I rested.	I'm tired. I rested.
Lijen sam. Vrijedan sam. Marljiv sam.	I'm lazy. Worth alone. I was diligent.	I'm lazy. I am worth. I was diligent.
Lijep sam. Ružan sam.	Beauty alone. I'm ugly.	Beauty alone. I'm ugly.

Još nekoliko primjera

06/2010

12/2010

Ja sam marljiv.

I am hard working.

I am industrious.

Ti si marljiv.

You're hardworking.

You are industrious.

On je marljiv.

He is hard working.

He is industrious.

Ona je marljiva.

She is diligent.

She is diligent.

Ono je marljivo.

It was hard.

It has been hard.

Mi smo marljivi.

We are diligent.

We are diligent.

Vi ste marljivi.

You are diligent.

You are industrious.

Oni su marljivi.

They are diligent.

They are diligent.

One su marljive.

They are hard working

They are diligent.

Ona su marljiva.

They are diligent.

They are diligent.

Sličnost rečenica

- Tekstovi koji nastaju metodom copy-paste su najbolji kandidati za automatsko prevođenje.
- Pogledajmo primjer nekoliko rečenica iz matematičkog kurikuluma za srednju školu.
- Odabrani su dijelovi koji sadržajno imaju veze s automatskim prevođenjem

B. Povezivanje

1.


- **1. ciklus:** B3. usporediti, grupirati i **razvrstati** objekte i pojave prema **određenom** kriteriju **u jednostavnim konkretnim situacijama**.
- **2. ciklus:** B3. usporediti, grupirati i **klasificirati** objekte i pojave prema **određenom** kriteriju.
- **3. ciklus:** B3. usporediti, grupirati i **klasificirati** objekte i pojave prema **zadanom ili izabranom** kriteriju.
- **4. ciklus (strukovno obrazovanje):** B3. usporediti, grupirati i **klasificirati** objekte i pojave prema **zadanom ili izabranom** kriteriju.
- **4. ciklus (gimnazije):** B3. usporediti, grupirati i **klasificirati** objekte i pojave prema **zadanom ili izabranom** kriteriju.


2.


3.

SDL Trados Translator's Workbench - brisati - Croatian -> English (United Kingdom)


File Settings View Options Tools Help


 usporediti, grupirati i **klasificirati** objekte i pojave prema određenom kriteriju.


 usporediti, grupirati i **razvrstati** objekte i pojave prema određenom kriteriju u jednostavnim konkretnim situacijama.

 usporediti, grupirati i razvrstati objekte i pojave prema određenom kriteriju u jednostavnim konkretnim situacijama.


< 60% > 60% Fuzzy Match Match 1 of 1


 usporediti, grupirati i klasificirati objekte i pojave prema zadanom ili **izabranom** kriteriju.


 usporediti, grupirati i klasificirati objekte i pojave prema **određenom** kriteriju.

 usporediti, grupirati i klasificirati objekte i pojave prema određenom kriteriju.

< 80% > 56% 80% Fuzzy Match Match 1 of 2

 usporediti, grupirati i **klasificirati** objekte i pojave prema **zadanom ili izabranom kriteriju**.

 usporediti, grupirati i **razvrstati** objekte i pojave prema **određenom kriteriju** u **jednostavnim konkretnim situacijama**.

 usporediti, grupirati i razvrstati objekte i pojave prema određenom kriteriju u jednostavnim konkretnim situacijama.

80% < 56% > 56% Fuzzy Match Match 2 of 2

C. Logičko mišljenje, argumentiranje i zaključivanje

- 1. ciklus: C3. zaključivati nepotpunom indukcijom i neformalnom dedukcijom s malim brojem koraka.
- 2. ciklus: C3. zaključivati nepotpunom indukcijom i neformalnom dedukcijom s malim brojem koraka.
- 3. ciklus: C3. pratiti i stvarati kraće lance matematičkih argumenata, zaključivati nepotpunom indukcijom i neformalnom dedukcijom, te primjenjivati analogiju, generalizaciju i specijalizaciju u jednostavnim situacijama.
- 4. ciklus (strukovno obrazovanje): C3. pratiti i stvarati kraće lance matematičkih argumenata, zaključivati nepotpunom indukcijom i neformalnom dedukcijom, te primjenjivati analogiju, generalizaciju i specijalizaciju u jednostavnim situacijama.
- 4. ciklus (gimnazije): C3. pratiti, stvarati i vrednovati lance matematičkih argumenata različitih vrsta, te primjenjivati analogiju, generalizaciju i specijalizaciju,

1.

2.

3.

4.

zaključivati nepotpunom indukcijom i neformalnom dedukcijom s malim brojem koraka.

zaključivati nepotpunom indukcijom i neformalnom dedukcijom s malim brojem koraka.

zaključivati nepotpunom indukcijom i neformalnom dedukcijom s malim brojem koraka.

< 100% > Exact Match Match 1 of 1

Concordance - brisati (1 hit)

Search << Previous Next >> Clear Options... Close

zaključivati nepotpunom indukcijom i neformalnom dedukcijom Help

zaključivati nepotpunom indukcijom i neformalnom dedukcijom s malim brojem koraka.

zaključivati nepotpunom indukcijom i neformalnom dedukcijom s malim brojem koraka.

Created on: 20.6.2010, 8:02 Created by: GORAN Changed on: Changed by: Last Used: 20.6.2010, 8:04 Usage: 1

SDL Trados Translator's Workbench - brisati - Croatian -> English (United Kingdom)

File Settings View Options Tools Help

pratiti stvarati vrednovati lance matematičkih argumenata različitih vrsta, te primjenjivati analogiju, generalizaciju i specijalizaciju.

pratiti stvarati kraće lance matematičkih argumenata, zaključivati nepotpunom indukcijom i neformalnom dedukcijom, te primjenjivati analogiju, generalizaciju i specijalizaciju u jednostavnim situacijama.

pratiti stvarati kraće lance matematičkih argumenata, zaključivati nepotpunom indukcijom i neformalnom dedukcijom, te primjenjivati analogiju, generalizaciju i specijalizaciju u jednostavnim situacijama.

< 46% > 46% Fuzzy Match Match 1 of 1

D. Rješavanje problema i matematičko modeliranje

- **1. ciklus:** D2. primijeniti matematičke pojmove i postupke u **različitim kontekstima**,
- **2. ciklus:** D2. primijeniti matematičke pojmove i postupke u **različitim kontekstima**,
- **3. ciklus:** D2. primijeniti matematičke pojmove i postupke u **različitim kontekstima**,
- **4. ciklus (strukovno obrazovanje):** D2. primijeniti matematičke pojmove i postupke u **svakodnevnom privatnom, profesionalnom i društvenom životu te u drugim odgojno-obrazovnim područjima**,



1.



2.

G. Algebra i funkcije

- **2. ciklus:** G3. rabiti opće brojeve (slova) u zapisivanju jednostavnih formula i izraza, te umjesto njih **uvrstiti konkretne vrijednosti**,
- **3. ciklus:** G2. **uvrstiti konkretne vrijednosti u formulu** i **izračunati vrijednost preostale veličine**
- **4. ciklus (strukovno obrazovanje):** G2. **uvrstiti konkretne vrijednosti u formulu** i **izračunati vrijednost preostale veličine**,
- **4. ciklus (gimnazije):** G2. **uvrstiti konkretne vrijednosti u formulu** (osobito u funkciju zadanu formulom), **izračunati vrijednost preostale veličine**, te izraziti jednu veličinu u formuli pomoću ostalih,



1.



2.

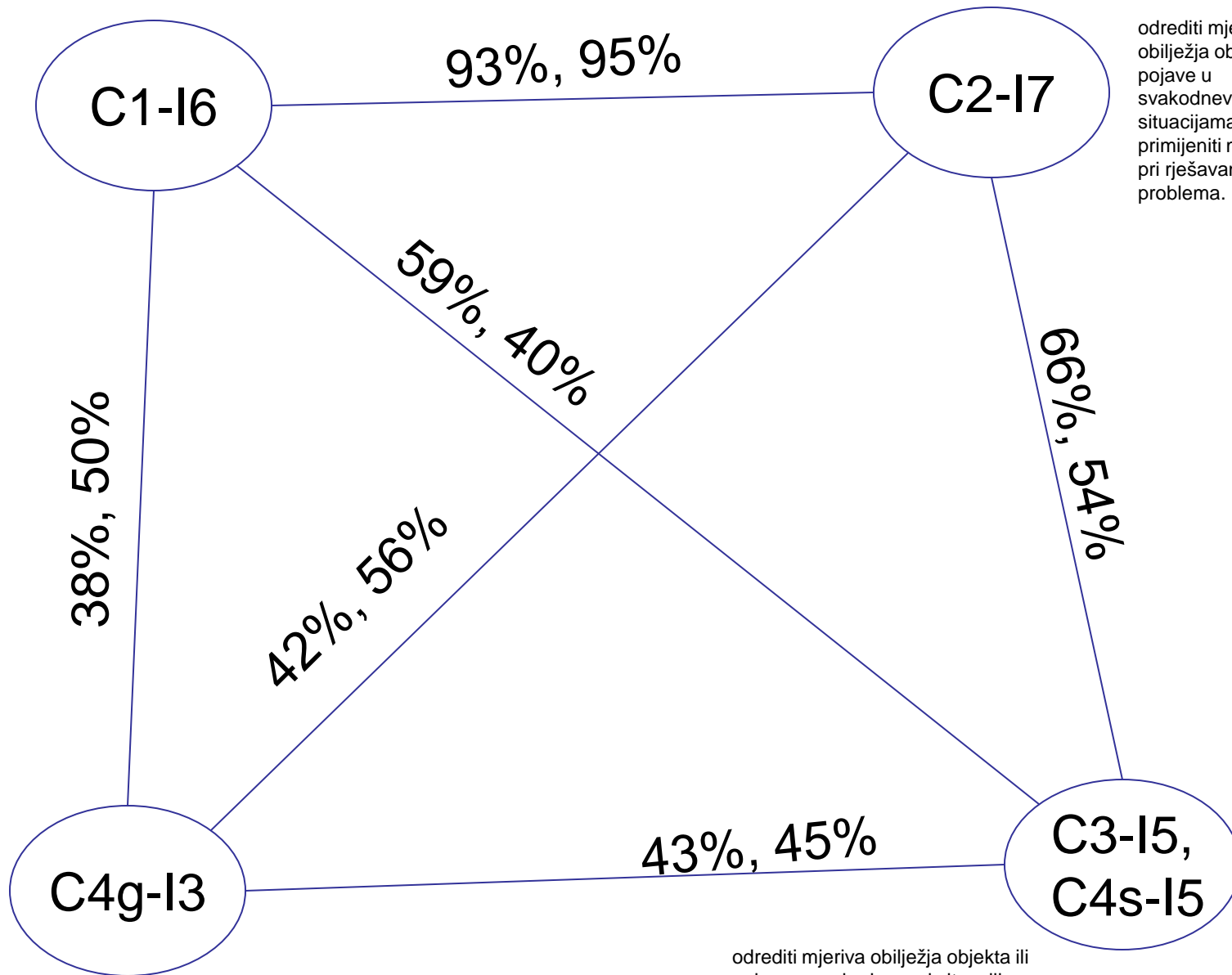


3.

I. Mjerenje

- **1. ciklus: I6. odrediti mjeriva obilježja jednostavnog objekta ili pojave u svakodnevnim situacijama i primijeniti mjerenje pri rješavanju problema.**
- **2. ciklus: I7. odrediti mjeriva obilježja objekta ili pojave u svakodnevnim situacijama i primijeniti mjerenje pri rješavanju problema.**
- **3. ciklus: I5. odrediti mjeriva obilježja objekta ili pojave u svakodnevnoj situaciji**, odabrati primjerene mjerne jedinice i mjerne uređaje, te **primijeniti mjerenje pri rješavanju problema.**
- **4. ciklus (strukovno obrazovanje): I5. odrediti mjeriva obilježja objekta ili pojave u svakodnevnoj situaciji**, odabrati primjerene mjerne jedinice i mjerne uređaje, te **primijeniti mjerenje pri rješavanju problema.**
- **4. ciklus (gimnazije): I3. odrediti mjeriva obilježja objekta ili pojave u svakodnevnoj situaciji** te **primijeniti mjerenje pri rješavanju problema** iz matematike, ostalih odgojno-obrazovnih područja i svakodnevnog života.

odrediti mjeriva obilježja jednostavnog objekta ili pojave u svakodnevnim situacijama i primijeniti mjerenje pri rješavanju problema.



odrediti mjeriva obilježja objekta ili pojave u svakodnevnim situacijama i primijeniti mjerenje pri rješavanju problema.

odrediti mjeriva obilježja objekta ili pojave u svakodnevnoj situaciji te primijeniti mjerenje pri rješavanju problema iz matematike, ostalih odgojno-obrazovnih područja i svakodnevnog života.

odrediti mjeriva obilježja objekta ili pojave u svakodnevnoj situaciji, odabrati primjerene mjerne jedinice i mjerne uređaje, te primijeniti mjerenje pri rješavanju problema.

Sličnost, udaljenost i ostala mjerenja

- Kod prevođenja sličnost nije simetrična funkcija budući da je jednostavnije izbaciti neki dio nego ga dopisati
- Udaljenost ne mora biti simetrična funkcija

Matematika:

Je li jednostavniji izraz

$$\frac{x^{10} - 1}{x - 1}$$

ili

$$x^9 + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$$



Gramatika

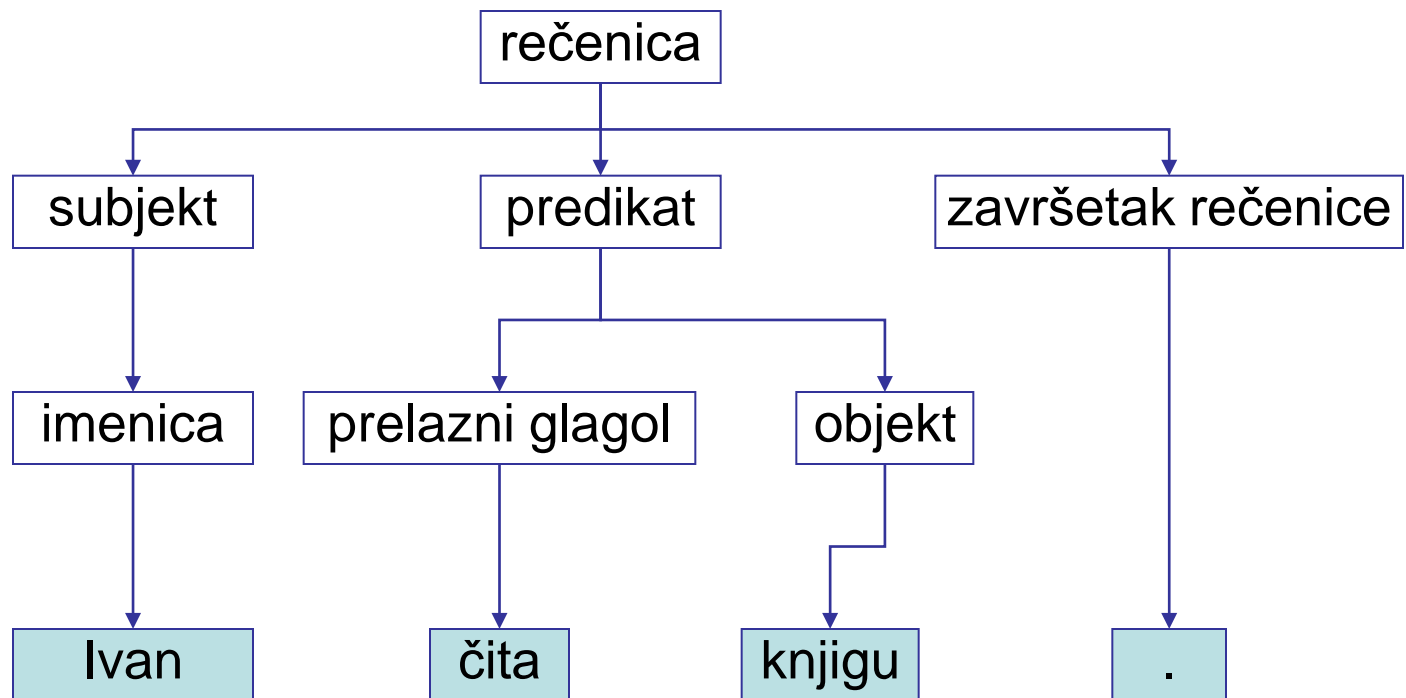
- Svaki govornik nekog jezika ima u glavi skup pravila za korištenje jezika. Taj skup pravila naziva se **gramatika**. Kod materinjeg jezika ova pravila nisu prikupljena svjesnim izučavanjem već slušanjem drugih govornika. Velik dio usvajanja gramatike dešava se tijekom djetinjstva. Kasnije učenje jezika uključuje viši stupanj formaliziranja gramatičkih pravila
- Formalna gramatika je **skup pravila za generiranje riječi** nekog jezika. Pravilo opisuje kako se slažu nizovi znakovi abecede (riječi, rečenice) koji zadovoljavaju sintaksu jezika. Gramatika ne opisuje značenje riječi.

Formalna gramatika

- Formalna gramatika je skup pravila za transformaciju znakovnih nizova. Uključuje početni simbol s kojim mora započeti svaka takva transformacija.
- Gramatikom može **generirati** jezik, tj. sve njegove riječi.
- Na temelju gramatike se mogu konstruirati automati koji **prepoznaju** riječi nekog jezika, tj. odrediti je li neka riječ (rečenica) gramatički ispravna.

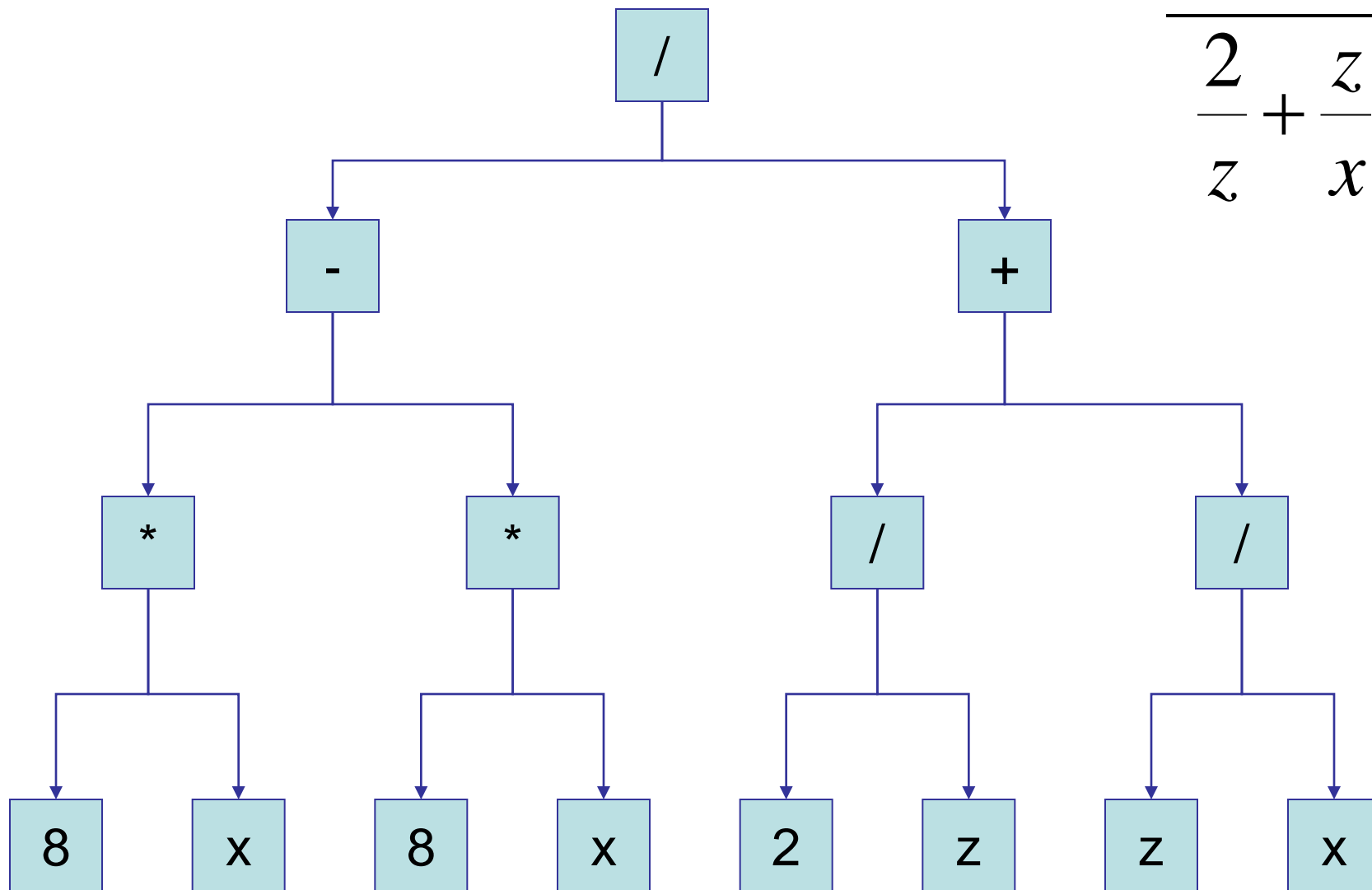
Sintaktička analiza (parsing)

- Prepoznavanje iskaza (rečenica u govornom jeziku) rastavljanjem na temelju gramatike tog jezika.
- Prvi korak u opisivanju značenja iskaza je rastavljanje na sastavne dijelove i promatranje tako rastavljenog oblika - stabla raščlambe.



Stablo raščlambe:

$$\frac{8x - 4z}{\frac{2}{z} + \frac{z}{x}}$$



Višedimenzionalnost i projekcija

- *odrediti mjeriva obilježja objekta ili pojave u svakodnevnim situacijama i primijeniti mjerenje pri rješavanju problema.*
- Označimo rečenicu:
- ((odrediti,odrediti,glagol, ε), (mjeriva**a**,mjeriv, pridjev, $\varepsilon > a$), (obilježja**a**,obilježje**e**,imenica, $e > a$), (objekta**a**,objekt,imenica, $\varepsilon > a$), (ili,ili,veznik-ili, ili), (pojave**e**,pojava**a**,imenica, $a > e$), (u,u,prijedlog-u,u), (svakodnevn**om**,svakodnevan**an**,pridjev, $an > nom$), (situacija**ma**,situacija,imenica, $\varepsilon > ma$), (i,i,veznik-i, i), (primijeniti,primijeniti,glagol, ε), (mjerenje,mjerenje,imenica, ε), (pri,pri,prijedlog-pri, pri), (rješavanju**u**,rješavanje**e**,imenica, $e > u$), (problem**a**,problem,imenica, $\varepsilon > a$))

Projekcija

- Niz sastavljen od drugih komponenti:

odrediti mjeriv obilježje objekt ili pojava u svakodnevnoj situaciji i primijeniti mjerenje pri rješavanju problema

- Niz sastavljen od trećih komponenti:

*glagol pridjev imenica imenica veznik-ili imenica prijedlog-u
pridjev imenica veznik-i glagol imenica prijedlog-pri imenica,
imenica*

- Niz sastavljen od četvrtih komponenti:

$\varepsilon \varepsilon > a$ $e > a$ $\varepsilon > a$ ili $a > e$ u $an > nom$ $\varepsilon > ma$ i $\varepsilon \varepsilon$ pri $e > u$ $\varepsilon > a$

Matematika:

Računanje modulo n , vektori, razlomci

Klase općenito

Automatski prijevod

- odrediti mjeriva obilježja objekta ili pojave u svakodnevnim situacijama i primijeniti mjerenje pri rješavanju problema.
- **mjeriva** determine **the characteristics of the object** or occurrence in everyday situations and apply **the measurement to solve** problems. (Google lipanj 2010)
- define measurable **attributes of the object** or occurrence in everyday situations and apply the measure **to solve problems**. (Google prosinac 2010)

- odrediti mjeriv obilježje objekt ili pojava u svakodnevnan situacija i primijeniti mjerenje pri rješavanje problem
- determine **mjeriv feature object** or occurrence in everyday situations and apply measurement **in solving** problems (Google lipanj 2010)
- define measurable **feature object** or occurrence in everyday situations and apply the measure **in solving the problem** (Google prosinac 2010)

Ima li automatsko prevođenje veze s matematikom?

- Slovo, riječ, jezik
- Značenje i višeznačne funkcije
- Prijevod i relacija prevođenja
- Nekomutativne operacije s riječima
- Faktorizacije
- Funkcije udaljenosti i sličnosti
- Gramatika i simboličko računanje
- Višedimenzionalnost i projekcija

SPECIJALNI DODATAK

- Google Translation Toolkit
<http://translate.google.com/toolkit>
- Google Translator Toolkit Data API ([Labs](#))
<http://code.google.com/apis/gtt/>

"What am I doing?"

"Što ja to radim?"



"I'm standing.

"Ja sam stajao.

"What is he doing?"

"Ono što je on radi?"



"He's sitting.

"On je sjedio.

"What is she doing?"

"Što je ona radi?"



"She's walking.

"Ona je hodanje.

"What is it doing?"

"Što je to radiš?"



"It's running.

"To je trčanje.

"What are they doing?"

"Što su oni radili?"



"They're playing."

"Oni igraju."

"What are we doing?"

To surf

"We're surfing the Internet.

"Ono što smo mi radiš?"

"Mi smo surfanje Internetom

- Hvala na pažnji. Nadam se da vam je bilo zanimljivo.
- Thank you for your attention. I hope you find it interesting.

<p>Hvala na pažnji.</p> <p>Nadam se da vam je bilo zanimljivo</p>	<p>Thanks on attention.</p> <p>Hope the that you is any interesting.</p> <p>G lipanj 2010</p>	<p>Thanks on attention.</p> <p>Hope the to you is any interestingly</p> <p>G prosinac 2010</p>
---	--	---