

Univerza v Ljubljani
Fakulteta za računalništvo
in informatiko



PRODUKCIJA MULTIMEDIJSKIH GRADIV (PMG)

BESEDILO

Borut Batagelj

V1.3
2023



Tekst

- Dvojna vloga:
 - Vizualna predstavitev jezika
 - Grafični element



```

1 function [Xm,U,L]=pca(X,K);
2 %PCA Principal Component Analysis.
3 % [Xm,U,L] = pca(X) creates an eigenspace from data matrix X.
4 % It returns mean image Xm, eigenvectors in columns of U and eigenvalues L.
5 % Each column of X is a vector representing an image.
6 % [Xm,U,L] = pca(X,K) returns only the first K eigenvectors and eigenvalues.
7
8 [M N]=size(X);
9 Xm=mean(X,2);
10 Xd=X-repmat(Xm,1,N);
11 if (N < M) %less images than image length
12   C=Xd'*Xd;
13   [V D Ut]=svd(C);
14   U=Xd*V;
15   U=U./repmat(sqrt(diag(D)') ,M,1);
16 else %more images than image length
17   C=Xd*Xd';
18   [U D Ut]=svd(C);
19 end;
20 L=diag(D)'/N;
21
22 if nargin>1
23   U=U(:,1:K);
24   L=L(1:K);
25 end;

```

ABCDEFGHIJKL
 MNOPQRSTUVWXYZ
 xpyzàâéîõabcdefghijkl
 klmnopqrstuvwxyzaáéîõ
 &1234567890(\$£.,!?)





Vsebina in oblika

- Vsebina in oblika sta v glavnem ločeni, ne pa povsem
 - Izgled lahko **poudari besedo**

The Right Hon was a tubby little chap who looked as if he had been poured into his clothes and had forgotten to say 'When!'

THE RIGHT HON was a tubby little chap who looked as if he had been *poured* into his clothes and had forgotten to say 'When!'



Množica znakov

- Abeceda
 - Množica vseh simbolov, ki jih uporabljamo:
- Znakovni nabor (*character set*)
 - Preslikava med znaki in števili
- Repertoar znakov
 - Vsi znaki iz abecede z definiranimi kodami
 - Anglija: $26 \cdot 2 + 10 + x$
 - Slovenija: $25 \cdot 2 + 10 + x$
 - Rusija: $33 \cdot 2 + 10 + x$
 - Japonska (Kanji): $1945 + 166 + \dots$, več kot 6000 znakov

a b c d e f g
 h i j k l m n
 o p q r s t u
 v w x y z

А Б В Г Д Е
 Ё Ж З И Й К
 Л М Н О П Р
 С Т У Ф Х Ц
 Ч Ш Щ Ъ Ы
 Ь Э Ю Я

良	農	業	商	議	選	員	記	運	轉	者	事
味	試	次	難	形	適	當	同	違	正	惡	点
念	殘	落	格	果	受	說	接	面	驗	合	指
約	決	旅	消	流	深	案	投	洗	打	扌	折
備	準	到	發	線	泊	特	絡	連	急	談	相
注	押	意	故	路	信	局	機	關	割	交	引
用	器	願	知	求	台	具	取	自	窓	由	營
期	產	個	價	品	資	銀	誌	雜	辭	服	紙

• Standardizacija !



Standardi - ASCII

- ASCII (American Standard Code for Information Interchange)
- Od leta 1970 (Lyndon B. Jonson leta 1986 odredil podporo)
- 7 bitna kodna tabela
 - $2^7=128$ kod
 - 8. bit za popravljanje napak
 - 95 znakov + 32 kontrolnih + del
- Ameriška angleščina

- Standard ISO 646 leta 1972
 - Vsebuje tudi nacionalne različice

32		33 !	34 “	35 #
36 \$		37 %	38 &	39 ‘
40 (41)	42 *	43 +
44 ,		45 -	46 .	47 /
48 0		49 1	50 2	51 3
52 4		53 5	54 6	55 7
56 8		57 9	58 :	59 ;
60 <		61 =	62 >	63 ?
64 @		65 A	66 B	67 C
68 D		69 E	70 F	71 G
72 H		73 I	74 J	75 K
76 L		77 M	78 N	79 O
80 P		81 Q	82 R	83 S
84 T		85 U	86 V	87 W
88 X		89 Y	90 Z	91 [
92 \		93]	94 ^	95 _
96 `		97 a	98 b	99 c
100 d		101 e	102 f	103 g
104 h		105 i	106 j	107 k
108 l		109 m	110 n	111 o
112 p		113 q	114 r	115 s
116 t		117 u	118 v	119 w
120 x		121 y	122 z	123 {
124		125 }	126 ~	



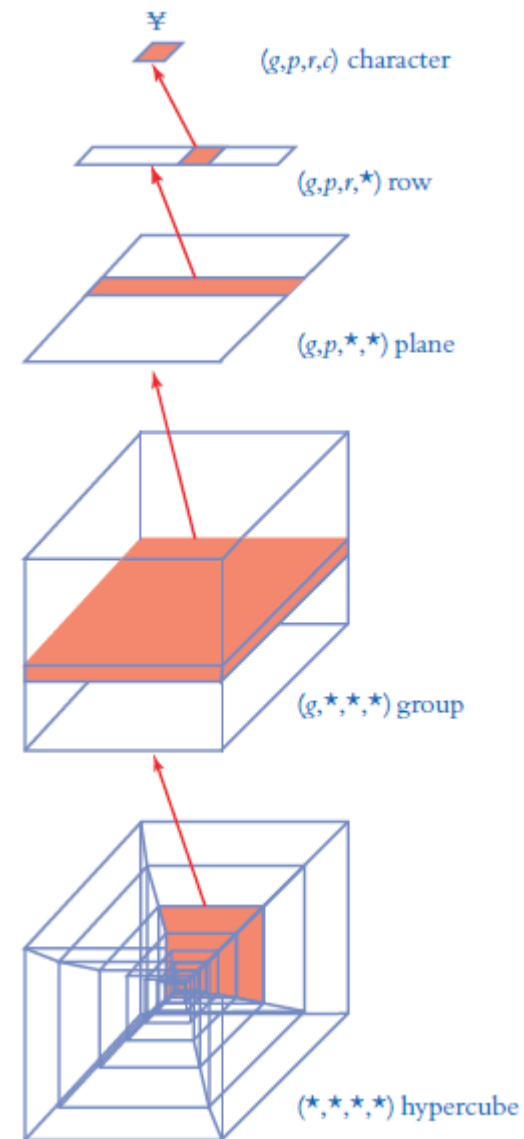
ISO 8859

- Standardni nabori za posamezne skupine jezikov
 - ISO Latin standard
- 8 bitne kodne tabele
 - Spodnja polovica = ASCII (ISO 646-US)
 - Zgornja polovica (več podskupin)
 - ISO 8859-1: ISO Latin 1
 - Zahodna Evropa
 - ISO 8859-2: ISO Latin 2
 - Centralna Evropa, naši znaki
- Še vedno ne omogoča uporabe več jezikov hkrati

160		161 i	162 ¢	163 £
164 □	165 ¥	166	167 §	
168 ``	169 ©	170 *	171 «	
172 ¬	173 -	174 ®	175 ´	
176 °	177 ±	178 º	179 ³	
180 ´	181 µ	182 ¶	183 ·	
184 ¸	185 ¹	186 º	187 »	
188 ¼	189 ½	190 ¾	191 ¸	
192 À	193 Á	194 Â	195 Ã	
196 Ä	197 Å	198 Æ	199 Ç	
200 È	201 É	202 Ê	203 Ë	
204 Ì	205 Í	206 Î	207 Ï	
208 Ð	209 Ñ	210 Ò	211 Ó	
212 Ô	213 Õ	214 Ö	215 ×	
216 Ø	217 Ù	218 Ú	219 Û	
220 Ü	221 Ý	222 Þ	223 ß	
224 à	225 á	226 â	227 ã	
228 ä	229 å	230 æ	231 ç	
232 è	233 é	234 ê	235 ë	
236 ì	237 í	238 î	239 ï	
240 ð	241 ñ	242 ò	243 ó	
244 ô	245 õ	246 ö	247 ÷	
248 ø	249 ù	250 ú	251 û	
252 ü	253 ý	254 þ	255 ÿ	

UCS – ISO 10646

- USC – Universal Character Set
- 32 – bitov = $2^{32} = 4.294.967.296$
- 4-dimenzionalna kocka
 - 256 skupin
 - 256 ravnin
 - 256 vrstic
 - 256 znakov v vrstici (8bit množica)
- Vsak znak je določen z (g,p,r,c)
- ISO Latin 1 = $(0,0,0,*)$





Unicode

- Razvoj v industrijskem konzorciju
- 16 bitov
- Drugačna struktura
 - CJK konsolidacija
 - Predstavitev Korejskih, Japonskih in Kitajskih znakov
- Kode za
 - Praktično vse (pomembne) (moderne) jezike
 - Ločila
 - Matematične simbole in znake
 - Posebne simbole
 - Akcente, ipd. (mogoče kombinirati z ostalimi znaki)
- 39000 kod porabljenih
- 6400 kod rezerviranih za privatno uporabo
- Združitev s standardom ISO
 - Unicode = ISO 10646 (0,0,*,*) =
= BMP (Basic Multilingual Plane)



Kodiranje

- Encoding
 - Naslednji nivo preslikave
 - Preslika kodo v zaporedje bajtov (za hranjenje in prenos)
- [QP](#) (Quoted Printable)
 - Kodiranje v zaporedja 7-bitnih znakov
- UCS-4: štirje bajti za ISO-10646 četvorčke
 - Najbolj univerzalen, ne najbolj ekonomičen
- UCS-2: brez zgornjih dveh ničel pri BMP = Unicode
 - Veliko bolj primeren za BMP



UCS Transformation Formats (UTF)

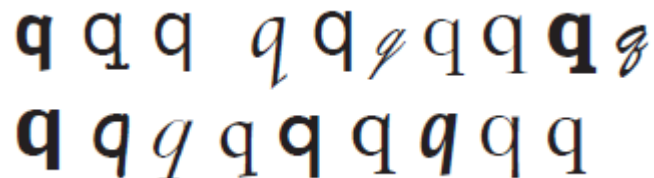
- UCS Transformation Formats (UTF)
 - UTF8
 - Izboljšan UCS-2
 - Zaporedje 8-bitnih polj = bajtov
 - Privzet standard v prog. jezikih, operacijskih sistemih, ...

Bits of code point	First code point	Last code point	Bytes in sequence	Byte 1	Byte 2	Byte 3	Byte 4	Byte 5	Byte 6
7	U+0000	U+007F	1	0xxxxxxx					
11	U+0080	U+07FF	2	110xxxxx	10xxxxxx				
16	U+0800	U+FFFF	3	1110xxxx	10xxxxxx	10xxxxxx			
21	U+10000	U+1FFFFF	4	11110xxx	10xxxxxx	10xxxxxx	10xxxxxx		
26	U+200000	U+3FFFFFFF	5	111110xx	10xxxxxx	10xxxxxx	10xxxxxx	10xxxxxx	
31	U+4000000	U+7FFFFFFF	6	1111110x	10xxxxxx	10xxxxxx	10xxxxxx	10xxxxxx	10xxxxxx

- UTF-7: UTF-8 prilagojen za zaporedja 7-bitnih znakov
- UTF-16: omogoča kodiranje izven UniCoda (BMPja)
- UTF-32: omogoča zapis s 4 bajti (21 bitov na razpolago)



Pisava



- Pisave (*angl. fonti*)
 - Različen izgled istega znaka
 - Glifi (*angl. glyphs*)
- Prenašajo se znaki, izrisujejo se kot glifi
 - Velika kompresija
- Dostop do različnih pisav:
 1. Glifi so shranjeni v sistemu
 - Najbolj enostavno, za multimedijo ni dovolj
 - Ni nujno, da imamo isto pisavo
 2. Pisava je vgrajena v dokument
 - Vsi dokumenti to ne omogočajo
 - Poveča velikost datotek
 3. Pisava se naloži z oddaljenega strežnika
 - Uporabnik mora imeti dostop
- Neobstoječa pisava se nadomesti z nadomestno
 - Neoptimalno (različen izgled, širina, znaki, ...)





Vrste pisav

- Pisave s fiksno ali spremenljivo širino

Monospaced Font: Courier
Each letter occupies the same amount of horizontal space, so that the text looks as if it was typed on a typewriter.

Proportional Font: Bembo
Each letter occupies an amount of horizontal space proportional to the width of the glyph, so that the text looks as if it was printed in a book.

- Pisave s serifi (zavihki) ali brez
 - Serif
 - Sans serif



Sans Serif Font: Univers
The letters of a sans serif (or sanserif) font lack the tiny strokes known as serifs, hence the name. They have a plain, perhaps utilitarian, appearance.



Vrste pisav

- Pokončna ali poševna oblika
 - Običajna pisava
 - „Italic“ pisava
 - „Slanted“ pisava
 - Transformirana pokončna
- Kaligrafična pisava
- Posebni tipi pisav
 - Rokopis
 - Igrive pisave

Fantasy Font: Jokerman

Fantasy fonts defy characterization, and often break all the rules. They are easily over-used.

Italic Font: Bembo Italic

The letters of an italic font slope to the right, and are formed as if they were made with an italic pen nib. Italics are conven-

Slanted Font: Lucida Bright Oblique

The letters of a slanted font share the rightward slope of italic fonts, but lack their calligraphic quality. Slanted fonts

Calligraphic Font: Apple Chancery
Calligraphic fonts usually resemble 'round hand' or 'copperplate' handwriting, unlike italic fonts.

Handwriting Font: Kidprint

Handwriting fonts are based on samples of real people's handwriting, so they are often quite idiosyncratic.



Vrste pisav

- Teža pisave
 - Poudarjena pisava
 - posebna skupina „Bold“
- Širina pisave
 - Stisnjena pisava (Condensed)
 - Razširjena pisava (Extended)
- Velikost pisave
 - (Skoraj) poljubna velikost
- Družine pisav
 - Različne različice iste pisave
- Uporaba pisav
 - Iz iste družine
 - Iz zelo različnih družin

Condensed
Condensed Italic
Bold Condensed
Bold Condensed Italic
Regular
Italic
Semibold
Semibold Italic
Bold
Bold Italic

Text that is set in Bembo goes well with *Bembo italic* and ***Bembo bold italic*** but it looks quite wrong mixed with Cheltenham and *Cheltenham italic*. (All this text is 12 pt.)



Izbira pisave

Za besedilo

- Nevsiljiva
 - ne odvrne pozornosti
 - ne posega v sporočilo
- Enostavna za branje
- Ne utruja
- Običajno: pokončna, serifna, normalne debeline

Za prikaz

- dekorativne/naslovne/prikazne pisave
- da pritegne



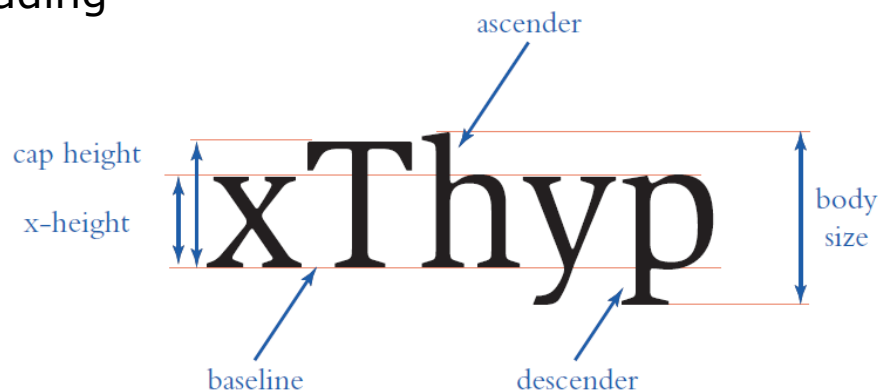
Terminologija

- Velikost pisave: pike (points = pt)
- Razdalja med vrsticami: pica = pc = 12pt
- Velikost črke (body size)
- Izhodiščnica (baseline)
- Razdalja med izhodiščnicama = leading
- Velikost znaka x=ex
 - Podaljšek navzgor
 - Podaljšek navzdol
 - H-višina: velikost velike črke
- em=širina črke M,
- en=širina črke N
- Kombinacija
 - Kerning; razdalja med znaki
 - Ligature: sestavljeni znaki iz več grafemov

kerned not kerned

AV AV

Ta Ta



fine fluffy soufflés
fine fluffy soufflés



Formati digitalnih pisav

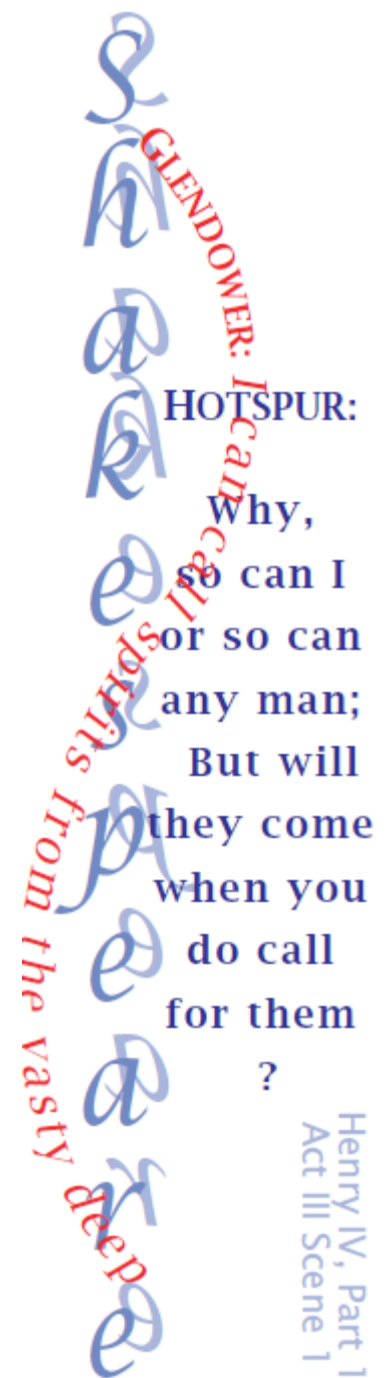
- Rasterske pisave (*bitmapped fonts*)
 - Lastnosti rasterskih slik - kvaliteta
 - Hitro enostavno prikazovanje
 - Glajenje robov
- Vektorske pisave (*outline fonts*)
 - **Adobe Type 1** (PostScript fonts)
 - Do 256 glifov
 - Glif je majhen program v jeziku PostScript, Biezjeve krivulje
 - **TrueType**
 - Do 256 glifov
 - Zaporedje točk, ki orisujejo glif s krivuljami drugega reda
 - **OpenType**
 - Do 65536 glifov, temelji na UniCode
 - Podpira Type1 in TrueType
 - Dodatne razširitve





Postavitve in oblikovanje

- Oblikovanje pisave (inline formatting)
 - Definira lastnosti pisave
- Oblikovanje odstavkov
 - Poravnava
 - Zamik teksta
 - Nov odstavek
 - Presledek med vrsticami
 - Okvirji okrog bloka
 - Stolpci, tabele
 - Postavitve blokov na stran





Urejanje

- WYSIWYG oz. „What you see is what you get“
 - Urejanje v vizualnem okolju
- Urejanje z označevalnim jezikom
 - Oznake za oblikovanje teksta (HTML, Latex, ...)
 - Tudi WYSIWYG sistemi generirajo označen tekst

```
Page Add media: Visual HTML  
b i link b-quote del ins img ul ol li code more lookup close tags  
  
<h3>EU FP7 ICT Cognitive Systems Large-Scale Integrating Project</h3>  
<h2>CogX - Cognitive Systems that Self-Understand and Self-Extend</h2>  
The high level aim of this project is to develop a unified theory of self-understanding and self-extension  
with a convincing instantiation and implementation of this theory in a robot. By self-understanding we  
mean that the robot has representations of gaps in its knowledge or uncertainty in its beliefs. By  
self-extension we mean the ability of the robot to extend its own abilities or knowledge by planning  
learning activities and carrying them out. The project involves six universities and about 30 researchers.  
<table border="0">  
<tbody>  
<tr>  
<td><a href="/results/dora/"></td>  
<td><a href="/results/george/"></td>  
</tr>  
</tbody></table>  
<div class="footer">The project began on 1st May 2008, and runs until June 30th 2012.  
  
If you have any queries about the project, please contact the project coordinator <a title="Jeremy  
Wyatt's home page" href="http://www.cs.bham.ac.uk/~jlw" target="_blank">Jeremy Wyatt</a>.  
  
 
```



Označevanje

- Vizualno označevanje
 - Oznake posamično vizualno oblikujejo izpis (pisavo, odstavek, ...)
- Strukturno označevanje
 - Oznake določajo logične elemente dokumenta (naslovi, sezname, tabele, ipd.)
 - Vsi takšni elementi v besedilu so enako oblikovani
 - Konsistentna postavitve v celotnem dokumentu
 - Hitro spreminjanje oblike celotnega besedila
 - Lahko definiramo različne postavitve za različne medije
 - Razdelitev na strukturo in izgled besedila (vsebino in obliko)
 - HTML in CSS
 - Poimenovanje posameznih elementov – hitro iskanje
 - Učinkovita raba predlog (stylesheets)



Vsebina in oblika

- Besedilo je sestavljeno iz vsebine, ki jo sporoča, in oblike, ki je uporabljena za prikaz

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.1//EN"
"http://www.w3.org/TR/xhtml11/DTD/xhtml11.dtd">

<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="en"
<head>
  <title>Deft Widgets</title>
  <meta http-equiv="Content-Type"
  content="application/xhtml+xml; charset=utf-8"/>
  <link rel="stylesheet" href="styles.css" type="text/css"
  media="screen" title="standard styles" charset="utf-8" />
</head>
<body>
<ul class="navbar">
  <li class="current">Tema</li>
  <li><a href="tis">Tis</a></li>
  <li><a href="rolod">Rolod</a></li>
  <li><a href="muspi">Muspi</a></li>
  <li><a href="merol">Merol</a></li>
</ul>
<div id="content">
  <ul class="sidebar">
    <li>Lorem ipsum dolor sit amet</li>
    <li>Consectetur adipisicing elit</li>
    <li>Sed do eiusmod tempor</li>
    <li>Incididunt ut labore</li>
    <li>Dolore magna aliqua</li>
  </ul>
  <h1>Deft Widgets</h1>
  <p>Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed
  do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim
  ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip
  ex ea commodo consequat. Duis aute irure dolor in reprehenderit in
  voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint
  occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit
  anim id est laborum.</p>
  <p>
  <p>Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed
  do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim
  ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip
  ex ea commodo consequat. Duis aute irure dolor in reprehenderit in
  voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint
  occaecat cupidatat non proident, sunt in culpa qui officia deserunt mollit
  anim id est laborum.</p>
  </div>
</body>
</html>
```

```
{
  margin: 0;
  padding: 0;
  font-size: 100%;
}
body {
  font-family: "Lucida Grande", Lucida, Helvetica, sans-serif;
  font-size: medium;
  width: 30em;
  padding: 0 2em;
  border-left: 15em solid #DAB2D5;
  background-color: #DDD;
  margin: auto;
}
h1 {
  font-family: Optima, Times, Georgia, serif;
  letter-spacing: 0.5em;
  font-size: 1.1em;
  font-weight: bold;
  text-align: center;
  padding-bottom: 0.8em;
  text-transform: uppercase;
}
p {
  text-align: justify;
  padding-bottom: 1em;
}
ul.sidebar {
  font-family: Optima, Times, Georgia, serif;
  list-style-type: none;
  float: left;
  margin-left: -15em;
  padding-top: 2.2em;
  width: 12em;
}
.sidebar li {
  padding-bottom: 1.2em;
}
ul.navbar {
  list-style-type: none;
  padding: 1.5em 0 1em 0;
  margin: 0 2em 0;
  border-bottom: 2px dotted #DAB2D5;
  text-align: center;
}
.navbar li {
  display: inline;
  padding: 0 1em;
  font-size: 1.2em;
}
.navbar li:first-child {
  padding: 0;
}
.navbar a {
  text-decoration: none;
  color: inherit;
}
.current {
  color: #DAB2D5;
}
```



- HTML dokument (za vsebino) in CSS dokument (za obliko)