

DIGIN₂₀₂₃



Univerza v Mariboru

Fakulteta za elektrotehniko,
računalništvo in informatiko



Uporaba tehnik za samodejno podnaslavljanje avdiovizualnih vsebin

Using auto-subtitle system methods in audiovisual content

Bogdan Dugonik

Motivacija

- Digitalne video tehnologije omogočile eksponentno naraščanje števila in obsega razpoložljivih video vsebin.
- Zakaj – podnaslavljanje video vsebine
(hrup, naglušnost, staranje, nekakovostna oprema za predvajanje, neprimerna akustična ureditev prostora)
- Kako – ročne ali avtomatske tehnike podnaslavljanja
– orodja za montažo videa z vgrajenim ASR (Auto speech recognition)
- Postopki in izvedba samodejnega podnaslavljanja videa



Podnapisi v videu

- Podnapisi ali podnaslovi **so skrajšani ali povzeti** zapisi govora avdiovizualnih vsebin.
- Podnapisi so lahko prevodi tujega jezika ali zapisi dialoga v avdiovizualnem delu.
- Umeščeni so nad spodnji rob slike in so časovno usklajeni z dialogom.
- Podnapisi AV vsebin imajo v Sloveniji dolgo tradicijo.
- Podnapisovanje velja za zamudno opravilo in velja za avtorsko delo.
- Obseg AV vsebin nenehno narašča in zahteva tudi uporabo strojnega razpoznavanja jezika.
- Strojno prepoznavanje jezika ASR z uporabo AI je za Slovenski jezik še nepopoln in zahteva precej ročnih korekcij.

Digitalna orodja v video produkciji

- 2000 – prehod iz analogne v digitalno obdelavo (zajem in montaža)
- 2002 – IT tehnologija omogoča prenos zvoka in videa po spletu (VoD) *
- 2003 – širokopasovna omrežja omogočajo več pasovne širine (pretočne video vsebine – streaming)
- 2006 – ustanovitev platforme YouTube, uporabnikom omogoča nalaganje in razširjanje videovsebin
- 2009 – predvajanje vsebin v 1080p; samodejno razpoznavanje govora (ASR)
- 2010 – video vsebine predvajane v živo (live stream) in 4K
- 2015 – predvajanje vsebin v 360° in 8K; Adobe Premiere vključi AI za razpoznavanje govora
- 2019 – YouTube je druga najbolj obiskana stran (za Googlom), 500 ur/na minuto
- 2022 – YT lokaliziran v 108 državah (od 2014 v Sloveniji), 15 M





Slovenščina



SHRANI OSNUTEK

OBJAVI



+ NAPIS

UREDİ BESEDILO

naša naslednja naloga je, da realiziramo popolni seštevalnik z multipleksijem

00:07:20

to je besedila naloge se pa glasi tako realizirajte popolni seštevalnik z uporabo dvojnega multipleksorja

00:09:20

00:09:20

ki ga imate seveda tudi na slikah v vajah. Tale dvojni multipleks v eles štiriinsemdeaset sto triinpetdeset ima v

00:11:20

00:11:20

enem samem integriranem veze torej v enem samem čipu dva enakovredna dela multiplekser a in multipleksr b

00:13:20

00:13:20

katerima so vhodi označini z in nič a i ena a idva a in i tri a oziroma nič b i nič b i ena ve idva b in itri b ter izhoda ipsilon a in ipsilon b

00:15:20

00:15:20

torej naslovna vhoda sta es nič in es ena in s tema naslovnima vodomoma sočasno naslavljammo multiplekse a kot tudi multipleksr b <spn>

00:17:20

00:17:20

00:19:20

00:00:00

← RAZVELJAVI

→ UVELJAVI



00:00:00

00:06:00

00:12:00

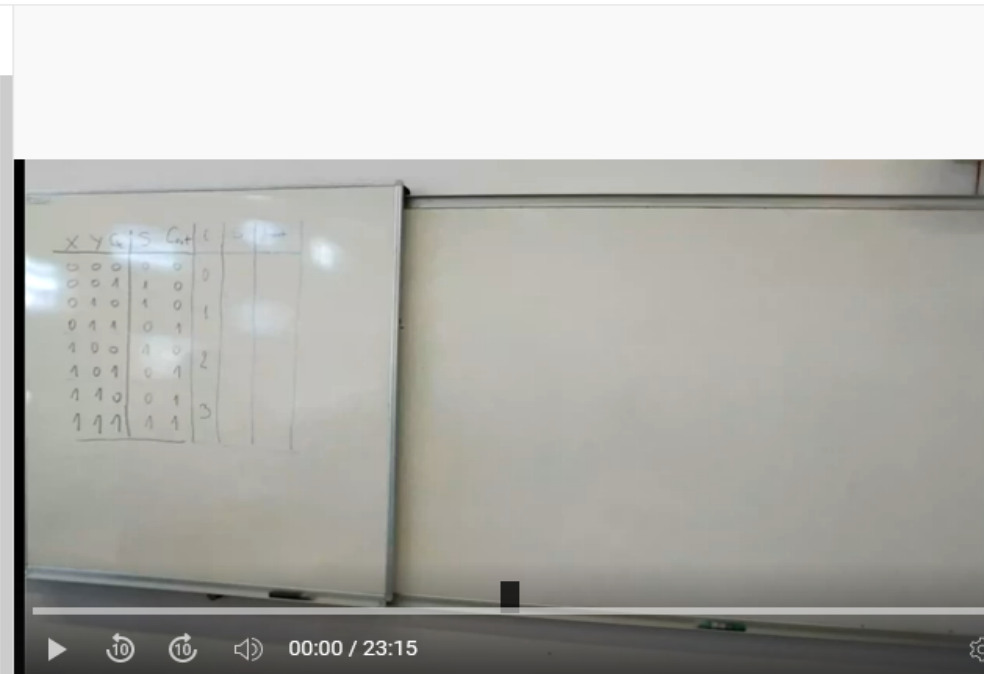
00:18:00

00:24:00

00:30:05



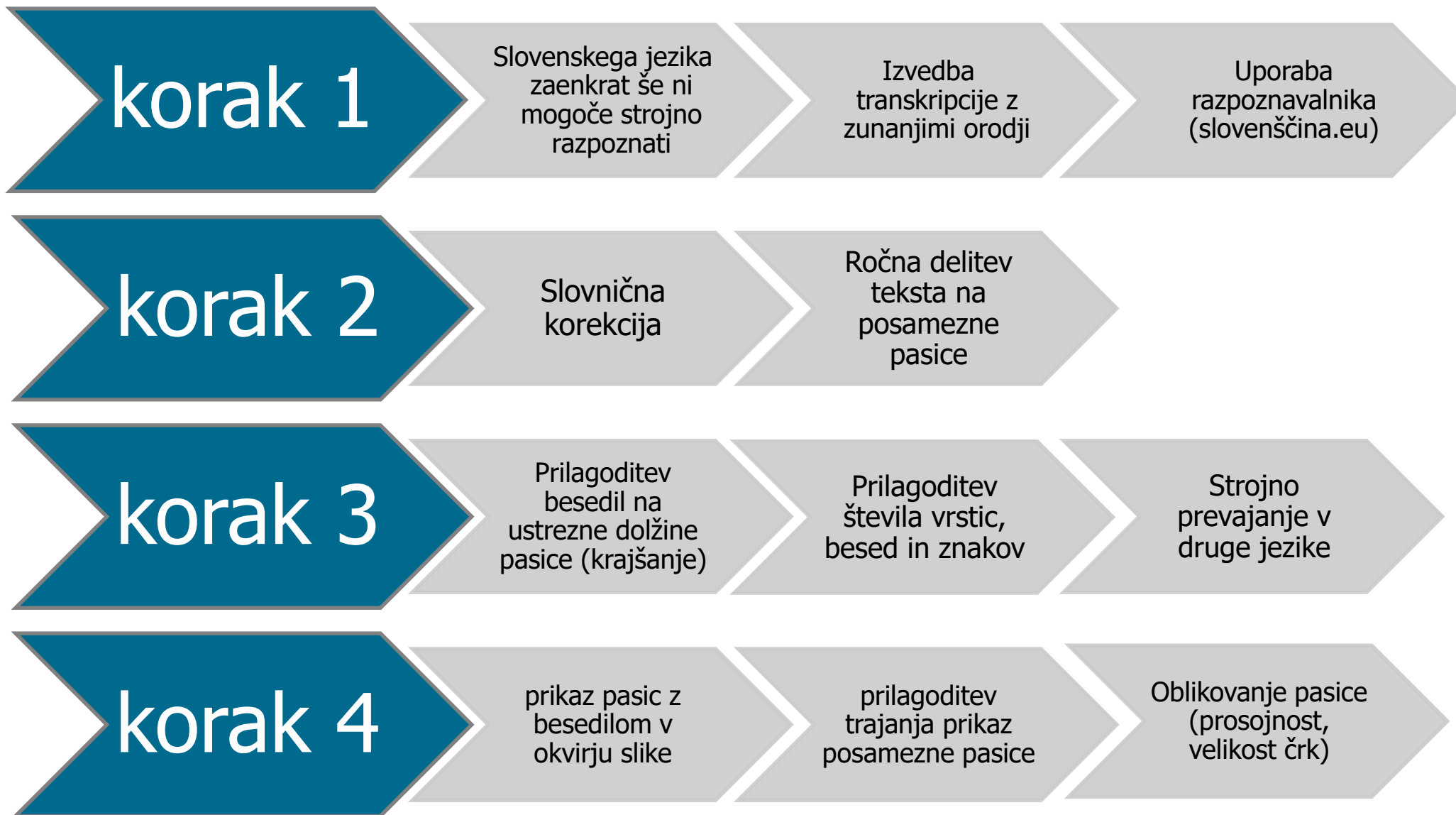
naša nas... to je bes... ki ga ima... enem sa... katerima ... torej nasl...



00:00 / 23:15

Z bližnjimi tipkami lahko podnapise vnesete hitreje.

 Premor med vnašanjem besedila





The screenshot displays the Adobe Premiere Pro interface. The main video preview window shows a person looking at a laptop screen displaying a 3D architectural model of a theater. The text "3sat" is visible in the top left of the video, and "za prostorski zvok?" is overlaid at the bottom. The timeline at the bottom shows a sequence of subtitle tracks with text in Slovenian, such as "Orkester", "potrebuje prvo...", "Kdo je...", "Arh...", "Ne... z zvokom...", "N... in...", "da v... Tu...", "Hrup... ob gla... po...", "Z... Le dobra... k vrhunski iz...", "V zad... ki...", "V... se je na...", and "Ena...".

On the left side, a settings menu is open, showing the following options:

- Language: English
- Speaker labeling: Spanish, German, French, Japanese, Portuguese (EU)
- Audio analysis: (no options listed)

On the right side, the Captions panel is visible, showing a list of subtitle tracks with their start and end times and the text of the subtitles. The current track is highlighted, showing the subtitle "za prostorski zvok?".



korak 1

Aktivacija funkcije ASR in izbira jezika za razpoznavanje

Transkripcija v podnaslove

Generiranje besedil na trak v časovnici

korak 2

Trak z besedili

Slovnični popravki besedil v oknu znotraj programa (samodejno za izbrane jezike)

Izvoz besedil v zunanji urejevalnik

korak 3

Prilagoditev besedila na ustrezne dolžine (krajšanje)

Prilagoditev števila vrstic, besed in znakov

Prevod v druge jezike

korak 4

namestitev pasice z besedilom v okvir slike

prilagoditev časa trajanja za prikaz posamezne pasice

Oblikovanje pasice za podnaslove



Untitled - Subtitle Edit 4.0.1

File Edit Tools Spell check Video Synchronization Auto-translate Options Networking Help

Format SubRip (.srt) Encoding UTF-8 with BOM

#	Start time	End time	Duration	Text

Audio to text

Generate text from audio via Whisper speech recognition Engine OpenAI

[Whisper website](#)

Languages and models

Choose language Slovenian Choose model large (2.9 GB) [Open models folder](#)

Translate to English

Auto adjust timings

Use post-processing (line merge, fix casing, punctuation, and more)

Transcribing audio to text... 1:05:58,809

Generate Batch mode Cancel



libmpv 2.1 00:00:02.720 / 00:39:15.455

Popolni seštevalnik realiziran z logičnimi vrati.mp4 1280x720 MP4

Start time 00:00:00,000 Duration 0,000 Text

< Prev Next >

Unbreak Auto br

Translate Create Adjust

Auto repeat Auto repeat on Repeat count (times) 2

Auto continue Auto continue on Delay (seconds) 2

Tip: Use <alt+arrow up/down> to go to previous/next subtitle

< Previous Play Next >

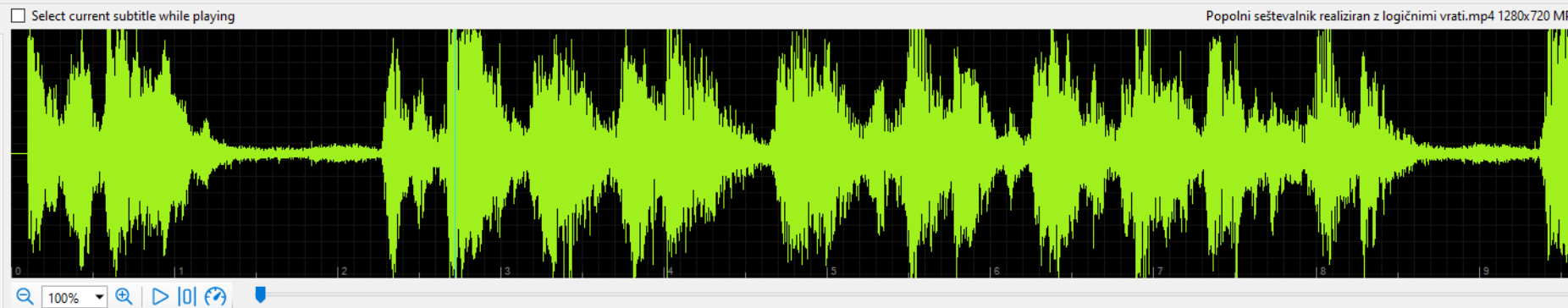
Pause

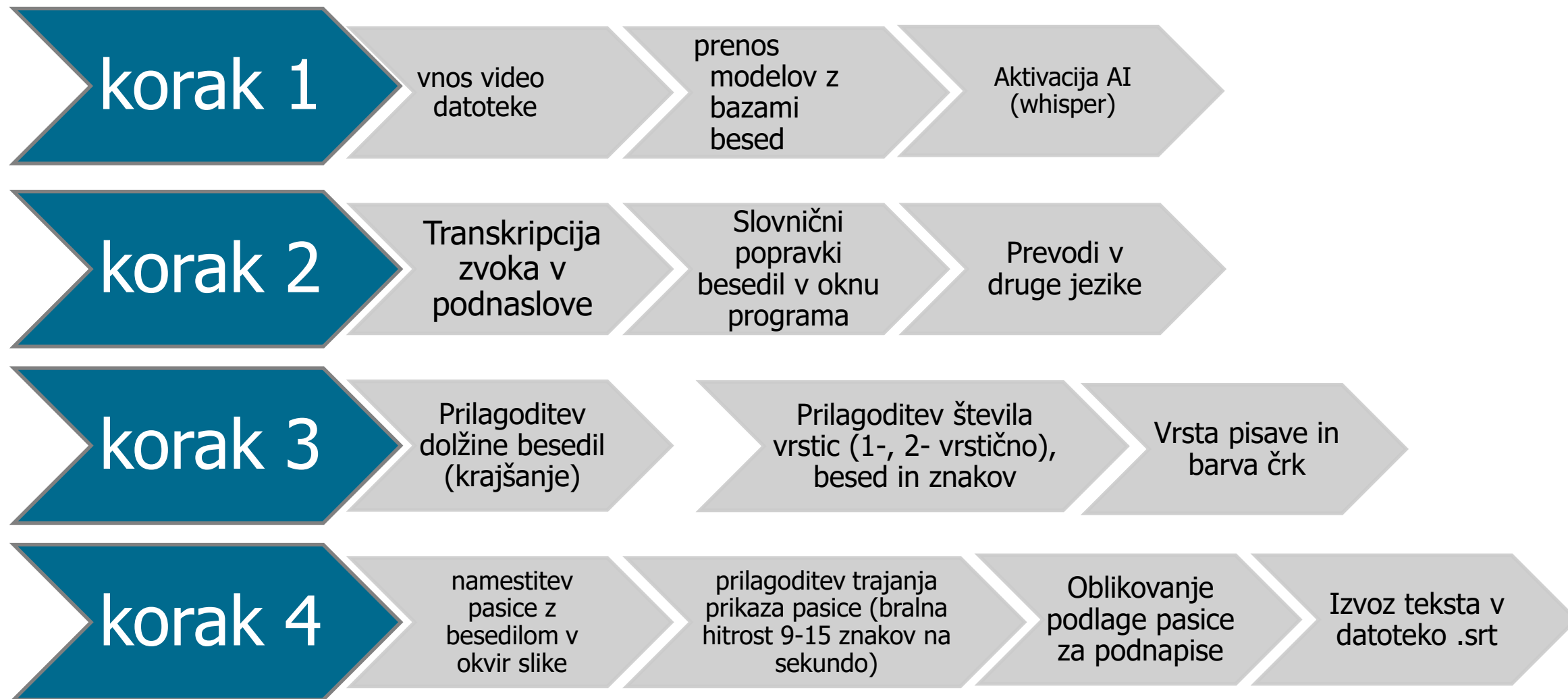
Search text online

Google it Google translate

The Free Dictionary

Wikipedia







Sklep

- Podnaslovi v avdiovizualnih delih vsekakor pripomorejo k lažjemu razumevanju vsebin gluhim in naglušnim, kot tudi starejšim osebam
- Izboljšajo razumljivost vsebin pri predvajanju v hrupnih ali akustično neurejenih prostorih
- Pogosto je zvok AV vsebin slabše kakovosti in posledično nerazumljiv
- Zmanjšajo napor pri spremljanju AV vsebin
- Zaradi vse večjega števila AV vsebin jih je nemogoče ročno podnasloviti, zato potrebujemo sistem, ki bodo omogočal kakovostno strojno podnaslavljanje.
- Hiter razvoj umetne inteligence bo zagotovo prinesel napredek v razvoju učinkovitih sistemov za strojno prepoznavanje jezika za hitro, kakovostno in tudi sprotno izvedbo podnaslavljanja avdiovizualnih vsebin.