

# Analizator teksta



## Što je Analizator teksta?

Analizator teksta, TextAn, je inovativna softverska komponenta kojom se omogućuje uvođenje višejezičnog komunikacijskog kanala nestrukturiranim tekstom između korisnika i aplikacija.

## Kako do Analizatora teksta?

Ukoliko ste zainteresirani za kupnju i ugradnju Analizatora teksta u Vašu aplikaciju ili uslugu kontaktirajte nas na [prodaja@capabilis.hr](mailto:prodaja@capabilis.hr)

# Analizator teksta

## *Kome je namijenjen Analizator teksta?*

---

Ugradbena komponenta TextAn je namijenjena kompanijama i pojedincima koji razvijaju softverske aplikacije i usluge temeljene na tekstualnoj komunikaciji korisnika (e-mail, SMS, trenutno poručivanje...) s aplikacijom ili uslugom.

## *Čemu je namijenjen Analizator teksta?*

---

TextAn je namijenjen prepoznavanju riječi i fraza te njihovog značenja iz slobodno kreiranog teksta u digitalnom obliku čime uvelike pomaže pružateljima usluga / davateljima aplikacija u realizaciji aplikacija za realne životne slučajeve.

TextAn prihvaća nestrukturirani, slobodno napisani, tekst u digitalnom obliku (e-mail, SMS, konverziju govora u tekst, poruke trenutnog poručivanja...), analizira ga, prepoznaje značenja i pretvara u strukturirani tekst pogodan za komunikaciju s aplikacijom, uslugom ili administratorom. Ovim se omogućuje da korisnik slobodno napisanim tekstom uputi zahtjev ili komandu aplikaciji, usluzi ili administratoru.

TextAn se može upotrijebiti i kao unificirani pristupni kanal tj. kao usmjerivač poruka prema različitim aplikacijama i uslugama ili različitim subjektima.

TextAn se može upotrijebiti za „zakonito presretanje“ (eng. lawful interception), „roditeljsku zaštitu djece“ i slično tj. za nadzor razmjene kratkih tekstualnih poruka (SMS), elektroničke pošte (e-mail), multimedijских poruka (MMS) ali i za nadzor glasovne komunikacije posredstvom specijaliziranih uređaja za konverziju govora u tekst (eng. speech-to-text).

# Analizator teksta

TextAn se može upotrijebiti za dubinsku analizu statičkih i dinamičkih sadržaja dostupnih na Internetu (forumi, društvene mreže itd.)

Obzirom da smanjuje formalizam te povećava komfor korištenja TextAn, svojim automatski prepoznavanjem slobodno napisanog teksta i njegovog značenja, olakšava penetraciju ICT usluga baziranih na razmjeni tekstualnih informacija. Općenito, TextAn se može primijeniti gdje god se radi o potrebi prepoznavanja riječi i grupa riječi tj. fraza i njihovog značenja, a da su navedene bilo kao strukturirani ili nestrukturirani tekst u digitalnom obliku.

## *Po čemu se Analizator teksta razlikuju od konkurentnih rješenja?*

---

Primjenom Analizatora teksta uklanjaju se nedostaci slijedećih, postojećih tehnoloških rješenja:

- Velik dio modernih aplikacija i usluga temelji se na komunikaciji tekstem (e-mail, SMS,..) zbog svoje jednostavnosti, dostupnosti i prihvatljivih cijena.
- Dobar dio aplikacija i usluga zahtijeva informatičku i gramatičku pismenost kod krajnjeg korisnika što nije uvijek slučaj.
- Aktualna komunikacija tekstem (e-mail, SMS) ima svoje zakonitosti: česte su manje greške u pisanju, rijetko tko se drži pravopisa, izostaju velika slova i interpunkcija, češća je primjena brojeva, itd.
- Nerijetko aplikacije i usluge koje se temelje na razmjeni tekstualnih poruka a istovremeno traže formalizam glede sintakse.
- Aplikacije i usluge sa strogim formalizmom imaju mnogo slabiju dinamiku prihvaćanja od onih sa manje naglašenim formalizmom.

# Analizator teksta

- Dodatni problem su aplikacije i usluge koje se ne koriste svakodnevno pa korisnici zaborave proceduru ili se boje da ne naprave nešto krivo i odustaju od usluge.
- Rješenja s mobilnim aplikacijama na pametnim telefonima rijetko će promptno pokriti sve potrebe tržišta usluga.

## *Koje su prednosti Analizator teksta?*

---

- Slobodan unos teksta bez straha od sintaksne ili gramatičke greške.
- Prihvatanje dijalekta, brojeva, kratica itd.
- Slobodan unos teksta pogodan za informatički slabije obrazovane.
- Slobodan unos teksta pogodan za rjeđe korištene usluge.
- Komunikacija s korisnikom putem e-Mail, SMS, MMS ili WAP.
- Komunikacija s korisnikom na jeziku korisnika.
- Prepoznavanje nestrukturiranog teksta i pretvaranje u zahtjev/komandu.

## *O čega se sastoji Analizator teksta?*

---

TextAn predstavlja komercijalnu komponentu pogodnu za ugradnju u postojeće ili nove ICT usluge sa digitalnim sadržajima, a sastoji se od slijedećih modula:

**Komunikator** - je modul koji predstavlja sučelje za komunikaciju sa okolinom. Okolina ovisi o vrsti i namjeni ICT usluge i o njezinoj arhitekturi. Modul komunicira sa drugim komponentama ICT usluge porukama HTTP / HTTPS protokola.

**Ekstraktor** - modul za izdvajanje ne-tekstualnih znakova i usmjeravanje elemenata teksta na daljnju obradu. Modul je zajednički za sve jezike.

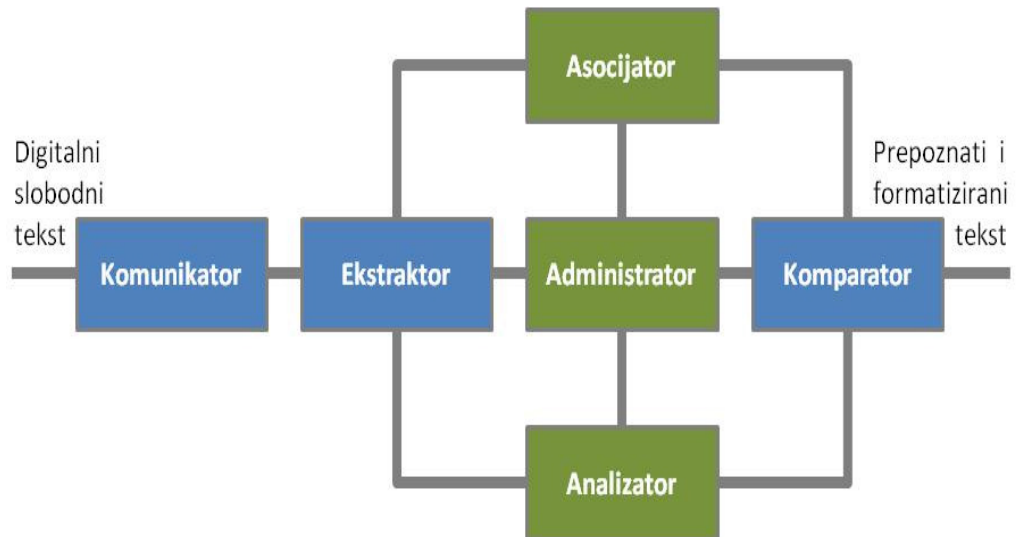
**Analizator** - modul koji provodi očitavanje, grupiranje i usmjeravanje digitalnog sadržaja a koristi se prethodno pronađenim značenjima nizova znakova koji su se pojavili u tekstu. Modul je izveden tehnologijom ternarnog stabla s dimenzijom prilagođenom brojnosti značenja koje usluga treba razvrstavati po kategorijama.

# Analizator teksta

Puna funkcionalnost modula postiže se učenjem nad cijelim poznatim skupom podataka.

**Administrator** - je modul preko kojeg se upravlja sveukupnim funkcionalnostima i prati rad cijele komponente. Upravljanje radom podrazumijeva oblikovanje podataka za učenje i pokretanje procesa učenja za pojedine module, nadzor ispravnosti komunikacijskih kanala, otkrivanje i eliminaciju konfliktnih uzoraka, te statistiku.

**Asocijator** – modul za asocijativno očitavanje značenja termina iz niza znakova. Modul je realiziran pomoću umjetnih neuronskih mreža, a punu funkcionalnost postiže postupkom učenja iz poznatih podataka. Svojstvo generalizacije omogućuje djelovanje komponente i nad skupom nepoznatih podataka. Modul je dizajniran tako da je generalizaciju moguće u potpunosti i po volji prilagoditi potrebnom opsegu. Svojstva modula prilagođavaju se ponavljanjem postupka učenja. Rezultat obrade pokazuje na značenje svakog pojedinog detektiranog niza znakova u nestrukturiranom tekstu, ukoliko ono postoji.



**Komparator** – modul za ocjenjivanje točnosti i usmjeravanje rezultata prepoznavanja. Ispravno prepoznato značenje se usmjerava prema usluzi, a svako potencijalno odstupanje od ispravnog značenja se usmjerava prema administratoru ili usluzi sa upozorenjem o razini potencijalne pogreške. Na temelju primljenog upozorenja pokreće se ručna ili automatska prilagodba modula prema definiranim pravilima usluge.

# Analizator teksta

## *Capabilis d.o.o.*

---

Capabilis d.o.o. je osnovana 2011. godine nakon uspješnog završetka tehnologijskog projekta sufinanciranog od strane Hrvatskog instituta za tehnologiju.

Kompanija je registrirana za pružanje usluga u području informacijske i komunikacijske tehnologije.

Istraživački i razvojni naponi Capabilis-a usmjereni su na stvaranje novih ideja, aplikacija i usluga koji donose novu vrijednost i rješavaju realne probleme na tržištu elektroničkih komunikacija.

Ako u području elektroničkih komunikacija imate:

Pitanje koje treba istražiti?

Ideju koju treba realizirati?

Problem koji treba softverski riješiti?

Onda je Capabilis Vaš najbolji partner!

### **Kontakt**

**Capabilis d.o.o.**  
Dragutina Golika 63  
10000 Zagreb  
Hrvatska  
[www.capabilis.hr](http://www.capabilis.hr)  
[info@capabilis.hr](mailto:info@capabilis.hr)