

FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY  
UNIVERZITA KOMENSKÉHO V BRATISLAVE

**Katalóg požiadaviek pre  
Anotátor pomenovaných entít  
(APE)**

*Alexander Szendy  
Juraj Kočan  
Marián Glatzner  
Michal Klein  
2016/2017*

# Obsah

<b>1 Úvod.....</b>	<b>3</b>
1.1 Ciele dokumentu.....	3
1.2 Rozsah projektu.....	3
1.3 Definície, skratky.....	3
1.4 Odkazy.....	3
1.5 Nasledujúce kapitoly.....	4
<b>2 Všeobecný popis.....</b>	<b>4</b>
2.1 Perspektíva projektu.....	4
2.2 Funkcie projektu.....	4
2.3 Charakteristika používateľov.....	4
2.4 Všeobecné obmedzenia.....	4
<b>3 Špecifické požiadavky.....</b>	<b>5</b>
3.1 Registrácia a prihlasovanie.....	5
3.2 Práva používateľa.....	5
3.3 Administrátor skupiny.....	5
3.4 Práva administrátora skupiny.....	5
3.5 Obsah anotovaného dokumentu.....	5
3.6 Vzťah dokument-skupina.....	5
3.7 Rozdelenie dokumentu.....	6
3.8 Skupiny.....	6
3.9 Tvorba skupiny.....	6
3.10 Default práva skupiny.....	6
3.11 Voľba skupiny.....	6
3.12 Anotácie.....	6
3.13 Referencie.....	7
3.14 Typy anotácií.....	7
3.15 Anotovanie.....	7
3.16 Zobrazovanie anotácií.....	7
3.17 Zobrazovanie zmien v dokumente.....	7
3.18 Viditeľnosť anotácií.....	7
3.19 Komentáre.....	7
3.20 Editovanie/mazanie komentárov.....	8
3.21 Zobrazovanie komentárov.....	8

3.22 Hľadanie v texte.....	8
3.23 Export.....	8
3.24 Aktivita v dokumente.....	9

# 1 Úvod

## 1.1 Ciele dokumentu

Cieľom tohto dokumentu je čo najbližšie popísať požiadavky zadávateľa projektu bez bližšieho určenia technickej špecifikácie ako bude projekt APE implementovaný. Slúži ako zoznam funkcií, ktoré hotový produkt musí spĺňať a na základe ktorých bude aj vyhodnotený. Dokument je určený pre všetkých ľudí, ktorí prídu priamo alebo nepriamo do kontaktu s týmto systémom alebo jeho vývojom.

## 1.2 Rozsah projektu

Našou úlohou je vytvoriť kolaboratívny textový editor pre jednoduchý text, v ktorom budú môcť viacerí používatelia spoločne anotovať jeden a ten istý dokument v reálnom čase naraz – zmeny vykonané jedným používateľom sa budú zobrazovať aj ostatným užívateľom.

## 1.3 Definície, skratky

- Skupina – ľudia, ktorí spoločne anotujú nejaký dokument
- Anotácia – časť textu, ktorá bude vyznačená
- Typ/kategória anotácie – každá anotácia má určitý typ (mesto, osobu, štát, čas, ...)
- NE – named entity, pomenovaná entita, zvyčajne ide o konkrétnu osobu, čas, mesto, organizáciu, ...
- UI – umelá inteligencia

## 1.4 Odkazy

Nasledujú odkazy na podobný softvér:

- <https://gate.ac.uk/>
- <http://brat.nlplab.org/index.html>

Príklad textu, ktorý sa bude môcť anotovať:

- [http://sagadb.org/egils\\_saga.en](http://sagadb.org/egils_saga.en)

## 1.5 Nasledujúce kapitoly

V nasledujúcich kapitolách bližšie špecifikujeme funkcie nášho systému, ako export dokumentov, prístupové práva pre používateľov a skupiny.

## 2 Všeobecný popis

### 2.1 Perspektíva projektu

APE bude mať využitie najmä v oblastiach UI, kde je často nevyhnutné program najskôr niečo naučiť – zvyčajne od človeka, ktorý rieši ten istý problém (v našom prípade ide o tzv. named entity recognition – či dané slovo/slová sú NE a ak áno, treba správne určiť kategóriu).

Keďže pôjde o kolaboratívny editor zobrazujúci zmeny v reálnom čase, anotáciu objemnejších textových dokumentov bude možné rozdeliť medzi viacero ľudí, čím sa zníži časová aj pracovná vyťaženosť človeka.

### 2.2 Funkcie projektu

Aplikácia bude rozdelená na dve časti: server a klient. Server bude slúžiť na spravovanie informácií týkajúcich sa skupín a šírenie komunikácie medzi používateľmi, zatiaľčo primárna funkcia klienta (s grafickým rozhraním) bude samotná anotácia dokumentu a zobrazovanie zmien pri anotácii.

Podrobnejšie informácie sú uvedené v špecifických požiadavkách [3].

### 2.3 Charakteristika používateľov

Používateľov delíme na dva typy:

- správca
- normálny používateľ (ďalej len používateľ)

Správca slúži na inštaláciu servera a jeho nasledovné spravovanie, nemá funkcie ako obyčajný používateľ.

Používateľ sa pripája na server cez klienta a jeho cieľom je anotovanie.

### 2.4 Všeobecné obmedzenia

Anotovať bude možné len jednoduché textové dokumenty. Dokumenty budú môcť mať rozličné konce riadkov ako aj kódovania (minimálne akceptované budú UTF-8 a ASCII).

## **3 Špecifické požiadavky**

### **3.1 Registrácia a prihlasovanie**

Po spustení aplikácie sa používatelia budú môcť zaregistrovať. Ak je používateľ už zaregistrovaný, tak pomocou prihlasovacieho mena a hesla sa bude môcť pripojiť na server. Používateľ si bude môcť určiť, či sa hneď po spustení aplikácie chce pripojiť na server na základe konfiguračného súboru.

### **3.2 Práva používateľa**

Práva používateľa sú určené podľa aktívnej skupiny. Práva sú nastavené po úspešnom pripojení do skupiny. Rozlišujeme medzi 3 základnými právami (tieto práva možno kombinovať):

- anotačné – umožňuje pridávať anotácie
- reanotačné – umožňuje preanotovať (meniť a mazať) anotácie iných
- editačné – umožňuje pridávať/meniť/mazať typy a podtypy anotácií v danej skupine

### **3.3 Administrátor skupiny**

Administrátor skupiny je jej tvorca. Každá skupina má práve jedného administrátora a ten je nemenný.

### **3.4 Práva administrátora skupiny**

Administrátor má všetky práva používateľa a takisto môže meniť štruktúru skupiny (pridať/meniť/odobrať: užívateľa alebo ich práva, heslo skupiny, názov, default práva). Administrátor nemôže meniť svoje práva.

### **3.5 Obsah anotovaného dokumentu**

Obsah dokumentu nemožno upravovať iným spôsobom ako anotovaním a komentovaním.

### **3.6 Vzťah dokument-skupina**

Dokument sa viaže ku skupine vždy v pomere 1:1 – jedna skupina anotuje práve jeden dokument, ktorý už nemôže zmeniť za iný.

### **3.7 Rozdelenie dokumentu**

Rozsiahle dokumenty budú rozdelené (podľa počtu viet) na viacero častí/stránok, medzi ktorými bude možné listovať. Dlhé texty sa budú dať aj scrollovať.

### **3.8 Skupiny**

Prihlásený používateľ budú môcť vytvárať skupiny alebo sa do nich pripájať za účelom kolaboratívnej tvorby.

### **3.9 Tvorba skupiny**

Ľubovoľný prihlásený používateľ bude môcť vytvoriť skupinu, kde bude môcť zvoliť nasledovné:

- názov skupiny
- dokument (na upload), ktorý bude skupina anotovať
- zoznam používateľov v skupine aj s ich prístupovými právami
- default práva (optional)
- heslo pre vstup do skupiny (optional)

### **3.10 Default práva skupiny**

Ak sú dané, tak sa nastaví používateľovi, ktorý sa chce pripojiť do skupiny, do ktorej zoznamu však nepatrí. Ak default práva nie sú zadané, do skupiny sa môžu pripojiť len ľudia v jej zozname.

### **3.11 Voľba skupiny**

Prihlásený používateľ si môže zvoliť svoju aktívnu skupinu zo zoznamu všetkých skupín. Aktívna skupina je vždy maximálne jedna. Používateľ má možnosť sa od skupiny aj odpojiť alebo sa prepnúť do inej – tá sa stane jeho aktívnou. Na pripojenie do skupiny, ktorá má heslo, je nevyhnutné ho správne zadať.

### **3.12 Anotácie**

Prihlásený používateľ bude môcť dokument svojej aktívnej skupiny anotovať, ak má na to právo. Anotácie sa môžu aj prekrývať.

### **3.13 Referencie**

Používateľ bude môcť špecifikovať, či je daná anotácia referenciou a na ktorú predošlú anotáciu sa vzťahuje (rodič). Referencia dedí od rodiča všetky jeho vlastnosti, pri zmene rodiča je patričná zmena vykonaná na všetkých jeho referenciách.

### **3.14 Typy anotácií**

Používateľ s editačnými právami bude môcť vytvárať nové typy a podtypy anotácií, alebo ich iným spôsobom upravovať (farba, tag, názov,...) či mazať.

### **3.15 Anotovanie**

Vyznačený text v dokumente môže používateľ anotovať stlačením kláves, kde si bude môcť vybrať napr.: typ anotácie (napr. “mesto”), popis, odkazy, .... Anotácie môžu byť zmenené/zmazané používateľmi na základe ich práv.

### **3.16 Zobrazovanie anotácií**

Anotácie v textovom editore budú farebne vyznačené podľa daného typu/kategórie. Farbu rôznych kategórií si môže užívateľ zvoliť. Prekryv anotácií bude nejakým spôsobom rozlíšený (napr. konkrétnou farbou alebo farba prekryvu bude určená zmiešaním farieb jednotlivých prekrývajúcich sa anotácií).

### **3.17 Zobrazovanie zmien v dokumente**

Zmeny, vykonané jedným používateľom sa okamžite zobrazia aj ostatným používateľom v danej skupine.

### **3.18 Viditeľnosť anotácií**

Používateľ si bude môcť určiť, ktoré typy anotácií sa mu budú podfarbovať v texte (napr. sa rozhodne, že anotácie typu “osoba” nebudú momentálne vyznačené). Táto zmena je len lokálna a dočasná – nezobrazuje sa ostatným používateľom v skupine.

### **3.19 Komentáre**

Používatelia budú mať možnosť pridať do textu komentáre.

### 3.20 Editovanie/mazanie komentárov

Komentáre môžu byť priamo upravené/zmazané len ich autorom.

### 3.21 Zobrazovanie komentárov

Po nadídení myšou/stlačením kláves nad okomentovaným textom sa komentár zobrazí používateľovi. Okomentovaný text bude v dokumente nejakým spôsobom vizuálne rozlíšený.

### 3.22 Hľadanie v texte

Aplikácia bude umožňovať používateľom hľadať slová v texte.

### 3.23 Export

Používateľ bude môcť exportovať dokument svojej aktívnej skupiny. Každý riadok vyexportovaného dokumentu bude mať nasledovný formát:

- číslo riadku/poradie    slovo    X:Z,...    A-B:...    odkazy

Kde **X** bude buď písmeno 'B' (begin – označuje začiatok anotácie) alebo 'I' (pokračovanie anotácie) prípadne nič, **Z** bude zoznam tagov daného typu anotácie (napr. PER pre persons, tagy si môže meniť používateľ s editačnými právami, viaceré tagy budú oddelené čiarkou). **A:B** je interval <A, B), ktorý určuje riadok slova alebo slov, na ktoré sa referencia vzťahuje, viacero intervalov je oddelených čiarkou. Pri anotáciách, ktoré nie sú referencie a pri neanotovanom texte je A aj B prázdne. Viaceré intervaly budú oddelené čiarkou. Nasledovať bude zoznam odkazov, pre danú anotáciu (pri viacslovných anotáciách je to vždy v riadku prvého anotovaného slova). Interpunkcia je takisto na samostatnom riadku.

Nasledujú príklady riadkov exportovaného dokumentu:

- 2      Bratislava    B:CIT    :    en.wikipedia.org/wiki/Bratislava

Viacslovný anotovaný text:

- 3      Barack      B:PER    :    en.wikipedia.org/wiki/Barack\_Obama
- 5      Obama      I            :

Referencia:

- 7      He      B:PER    1-2

Neanotované slovo:

- 11    foo    :            :

Viackrát anotované jedno slovo:

- 13    bar    B:NAME1,NAME2    :



A nakoniec referencia odkazujúca na entitu, ktorá je na viacerých miestach:

- 14 It B:CIT 123-124:2-3

### **3.24 Aktivita v dokumente**

Aktivita v skupine (anotovanie textu a rôzne iné zmeny) sa všetkým používateľom budú zobrazovať pomocou logu, kde budú všetky dôležité informácie ako čas zmeny, užívateľ, ktorý danú zmenu vykonal, samotná zmena – komentár, zmazanie, anotácia, ...).