

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

1 Úvod.....	3
1.1 Ciele dokumentu.....	3
1.2 Rozsah projektu.....	3
1.3 Definície, skratky.....	3
1.4 Odkazy.....	4
1.5 Nasledujúce kapitoly.....	4
2 Všeobecný popis.....	4
2.1 Perspektíva projektu.....	4
2.2 Funkcie projektu.....	4
2.3 Charakteristika používateľov.....	4
2.4 Všeobecné obmedzenia.....	5
3 Špecifické požiadavky.....	5
3.1 Pripojenie.....	5
3.1.1 Konfigurácia pripojenia.....	5
3.1.2 Konfiguračný súbor.....	5
3.2 Registrácia a prihlasovanie.....	5
3.2.1 Registrácia.....	5
3.2.2 Prihlasovanie a odhlasovanie.....	5
3.2.3 Automatické prihlasovanie.....	5
3.3 Skupiny.....	6
3.3.1 Tvorba skupiny.....	6
3.3.2 Aktívna skupina.....	6
3.3.3 Voľba skupiny.....	6
3.3.4 Prístup do skupiny.....	6
3.3.5 Default práva skupiny.....	6
3.3.6 Zrušenie skupiny.....	6
3.3.7 Aktivita v skupine.....	6
3.3.8 Čistenie logu.....	7
3.3.9 Export logu.....	7
3.4 Práva používateľa.....	7
3.4.1 Nastavovanie práv.....	7
3.4.2 Typy práv.....	7
3.5 Administrácia skupiny.....	7
3.5.1 Administrátor.....	7
3.5.2 Práva administrátora skupiny.....	7
3.6 Anotovaný dokument.....	8
3.6.1 Zmena obsahu anotovaného dokumentu.....	8
3.6.2 Zobrazovanie zmien v dokumente.....	8

3.6.3 Vzťah dokument-skupina.....	8
3.6.4 Rozdelenie dokumentu.....	8
3.7 Anotácie.....	8
3.7.1 Pridávanie/mazanie/úprava anotácií.....	8
3.7.2 Prekryv anotácií.....	8
3.7.3 Referencie.....	8
3.7.4 Zmena referencií.....	8
3.7.5 Typy anotácií.....	9
3.7.6 Vyznačenie anotácií a referencií v texte.....	9
3.7.7 Zobrazovanie anotácií.....	9
3.8 Komentáre.....	9
3.8.1 Pridávanie/editovanie/mazanie komentárov.....	9
3.8.2 Zobrazovanie komentárov.....	9
3.9 Hľadanie v texte.....	9
3.10 Export.....	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

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 Pripojenie

3.1.1 Konfigurácia pripojenia

Používateľ si po naštartovaní aplikácie bude môcť konfigurovať pripojenie na server, na ktorý sa chce pripájať.

3.1.2 Konfiguračný súbor

Informácie o pripojení sa ukladajú do konfiguračného súboru. Pri opätovnom spustení aplikácie sa používateľ pripája na server uložený v tomto súbore.

3.2 Registrácia a prihlasovanie

3.2.1 Registrácia

Používateľ sa bude môcť zaregistrovať na vybraný server. Pri registrácii bude potrebné uviesť používateľské meno a heslo.

3.2.2 Prihlasovanie a odhlasovanie

Po registrácii sa používateľ bude môcť, za pomoci prihlasovacieho mena a hesla, pripojiť na daný server. Prihlásený používateľ sa bude môcť zo servera aj odhlásiť.

3.2.3 Automatické prihlasovanie

Používateľ si bude môcť určiť, či sa hneď po spustení aplikácie bude chcieť automaticky pripojiť na server na základe informácií z konfiguračného súboru.

3.3 Skupiny

3.3.1 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)

3.3.2 Aktívna skupina

Skupina sa stáva aktívnou, ak ju používateľ vytvoril, alebo sa do nej pripojil. Aktívna skupina je maximálne 1.

3.3.3 Voľba skupiny

Prihlásený používateľ si môže zvoliť svoju aktívnu skupinu zo zoznamu všetkých skupín. Používateľ má možnosť sa od skupiny aj odpojiť alebo sa pripojiť do inej.

3.3.4 Prístup do skupiny

Prístup do skupiny môže byť chránený heslom. Ak skupina má nastavené heslo, na pripojenie do nej je nevyhnutné ho správne zadať.

3.3.5 Default práva skupiny

Skupina môže mať nastavené defaultné práva. Ak sú nastavené, tak sa užívateľom, ktorí sa do nej pripojili, nastavlia. Ak nastavené nie sú, do skupiny smú vstúpiť len používatelia v jej zozname.

3.3.6 Zrušenie skupiny

Používateľ bude kedykoľvek môcť zmazať skupinu, ktorú sám vytvoril. Pri zrušení skupiny je jej dokument taktiež zmazaný.

3.3.7 Aktivita v skupine

Aktivita (anotovanie textu a rôzne iné zmeny) sa bude zobrazovať všetkým používateľom v skupine pomocou logu, kde budú dôležité informácie, napr.

čas zmeny, užívateľ, ktorý danú zmenu vykonal, samotná zmena – komentár, zmazanie, anotácia,

3.3.8 Čistenie logu

Log bude môcť používateľ aj zmazať. Log sa vždy viaže na aktívnu skupinu a pri jej zmene sa sám zmaže.

3.3.9 Export logu

Používateľ bude mať možnosť aktuálny log aj exportovať.

3.4 Práva používateľa

3.4.1 Nastavovanie práv

Práva používateľa sa nastavujú hneď po úspešnom pripojení do skupiny (buď aké má v zozname skupiny alebo jej defaultné).

3.4.2 Typy práv

Rozlišujeme medzi 3 základnými typmi právami (možno ich 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

3.5 Administrácia skupiny

3.5.1 Administrátor

Používateľ, ktorý vytvára skupinu sa stáva jej administrátorom. Každá skupina má práve jedného administrátora a ten je nemenný.

3.5.2 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.

3.6 Anotovaný dokument

3.6.1 Zmena obsahu anotovaného dokumentu

Obsah dokumentu nemožno upravovať iným spôsobom ako anotovaním a komentovaním.

3.6.2 Zobrazovanie zmien v dokumente

Zmeny vykonané používateľom sa okamžite zobrazujú všetkým ostatným používateľom, ktorí upravujú daný dokument.

3.6.3 Vzťah dokument-skupina

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

3.6.4 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.

3.7 Anotácie

3.7.1 Pridávanie/mazanie/úprava anotácií

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.

3.7.2 Prekryv anotácií

Ľubovoľnú časť textu môže byť anotovaná aj viackrát.

3.7.3 Referencie

Používateľ bude môcť špecifikovať, či je daná anotácia referencia a na ktorú predošlú anotáciu sa vzťahuje.

3.7.4 Zmena referencií

Referenciám možno iba zmeniť objekt, na ktorý referujú (rodič), prípadne ich vymazať. Referencia dedí od rodiča všetky jeho vlastnosti, pri zmene rodiča je patričná zmena vykonaná na všetkých jeho referenciách.

3.7.5 Typy anotácií

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

3.7.6 Vyznačenie anotácií a referencií v texte

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 zmiešaním farieb jednotlivých prekrývajúcich sa anotácií). Referencie budú mať vždy fixnú farbu.

3.7.7 Zobrazovanie anotácií

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

3.8 Komentáre

3.8.1 Pridávanie/editovanie/mazanie komentárov

Každý používateľ v skupine bude mať možnosť pridať do textu komentáre. Komentáre môžu byť priamo upravené/zmazané len ich autorom.

3.8.2 Zobrazovanie komentárov

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

3.9 Hľadanie v texte

Aplikácia bude umožňovať používateľom hľadať slová v texte.

3.10 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, referencie budú mať nemenný tag). **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.

Samotná dvojbodka v 3. stĺpci označuje, že slovo nie je anotované a vo 4. stĺpci, že slovo nie je referencia (kvôli lepšiemu parsovaniu).

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:REF 1-2

Neanotované slovo:

- 11 foo : :

Viacrát anotované jedno slovo:

- 13 bar B:NAME1,NAME2 :

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

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