

Pojednání k tématu disertační práce

Webové služby a komponentní technologie

2010 / 2011

Ing. Jaroslav Dytrych, xdytry00@stud.fit.vutbr.cz
Fakulta informačních technologií
Vysoké učení technické v Brně

Obsah

1 Úvod	3
2 Stav poznání v oblasti tématu práce	3
2.1 Anotace	3
2.2 Použití tagování pro spolupráci	4
2.3 Kolaborativní prostředí pro tagování	4
2.4 Strukturování znalostí	4
2.5 Způsoby vytváření anotací	4
2.6 Prostředí pro spolupráci v reálném čase	5
2.7 Existující řešení pro anotování	5
3 Cíle práce	8
3.1 Nové způsoby anotování	8
3.2 Anotování kdekoliv a kdykoliv	9
3.3 Nové možnosti využití anotací	9
4 Zvolené metody řešení	9
4.1 Formát anotace	10
4.2 Protokol pro přenos anotací	10
4.3 Server pro práci s anotacemi	11
4.4 Klienti	11
5 Shrnutí dosavadních výsledků řešení	13
5.1 Návrh formátu anotace	13
5.2 Návrh protokolu pro přenos anotací mezi klientem a serverem	15
5.3 Přenos anotací mezi servery	15
5.4 Server	16
5.5 Klienti	16
5.6 Využití anotací pro workflow	26
6 Budoucí vývoj	27
6.1 Technický vývoj	27
6.2 Plánovaná publikační činnost	27
6.3 Plánovaná zahraniční spolupráce	27
7 Závěr	27
Literatura	33
Přílohy	34

1 Úvod

V této práci se zabývám kolaborativním anotováním v různých prostředích a rozšířením možností využití anotací.

Ve 2. kapitole vysvětlím základní pojmy a stav poznání v oblasti tématu práce a souvisejících. Ve 3. kapitole jsou uvedeny cíle méj disertační práce, které zahrnují především nové způsoby anotování kdekoli a kdykoli a nové možnosti využití anotací. Zvolené metody řešení jsou následně uvedeny ve 4. kapitole, kde se zabývám nutností vytvoření nového formátu anotace, protokolů pro přenos anotací a požadavky na navržená řešení. Jsou zde uvedeny i informace o klientech a serveru pro práci s anotacemi. Následně jsem v 5. kapitole navrhl nový formát anotace a protokol pro komunikaci mezi klientem a serverem. Navrhl jsem také jednoduchý server pro práci s anotacemi a vzhled a chování klientů, kteří budou tvořeni editory anotací. V závěru kapitoly se potom věnuji návrhu formátu anotace pro workflow a problematice podpory workflow na serveru a v klientech.

V 6. kapitole popisují plánovaný vývoj své disertační práce a to nejenom z technického hlediska, ale také z hlediska spolupráce se zahraničními projekty. Uvádím zde také informace o plánované publikační činnosti. Závěrečná kapitola shrnuje aktuální stav práce i budoucí vývoj.

2 Stav poznání v oblasti tématu práce

Disertační práce je zaměřena na anotování v různých prostředích. V této kapitole tedy uvedu, co je to anotace, a následně se zaměřím na jeden z typů anotování, kterým je značkování (tagging). Potom popíši aktuální využití značkování ve sdílených prostředích a pojmy, které jsou s touto problematikou úzce spjaty.

Protože se budu zabývat i problematikou kolaborativního anotování v reálném čase, zaměřím se i na prostředí pro tento druh spolupráce, jejich výhody, nevýhody a související problémy.

2.1 Anotace

Historicky je slovo anotace definováno jednak jako akt samotného anotování a jednak jako poznámka přidaná jako vysvětlení k nějaké literární práci. Více o historii slova anotace lze nalézt v [2]. Dále budu slovem anotace označovat doplňující informaci přidanou k textu bez omezení významu této informace či charakteru anotované části textu.

Anotace se nejčastěji vyskytují ve formě jednoduchých poznámek přidávaných do textu. Tyto byly připisovány a vkládány do papírových dokumentů už před stovkami let. V souvislosti s přechodem na digitální formu přišla i elektronická forma anotace, tzv. digitální anotace, což je dle definice objekt související s obsahem jednoho nebo více jiných objektů [8]. Pro elektronické dokumenty dnes existuje velké množství aplikací umožňujících vložení poznámky, která se zobrazí např. ve formě zvýraznění textu a bublinové nápovědy při najetí myši na daný text.

Anotace umožňují uživatelům přirozeně spojovat osobní obsah s informačními zdroji. Mají široké spektrum použití od obohacování informačních zdrojů o vysvětlující informace až po přenášení a sdílení myšlenek a znalostí o anotovaném subjektu [7]. Mohou tvořit obsahovou vrstvu věnovanou vyjasňování významu informačního zdroje na nižší vrstvě [3]. Nejedná se tedy pouze o komentáře, anotace lze považovat i za hodnotnou autonomní intelektuální práci [2].

Pomocí anotací lze také doplnit strukturované informace k nestrukturovanému textu. Díky tomu začaly být anotace využívány k mnoha dalším účelům. Mezi tyto patří např. propojení dokumentů, doplnění sémantiky k vybrané části textu [4], [35] apod. Doplněná informace může být jak formálního, tak i neformálního charakteru [14]. Anotace lze potom rozdělit na [14]:

- textové – obsahují textový komentář,
- odkazové – odkazují na jiný dokument (informace je ve formě cíle odkazu),
- sémantické – přiřazuje elementy ze specifikovaného modelu s hodnotami z kontrolovaného slovníku.

V [14] je následně uvedeno i jemnější členění anotací dle různých kritérií, která často závisejí i na jejich konkrétním využití. Různé dimenze anotací jsou i náplní dalších prací (viz např. [29]). Problematika anotací je rozsáhlá a pro popis všech jejich aspektů není v tomto textu prostor. Nyní tedy přes konkrétní příklad využití anotací přejdu k další důležité oblasti týkající se této práce.

Formální sémantické anotace se využívají v sémantických wiki, kde obsahují vztahy s formálními definicemi pojmů ve znalostních databázích (kontrolovaný slovník). Doplněním anotace k určitému termínu v textu tak můžeme tento termín jednoznačně formálně popsat.

Dalším důležitým pojmem je tzv. tag (značka či štítek), což je krátká, obvykle jednoslovná, textová anotace. Tagging (tagování či značkování) je tedy specifický druh anotování, který spočívá v doplňování krátkých anotací k textu a je považován za nejjednodušší způsob přidávání metadat k fragmentům textu. Více lze nalézt v [57].

2.2 Použití tagování pro spolupráci

Pokud je tagging provozován kolaborativně, jedná se o tzv. social tagging. Tagy přiřazují různí uživatelé podle svého osobního názoru. Tagy mohou označovat nejenom druh či vlastnost termínu, ale také osobní preference či názory tagujících uživatelů. Některé anotace potom spíše než informace o anotovaném obsahu nesou informace o uživateli, jeho zájmech a preferencích, osobních názorech a dojmech [42]. Tagy lze tedy chápat jako propojující elementy mezi zdroji a uživateli, resp. mezi zdroji a pojmy v mysli uživatele [52].

Uživatelé mohou označit stejnou věc vzájemně se vylučujícími termíny. Pokud je jeden termín označen více protichůdnými tagy, můžeme podle počtu jednotlivých tagů určit, který tag bude zachován.

Tagy se využívají pro kategorizaci obsahu, navigaci a vyhledávání. Jako jeden z prvních je začal využívat Flickr [19], ve kterém slouží ke kategorizaci fotografií. Dalším příkladem využití je Twitter [54], kde ve formě hashtags (klíčová slova vyznačená křížkem) slouží ke kategorizaci zpráv (Tweets). Tagy se využívají i v blozích a dalších systémech, kde je potřeba kategorizovat obsah. Dle [70] jsou anotace jedním ze základních prvků sémantického webu.

Z množství tagů lze určit popularitu tagovaného pojmu. Ta je následně využitelná k různým účelům od nabízení při vytváření dalších tagů až po vyhledávání dokumentů, které se daným pojmem zabývají.

Vzhledem k tomu, že tag je krátká textová anotace, může být interpretován různými způsoby (slovo či fráze má více významů). Proto jsou obvykle sdíleny ve skupinách uživatelů se stejnými zájmy, přičemž každá skupina si vytváří určitý slovník tagů, jejichž významy jsou dány zvyklostmi dané skupiny.

2.3 Kolaborativní prostředí pro tagování

Jak známo, wiki je web, který umožňuje uživatelům přidávat a měnit obsah. Wiki stránky jsou vytvářeny kolektivně pomocí jednoduchého značkovacího jazyka [64].

Sémantická wiki je wiki rozšířená o technologie sémantického webu [56]. Kromě samotných wiki stránek tedy obsahuje i model znalostí, které jsou na nich popsány. Umožňuje tak formálně zachytit identitu dat a vztahy mezi nimi. Mezi nejznámější software pro sémantické wiki patří Semantic MediaWiki [49], což je rozšíření MediaWiki [31], které umožňuje ukládat strukturovaná data do wiki stránek (anotovat wiki stránky). Více lze nalézt v [43], [62] a [56].

Pojem sémantická wiki má dvojí chápání. Lze jej chápat nejenom jako wiki rozšířenou o technologie sémantického webu, ale i jako software pro vytváření sémantických dat způsobem běžným pro wiki (angl. wiki way [16]) [45], [44].

2.4 Strukturování znalostí

Z termínů folks (lidé) a taxonomie (klasifikace) vznikl termín folksonomie [55]. Dle [70] je sociální anotace formou folksonomie, která označuje internetovou metodu pro kolaborativní generování nelimitovaných textových popisů, které kategorizují obsah webu.

Pokud lidé označí nějaký pojem tagem, lze tento pojem chápat jako entitu náležící do množiny entit označených daným tagem (obdoba klasifikace). Tagy lze potom chápat jako označení typů entit.

Prostředkem pro sdílení konceptualizace domény jsou ontologie. Jedná se o sémantické modely popisující tématickou část světa [46]. Ontologie definují vztahy mezi entitami v dané doméně.

Pokud pomocí tagů definujeme entity a pomocí ontologií vztahy mezi nimi, vytvoříme model domény. Pokud je některá entita označena novým (dosud neznámým) tagem a tento tag začnou lidé využívat, vytvoří novou množinu (typ) entit. Entity označené tímto tagem takto získají novou sémantiku, která je označována jako objevující se sémantika (emerging semantics [12]).

2.5 Způsoby vytváření anotací

Anotace lze vytvářet:

- manuálně,
- poloautomaticky,
- automaticky.

Manuální tvorba anotací je prováděna lidmi, přičemž jedním z přístupů k této činnosti je social tagging. Získané informace potom mohou být např. ve formě folksonomií či kolaborativně vytvořených ontologií a nově objevené sémantiky (emerging semantics).

Při poloautomatickém vytváření systém uživatelé nabízí anotace či jejich části a uživatel je pouze schvaluje. Nabídky mohou být nejčastěji využívané tagy, ale také komplexnější anotace. K získání takových anotací existují různé systémy

pro extrakci informací z nestrukturovaného textu. Jednou z technik, které se v těchto systémech využívají, je Screen scraping.

Screen scraping je podle Wikipedie [63] technika, při které počítačový program extrahuje data ze zobrazovaného výstupu jiného programu. Hlavní rozdíl mezi screen scrapingem a běžným parsováním je, že výstup, který je scrapován, byl určen pro konečné zobrazení uživateli (člověku), spíše než ke vstupu do jiného programu, a proto obvykle není zdokumentovaný, ani strukturovaný pro pohodlné parsování [17].

Web scraping je speciální variantou screen scrapingu, která se zabývá extrakcí dat z webových stránek. Stránky jsou sice popsány ve značkovacím jazyce, ale ten zde obvykle spíše než strukturu dat popisuje způsob zobrazení koncovému uživateli, pro kterého je stránka určena. Dále se budu zabývat především web scrapingem a pro Scaper tedy využiji definici z [48]: „Scaper je počítačový program, který kopíruje strukturované informace z webových stránek do databáze.“

Automatické vytváření anotací lze využít např. tam, kde je třeba automaticky doplňovat databázi znalostí vytvářenou lidmi. Anotace jsou potom vytvářeny stejně, jako u poloautomatického vytváření, ale nejsou ihned kontrolovány a potvrzovány lidmi.

2.6 Prostředí pro spolupráci v reálném čase

Protože se ve své práci budu zabývat i kolaborativním anotováním v reálném čase, které umožní rychlejší a efektivnější vytváření anotací, je třeba zde zmínit i specifika tohoto druhu spolupráce.

Spolupráce v reálném čase znamená, že spolupracující uživatelé pracují paralelně. Je třeba řešit souběžný přístup, při kterém jsou změny prováděny jedním uživatelem ihned promítnuty do stavu aplikace druhého uživatele, resp. všech ostatních spolupracujících uživatelů.

Pokud by se každá změna projevila až poté, co se rozšíří mezi všechny spolupracující uživatele, byla by aplikace uživatelsky nepřívětivá (např. zobrazení znaku několik sekund po stisknutí klávesy). Je tedy nutné zvolit určitý kompromis a některé operace provést u daného uživatele a poté zaslat informaci ostatním. V tomto případě musejí být ošetřeny konflikty, kdy dva uživatelé provádějí vzájemně se vylučující operace (např. jeden uživatel edituje text, zatímco druhý jej maže).

V prostředí webu přenos stavových informací v reálném čase vyžaduje využití technologií, které umožní asynchronní přenos informací mezi klientem a serverem. Pro asynchronní komunikaci je velmi populárním přístupem AJAX (Asynchronous JavaScript + XML, více lze nalézt v [20]), ten však umožňuje pouze asynchronní komunikaci ze strany klienta a nikoliv ze strany serveru. Server tedy vždy musí čekat na požadavek od klienta. Aby bylo možné ihned přenést aktuální informace ke klientovi, je třeba využít Comet [41], což je přístup umožňující obousměrně asynchronní komunikaci webového prohlížeče se serverem.

V případě distribuovaného prostředí, které zahrnuje více serverů, je třeba řešit i komunikaci mezi servery. Vzhledem k tomu, že komunikace mezi servery není tak výrazně omezena, jako komunikace mezi serverem a klientem, lze využít efektivnější transportní protokol. Mezi vhodné protokoly patří např. XMPP (Extensible Messaging and Presence Protocol [67]), který byl speciálně navržen pro přenos informací o přítomnosti uživatelů (presence) a zpráv (instant messaging) a pro komunikaci a spolupráci v reálném čase.

Příkladem platformy pro spolupráci v reálném čase je Google Wave [21], který je jednou z prvních webových platforem, u kterých jsou změny přenášeny s velmi nízkou granularitou (napsání jednoho písmene či provedení libovolné akce). Jedná se o distribuovanou platformu, kde každý klient je přihlášen ke svému domovskému serveru a pro komunikaci mezi servery se využívá Google Wave Protokol [22].

Dalšími příklady systémů pro spolupráci v reálném čase jsou Novell Pulse (kompatibilní s Google Wave), kancelářský balík Google Docs a Microsoft Office Live.

2.7 Existující řešení pro anotování

Pro anotování v prostředí webu existuje celá řada nástrojů, přičemž první z nich vznikly před více než 15 lety (např. [26] a [47]). Pro ukázkou jsem zvolil několik nástrojů, u kterých jsou patrné obvyklé přístupy, možnosti anotování a sdílení anotací. Nástroje jsou většinou uzavřené a k anotacím nelze volně přistupovat a strojově je zpracovávat. Anotace nelze přenášet do jiných prostředí a dále je využívat. Kombinace obecného anotování textovou poznámkou a tagování, u kterého jsou přiřazovány krátké značky, rovněž není samozřejmostí, i když je podporována u celé řady nástrojů. Podpora pro strukturování anotací (možnost využití anotace v jiné anotaci) je u existujících nástrojů výjimkou, a pokud je dostupná, nástroje mají jiná výrazná omezení.

Existující nástroje lze rozdělit do tří kategorií [40]:

- serverové,
- proxy,
- rozšíření (na klientovi).

U serverových systémů je vše prováděno na serveru, ke kterému uživatel přistupuje přes webové uživatelské rozhraní. Není tedy třeba instalovat žádný anotační program, ale lze anotovat pouze dokumenty umístěné na specifickém serveru. Obvykle se potom jedná spíše o vestavěnou anotační funkci nějakého systému, spíše než o samostatný anotační systém.

U proxy řešení je využit server, který načte obsah z jiného zdroje (obvykle jiného serveru) a ten následně umožní anotovat. Veškerá anotační funkcionalita je zde opět na serveru (není nutné nic instalovat). Nevýhodou však je, že server musí mít přístup k danému zdroji (klient přes něj např. musí zaslat přihlašovací údaje). Anotací funkcionalita je navíc kompletně zabudována přímo do načtené webové stránky, což může zavádět např. problémy při tisku dokumentu.

Pokud je využito rozšíření, uživatel si musí nainstalovat součást anotačního systému do nějakého svého programu (obvykle webového prohlížeče). Práce s anotacemi pak může probíhat odděleně od práce s dokumentem (anotační systém nemusí měnit obsah načítaného dokumentu) a lze anotovat cokoli, k čemu má uživatel přístup. Nevýhodou je zde nutnost instalace.

Následují příklady existujících anotačních nástrojů. V závěru kapitoly se potom budu zabývat přenosem anotací mezi klientem a serverem.

Annozilla

Annozilla [33] je rozšíření webového prohlížeče Mozilla Firefox, které umožňuje anotovat obsah webových stránek. Anotace jsou ukládány na lokální či vzdálený server a pro identifikaci anotovaného místa v dokumentu je zde využit XPointer. Jedná se o jednoduché textové anotace, které mají definovaný typ a jazyk. Typ anotace je vybírán z několika předdefinovaných možností, mezi které patří např. komentář, vysvětlení, otázka, souhlas / nesouhlas apod. Přítomnost anotací v textu je zobrazována formou ikonky a samotné anotace jsou zobrazovány v postranním panelu. Pro přenos anotací na server se využívá protokol vyvinutý v rámci projektu Annotea (viz níže).

Výhodou je jednoduchost a uživatelská přívětivost řešení. Mezi nevýhody patří především nemožnost tvorby strukturovaných anotací, omezená množina typů, které nepřirazují sémantiku anotovanému textu (nelze je využít jako tagy) a omezení na webové stránky (statický charakter dokumentu).

Amaya

Editor webových stránek Amaya [39] od W3C má v sobě zabudovanou anotační funkcionalitu. Jedná se o součást projektu Annotea [24]. Anotace jsou externí (mimo dokumenty) a mohou být uloženy lokálně (v počítači uživatele) nebo na jednom či více anotačních serverech. Lokálně uložené anotace jsou privátní, vzdáleně uložené jsou veřejné.

Vzhledem k tomu, že editor je stále ve vývoji, je zde řada omezení a nevýhod. Filtr pro anotace určitého typu, od určitého uživatele nebo z určitého serveru je implementován v klientovi (načíst se musí vše). Vlastnosti anotací (včetně typů anotací) musejí být popsány v RDFS, což vylučuje dynamické strukturování anotací. Pracovat lze pouze s dobře strukturovanými dokumenty (XML či XHTML). Více informací lze nalézt v [24] a [25].

SharedCopy

SharedCopy [50] umožňuje vytvořit kopii libovolné webové stránky a tuto kopii anotovat. Instalace aplikace spočívá v jednoduchém vytvoření záložky se speciálním odkazem, který obsahuje JavaScriptový kód, který vloží potřebné části aplikace do aktuálně zobrazené stránky. Anotace jsou textové komentáře a jednoduché grafické prvky a jsou ukládány společně s kopiemi stránek na serveru. Sdílení anotací je realizováno pomocí rozesílání vygenerované URL. Anotovaný text je zobrazen zvýrazněný a samotný textový obsah anotace se zobrazuje v panelu aplikace, ve kterém mohou spolupracovníci na anotace reagovat. Grafické prvky jsou zobrazovány přímo v anotované kopii stránky. Pokud se původní stránka změní, systém poskytuje možnost porovnání anotované kopie s aktuální verzí originálu.

Výhodou SharedCopy je jednoduchost použití bez nutnosti registrace a instalace SW. Vizualizace je rovněž dobře vyřešena. Nevýhodou je uzavřené řešení pro přenos a ukládání anotací a složitější sdílení. Anotace také nejsou využitelné mimo danou stránku.

ShiftSpace

ShiftSpace [51] je rozšíření do webového prohlížeče Mozilla Firefox, které umožňuje anotovat obsah webových stránek textovými poznámkami. Anotace jsou ukládány na server společně s kopiemi anotovaných stránek, přičemž kopie stránky může být modifikována. Modifikaci stránky lze chápat jako součást anotací na této stránce. Sdílení je realizováno obdobně, jako u SharedCopy.

ShiftSpace jsem zde uvedl pro úplnost, protože je často uváděn jako jeden z nejznámějších nástrojů k anotování. Tento nástroj se však nejvíce vzdaluje od zaměření této práce, protože kromě anotování umožňuje i modifikace anotovaného textu. Modifikace textu, které jsou využity k doplnění informací v anotacích, potom narušují koncept anotování, kdy anotace není pouze přidaná informace, ale stává se obtížně odlišitelnou součástí originálního textu. Anotace potom nejsou využitelné mimo anotovanou kopii dokumentu, tedy ani v originálním textu.

Stickis

Stickis [53] je zásuvný modul pro webové prohlížeče Mozilla Firefox a Microsoft Internet Explorer. Umožňuje vytváření anotací s formátovaným textem, ve kterém mohou být i obrázky a videa. Anotace mohou obsahovat i tagy, takže lze stránku současně tagovat. Anotuje se celá stránka či více stránek současně (např. anotování vztahu mezi stránkami). Anotace jsou ukládány na serveru a sdíleny formou kanálů, ke kterým se uživatelé přihlašují. Na anotace lze i reagovat, přičemž je upozorněn autor původní anotace. Stickis podporuje také řízení přístupu k anotacím a dělí je na veřejné, s omezeným přístupem a privátní.

Přehled anotací ke stránce je zobrazován v postranním panelu a samotné anotace v oknech nad stránkou. V postranním panelu mohou být zobrazeny pouze nové anotace a přečtené anotace, které již uživatele nezajímají, mohou být trvale skryty.

Výhodami jsou možnost anotování více stránek současně, možnost tagování a sdílení formou kanálů. Nevýhodami jsou nemožnost anotování konkrétního fragmentu textu, uzavřený formát ukládání anotací a omezení na konkrétní webové prohlížeče.

Crocodoc

Crocodoc [15] je webová aplikace, která umožňuje anotovat nejenom webové stránky, ale také obrázky, dokumenty Microsoft Word a PowerPoint a dokumenty ve formátu PDF. U webových stránek Crocodoc nejprve vytvoří obrázek (snapshot) a ten je následně anotován. Anotace jsou tvořeny textovými poznámkami, ručně kreslenými obrázky a zvýrazňováním textu. Vše je společně s anotovaným dokumentem ukládáno na serveru v uzavřeném formátu. Crocodoc umožňuje i sdílení anotovaných dokumentů.

Výhodou je možnost anotovat dokumenty v různých formátech. Nevýhodami jsou uzavřenost řešení a nemožnost tagování. Dokumenty s anotacemi nelze převádět mezi jednotlivými formáty.

MyStickies

MyStickies [66] je rozšíření webového prohlížeče, které umožňuje vkládání jednoduchých anotací do webových stránek. Anotace obsahují text a tagy a jsou vizualizovány ve formě žlutých samolepících papírků přilepených ke stránce.

Výhodami MyStickies jsou jednoduchost použití a možnost jednoduchého anotování a tagování současně. Velmi výraznou nevýhodou tohoto systému je však nemožnost sdílení anotací. Tato funkcionality je prozatím na seznamu budoucích rozšíření. Anotace rovněž nelze strukturovat.

Bundle Editor

Bundle Editor [69] nepatří mezi webové aplikace, ale jedná se o program pro demonstraci využití strukturovaných anotací při kooperativní tvorbě dokumentů. Umožňuje strukturování anotací jejich seskupováním do balíků. Anotace je možné filtrovat, řadit apod. Program je určen pro anotování textových dokumentů, do kterých jsou ukládány i anotace.

Program není veřejně dostupný a určený k širšímu využití. Využitá koncepce strukturování anotací se však ukázala jako efektivnější a uživatelsky přívětivější než využití prostých nestrukturovaných anotací [68].

Další editory

Existuje celá řada dalších řešení, která jsou určena převážně pro digitální knihovny (např. DiLAS [1]) nebo pro výzkumné účely. Příklady několika systémů lze nalézt v [1], [5] a [40].

Za samostatnou zmínku stojí PREP Editor [34], který je jedním ze starších editorů s anotační funkcionalitou podporující spolupráci v reálném čase. Je určen pro kolaborativní psaní dokumentů, z čehož vyplývá i omezení anotační funkcionality na potřeby specifické pro tuto činnost. Při spolupráci umožňuje manuální či automatické zaslání změn v dokumentu, přičemž u automatického je možné nastavení granularity výběrem ze tří možností: sloupec (column), kus textu (chunk) nebo stisk klávesy (keystroke). V jiných aspektech je však toto řešení z roku 1994 zastaralé.

Různé systémy podporují různou inovativní funkcionalitu, přičemž některé funkce jsou velkým přínosem (lepší či rychlejší spolupráce, lépe strojově zpracovatelné anotace, uživatelská přívětivost apod.). Většina systémů je však vytvářena se zaměřením na řešení určitého problému, což vede nejenom k omezení využitelnosti takového systému, ale i k častému zanedbání některého z klíčových požadavků na anotační systém [36]. Jako konkrétní příklad takového systému lze uvést FAST (Flexible Annotation Service Tool) [6]. Jedná se o systém s univerzálním využitím nezávislý na platformě s mnoha výhodami [5]. Vzhledem k tomu, že jádro je zde zcela izolované od konkrétních systémů pro správu obsahu, aby bylo univerzálně využitelné, nemá žádné informace o konkrétním obsahu, ale pouze o anotacích. Není zde tedy žádná informace o obsahu anotované verze dokumentu, ale pouze odkaz na umístění v daném dokumentu. V případě potřeby je obsah dokumentu načten přes modul, který však potřebuje spolupráci se systémem pro správu obsahu, tedy s původním zdrojem informací. Pokud potom chceme anotovat webové stránky, musí mít anotační proxy přístup k danému webserveru, což je značné omezení.

Existující řešení pro přenos anotací

Jak jsem uvedl výše, většina existujících nástrojů pro anotování je uzavřených a formáty a protokoly pro přenos anotací mezi klientem a serverem jsou nedostupné. Norma pro ukládání anotací je aktuálně ve vývoji (projekt Annotea organizace W3C [59]) a využívá ji pouze malé množství projektů (např. Annotzilla). Mezi její nevýhody patří omezení na XML a XHTML dokumenty [24] a přenos anotací pomocí HTTP (komplikace pro přenos v reálném čase a server musí podporovat zprávy PUT, DELETE apod.).

Pro přenos anotací před určitou dobou firma IBM vyvinula API [18] a implementovala jej v projektu InsightLink. K implementaci však není dostupný dostatek informací a není zde popsán ani formát anotace (pouhé konstatování, že je ve formátu XML), ani protokol pro přenos. Pro využití je tedy nutné implementovat knihovnu, což v heterogenním distribuovaném systému zavádí nutnost implementace pro různé programovací jazyky a různá prostředí. Pro moji práci je tedy takové řešení nevhodné a je lepší využít jednotný komunikační protokol.

3 Cíle práce

Hlavními nevýhodami dnešních systémů je, že neumožňují současně spolupracovat v reálném čase a využít strukturovaných anotací. Velkou překážkou ve využití je i nutnost pracovat s anotacemi pouze v jednom prostředí, kterým je obvykle webový prohlížeč, nebo editor dokumentů. Dokumenty, které se v průběhu anotování mění, jsou pro velké množství systémů rovněž nevyřešený problém.

Hlavní cíle mé práce jsou ukázat, že strukturované anotace umožní lepší popis sémantiky a snažší strojové zpracování anotací (např. podpora automatického odvozování informací), přičemž pokud tyto anotace využijeme při kolaborativním anotování v reálném čase, výrazně se tím usnadní a urychlí spolupráce při tvorbě dokumentů, folksonomií a dalších kolaborativních činnostech. Dále bych chtěl ukázat, že umožníme-li souběžné anotování v různých prostředích, zvýšíme tím uživatelskou přívětivost a možnosti využití anotačního systému. Takto bude mimo jiné umožněno i jednoduché propojení informací mezi interními dokumenty a širokým spektrem informací na webu. Přínosem bude rovněž současná podpora poloautomatického anotování, např. s využitím IKS FISE [23], a možnost anotovat i měnící se dokument, přičemž budou obsažené anotace automaticky či poloautomaticky aktualizovány. Zvýšením uživatelské přívětivosti lze dosáhnout širšího využití systému a získání většího množství anotací, ze kterých bude možné získat kvalitnější, relevantnější a úplnější informace. Takto dosáhneme nejenom nárůstu množství znalosti ve strojově dotazovatelné podobě, které budou využitelné pro podporu rozhodování, ale dosáhneme i širších možností využití anotací, např. anotací pro workflow (viz níže).

K dokázání těchto hypotéz chci navrhnout univerzální řešení pro kolaborativní anotování v různých prostředích. Toto řešení by mělo být dostatečně obecné, aby jej bylo možné využít k různým účelům (včetně těch, které jsou uvedeny v [35]) a dosáhnout tak i rozšíření možností využití anotací.

Využitelnost navrženého řešení budu demonstrovat návrhem a implementací serveru pro práci s anotacemi a klientů, kteří budou tvořit editory anotací pro různá prostředí.

3.1 Nové způsoby anotování

Většina existujících nástrojů nabízí možnost anotování textovou poznámkou, případně tagy. Strukturované anotace jsou spíše výjimkou a ve většině nástrojů nejsou dostupné. Jedním z dílčích cílů méj práce je navrhnout anotační nástroj, který umožní využití strukturovaných anotací spolu s klasickými poznámkami a tagováním. Bude tak možné vytvářet jak jednoduché anotace, na které jsou uživatelé zvyklí z existujících nástrojů, tak i strukturované anotace s atributy různých datových typů, vnořenými anotacemi a vazbami na jiné anotace.

Některé současné systémy umožňují kolaborativní práci s anotacemi, ale souběžný přístup obvykle není dobře ošetřen. Anotace jsou aktualizovány pouze v okamžiku načtení či obnovení dokumentu a následně může dojít k situaci, kdy někdo edituje neaktuální verzi anotace, případně ke ztrátě dat. Většina systémů navíc pracuje pouze se statickými dokumenty či snímky dokumentů z určitého časového okamžiku. Pokud se dokument mění, může být problém s aktualizací anotované verze dokumentu. Součástí mého řešení bude i řešení pro kolaboraci v reálném čase, kdy budou ihned přenášeny nejenom anotace, ale i změny v anotovaném dokumentu. V případě dokumentu, který je právě editován, potom může server automaticky či s podporou uživatele aktualizovat anotovanou verzi dokumentu, včetně anotací, které s ní souvisí.

3.2 Anotování kdekoliv a kdykoliv

Dostupné nástroje umožňují anotovat dokumenty v určitém formátu. Není však možné dokument anotovat ve více formátech současně (např. tvůrce webové stránky ji anotuje v textovém editoru ve formátu prostého textu či v syntaxi wiki, zatímco návštěvník této stránky ji anotuje ve webovém prohlížeči ve formátu HTML). Jedním z cílů méj práce je navrhnout řešení, které umožní anotovat text v různých formátech současně a převádět anotace mezi jednotlivými formáty dokumentu.

Většina nástrojů také umožňuje anotovat v určitém prostředí. Tím je obvykle webový prohlížeč. Není však možné anotovat v různých prostředích současně (např. ve webovém prohlížeči a v textovém editoru). Součástí méj práce bude i řešení této situace. Uživatel potom bude mít možnost anotovat text v textovém editoru na svém pracovním počítači a pokračovat v jednoduchém editoru s webovým rozhraním na chytrém mobilním telefonu v průběhu konference. Anotovat tedy bude možné kdekoliv (v libovolném prostředí) a kdykoliv (kdykoliv má uživatel přístup k síti, sám nebo současně se spolupracovníky).

3.3 Nové možnosti využití anotací

Na základě anotací bude možné dokumenty automaticky zpracovávat, indexovat, třídít, propojovat související informace a předávat mezi systémy. Jako součást demonstrace výsledků využiji ve své práci propojení systémů a správu workflow. Např. pokud při editaci dokumentace k projektu řešitel napíše, jaké prostředky mu chybí k další práci, a tuto informaci doplní příslušnou anotací, na základě anotace může být automaticky spuštěn proces schvalování přidělení prostředků a jejich následného přidělení.

Informace poskytnuté v anotacích mohou být interpretovány dle systému, který informace přijal. Např. v dokumentu s informacemi o požadovaných prostředcích, který dorazí k vedoucímu projektu, budou anotace označující prostředky využity k jejich zvýraznění a k rozpoznání, že se jedná o dokument s žádostí o přidělení prostředků. Pokud vedoucí žádost schválí, může doplnit další anotace s případnými omezeními pro poskytované prostředky a dokument předá systému, který na základě původních i přidaných anotací zjistí typy a aktuální umístění poskytovaných prostředků a zašle informace osobám či dalším systémům schopným prostředky přidělit. Kopie zaslaná řešiteli umožní zjištění aktuálního stavu žádosti a získání informací o tom, které prostředky byly schváleny.

4 Zvolené metody řešení

Po prostudování existujících řešení jsem zjistil, že bude třeba navrhnout nový formát anotace a protokoly pro přenos anotací. Následně bude možné navrhnout server pro práci s anotacemi a klienta, který bude poskytovat uživatelské rozhraní.

Vývoj v oblasti anotací se dle [61] dělí na dvě hlavní kategorie. Jednou je vývoj datových schémat a druhou vývoj anotačních systémů. Vzhledem k výše uvedenému bude moje řešení v obou těchto oblastech, přičemž návrh formátu anotace spadá do první oblasti a návrh protokolu pro přenos anotací, klienta a serveru do druhé. Popsaný protokol pro přenos anotací jsem proto navrhl tak, aby jej bylo možné využít i s jiným formátem anotace (s omezením na formát XML).

Klienti a server budou společně tvořit nový anotační systém. Při návrhu tohoto systému zvážím nejenom nové požadavky dle cílů práce, ale i aspekty uvedené v literatuře, jako např. základní operace a procedury, které by měly

anotační systémy poskytovat dle [8], a požadavky na anotace a anotační systémy dle [36]. Pro návrh uživatelského rozhraní potom budou hodnotným zdrojem požadavky na uživatelské rozhraní anotačního systému uvedené v [65]. Takto docílím vytvoření systému, který překoná známá omezení a nevýhody existujících systémů.

4.1 Formát anotace

Protože existující formáty anotací neposkytují dostatečnou flexibilitu a není do nich možné jednoduše uložit všechny potřebné informace, bude nutné navrhnout nový formát strukturované anotace. Tento formát musí umožnit jak uložení jednoduchých anotací, které jsou uživatelé zvyklí vytvářet v jiných systémech, tak i ukládání komplexních strukturovaných anotací.

Vzhledem k požadavku na univerzálnost je nutné, aby bylo možné anotace využívat i jako tagy a ve vytvořeném systému tagovat. Navržený formát tedy musí umožnit i tuto alternativu.

Anotace budou strukturovány pomocí atributů (obdobně jako v [9]). Oproti [69] tedy bude možné atributy využít nejenom k vytváření vazeb mezi anotacemi, ale také k uložení hodnot jednoduchých datových typů, což umožní vytvořit i vnitřní strukturu anotace.

U některých systémů je při přidání nového atributu anotace nutné upravit klienty (např. Annotea [25]). Této situaci je nutné se při návrhu vyhnout a umožnit jednoduché přidání libovolných atributů, což ovlivní nejenom návrh formátu anotace, ale následně i návrh klientů.

Vzhledem k tomu, že dle [13] je prokázáno, že největším problémem anotačního systému je osiřování anotací (situace, kdy po změně dokumentu již nelze najít původní umístění anotace), formát anotace musí být navržen tak, aby obsahoval dostatečně robustní popis pozice anotovaného fragmentu.

4.2 Protokol pro přenos anotací

Anotace je možné ukládat zcela nezávisle na anotovaných dokumentech, nebo společně s kopiemi dokumentů vytvořenými v době anotování. Pokud se při využití prvního přístupu dokument mění, anotace rychle ztrácejí platnost, protože po mnoha změnách již nemusí být možné určit původní pozici anotace. Při využití druhého přístupu může dojít k situaci, kdy se anotovaná kopie vyvíjí nezávisle na původním dokumentu. Tato situace je dle [36] špatná, protože anotace by měla vždy anotovat dokument samotný a ne něco jiného. Tento problém vyřeším tak, že společně s anotacemi bude uložena i kopie dokumentu, která bude průběžně aktualizována, včetně všech obsažených anotací. Při menších změnách je vyšší pravděpodobnost správné aktualizace anotace, a proto bude vhodné při editaci dokumentu ihned přenášet jeho změny.

Existující protokoly pro přenos anotací však neumožňují přenos změn anotovaného textu a dalších potřebných informací, navíc nejsou vhodné pro navržený způsob anotování. Součástí řešení tedy musí být i návrh nového protokolu pro přenos anotací.

Při komunikaci mezi klientem a serverem je třeba přenášet jiné informace, než při komunikaci mezi servery pro práci s anotacemi. Klienti navíc budou se serverem komunikovat pomocí protokolu HTTP, který není navržen pro obousměrnou asynchronní komunikaci, což přináší určité problémy. Z těchto důvodů pro komunikaci mezi servery navrhu jiný protokol, než pro komunikaci mezi klientem a serverem.

Protokol pro komunikaci mezi klientem a serverem

Navržený protokol pro komunikaci mezi klientem a serverem musí umožnit nejenom přenos samotných anotací oběma směry, ale také přenos typů anotací. Obdobně jako u [26] musí být možné jej bez další vrstvy přenášet v HTTP. Před zahájením anotování bude nutné provést synchronizaci dokumentu a v průběhu anotování přenášet změny v tomto dokumentu. Mezi přenášené informace bude patřit i nastavení klienta a serveru.

Uživatel může požadovat pouze anotace z určitého zdroje, kterým může být jiný uživatel, skupina uživatelů či jiný zdroj. Protokol tedy musí umožnit přihlašování a odhlašování od zdrojů anotací.

Anotace mohou být uživateli nabídnuty serverem, který je získá pomocí automatické extrakce informací z textu. Uživatel je potom může pouze schvalovat. Nabízené anotace musejí být přenášeny odděleně od uložených anotací.

Protokol musí umožnit i jednoduchou správu sezení a možnost budoucího rozšíření o správu přístupu k anotacím, která v první verzi protokolu není potřebná.

Protokol pro komunikaci mezi servery

Distribuované systémy nemají centrální autoritu ke koordinování změn. Aby v mnou vytvářeném systému nenastaly související problémy, je nutné dbát na korektní práci se změnami v dokumentu a souvisejícími změnami v anotacích nezávisle na serveru, ke kterému je uživatel připojen.

Vzhledem k tomu, že v distribuovaném prostředí může být dokument (včetně všech anotací, které s ním souvisejí) uložen na libovolném serveru, je při synchronizaci před zahájením anotování nutné jej vyhledávat nejenom v lokální databázi, ale také v databázích všech spolupracujících serverů. Pokud bude dokument nalezen na jiném serveru, je nutné s ním pracovat přímo na tomto serveru. Protokol tedy musí umožnit vyhledávání dokumentu a vzdálenou manipulaci s tímto dokumentem a souvisejícími anotacemi.

Protože všichni uživatelé by měli mít přístup ke kompletnímu stromu typů anotací, který je v každé skupině uživatelů sdílený, protokol musí umožnit vzájemnou synchronizaci stromů typů po připojení dalšího serveru a přenos změn ve stromech typů mezi všemi servery.

Když bude server poskytovat pokročilejší služby, může např. vyhledávat dokumenty podle anotací apod. Rovněž k této funkcionalitě musí být podpora v protokolu. Protože může být průběžně přidávána i další funkcionalita, protokol musí být snadno rozšiřitelný.

Servery se musejí vzájemně autentizovat a udržovat informace o aktivitě protistrany (udržování seznamu aktivních spolupracujících serverů).

Pro přenos navrženého protokolu jsem zvolil protokol XMPP [67], který nabízí vhodnou funkcionalitu.

4.3 Server pro práci s anotacemi

Nejprve navrhnu jednoduchý server pro práci s anotacemi, který bude poskytovat pouze minimální nutnou funkcionalitu. Tato funkcionalita zahrnuje:

- ukládání a aktualizace (synchronizace, zaznamenávání změn) kopií dokumentů,
- ukládání, aktualizace a mazání anotací,
- správu typů anotací,
- nabízení jednoduchých anotací (hledání slov v jednoduchém slovníku),
- správu uživatelských účtů s webovým rozhraním,
- správu uživatelských nastavení,
- základní komunikaci s klientem,
- základní práci s anotacemi.

Komunikace s klientem pokrývá rozsah komunikace základním navrženým protokolem. Nejsou zde začleněna rozšíření pro získání profilů a fotografií uživatelů a dalších doplňujících informací.

Základní práce s anotacemi zahrnuje základní kontrolu syntaxe anotací, distribuci změn mezi klienty přihlášenými k odběru apod. Nejsou zde obsaženy žádné pokročilé kontroly správnosti, ani kontroly zachování platnosti anotací při změnách dokumentu mimo rozsah samotného anotovaného fragmentu a cesty k němu.

Po implementaci jednoduchého serveru ověřím funkčnost navrženého konceptu a komunikačního protokolu. Následně navrhnu komplexnější variantu serveru, která umožní pokročilejší práci s anotacemi. Protokol pro komunikaci s klientem bude doplněn o jednoduchá rozšíření, která umožní přenos profilů uživatelů a dalších doplňujících informací. Pokročilejší server implementuji a demonstruji na něm pokročilé možnosti využití anotací.

V poslední fázi práce navrhnu a implementuji variantu serveru, která bude umožňovat nasazení v distribuovaném prostředí. Bude tedy komunikovat s ostatními servery a udržovat potřebné informace (viz výše).

4.4 Klienti

Klienti budou editory anotací, které poskytnou (obvykle grafické) uživatelské rozhraní. U klienta navrhnu především koncept, základní návrh uživatelského rozhraní a základní a volitelnou funkcionalitu, protože předpokládám implementace od různých autorů, kteří vytvoří doplňky do celé řady existujících aplikací (svých či od třetích stran). Jedním z prvních klientů bude doplněk do textového editoru TinyMCE [32], který je vyvíjen v rámci diplomové práce na FIT VUT v Brně.

Klienty lze dle aplikace rozdělit na 3 druhy:

- doplňky do textových editorů,
- doplňky do prohlížečů dokumentů,
- další klienti.

Doplňek do textového editoru je specifický tím, že pracuje s měnícím se textem. Uživatel může kdykoliv zasáhnout do kterékoliv části textu a jednotlivé fragmenty přidávat, upravovat či mazat. Při práci s anotacemi je potom třeba na tyto akce odpovídajícím způsobem reagovat. Pokud je anotovaný fragment modifikován, je třeba, aby uživatel rozhodl, zda bude anotace upravena nebo vymazána. Zachování původní anotace není možné, protože anotovaný text je její součástí. Vždy je tedy třeba upravit alespoň anotovaný fragment. Pro vizualizaci to znamená např. umístění značky na místo počátku původního fragmentu a zobrazení anotace s výběrem akce při najetí myši na tuto značku. Textový editor může pracovat jak se strukturovaným, tak i s nestrukturovaným textem. Každý strukturovaný text může být převeden do formátu XML a pozice fragmentu v textu vyjádřena pomocí XPath, offsetu a délky. Server bude jako výchozí formát využívat (X)HTML a každý klient, který pracuje s jiným strukturovaným formátem, tedy musí pro účely synchronizace dokument převést do tohoto formátu a následně přepočítávat pozice anotací. Strukturovaný text lze převést i na nestrukturovaný (linearizovat) a pozici fragmentu v textu přepočítat pouze na offset a délku. Při převodu je však důležité využít jednotný algoritmus, aby při zpětném převodu pozice nebyla určena chybně. Server tedy nejprve při synchronizaci porovná, zda se linearizace od klienta shoduje s jeho linearizací, a následně sám vypočítává pozice anotovaných fragmentů ve strukturovaném textu. Takto může pracovat s daty od různých doplňků nezávisle na tom, v jaké formě uživatel s textem pracuje. Stejný text je tedy možné anotovat v jednoduchém editoru i ve webové stránce.

Doplňek do prohlížeče dokumentů pracuje se statickým textem. Typickým příkladem je doplňek do webového prohlížeče, který umožňuje anotovat zobrazenou webovou stránku. U webové stránky je třeba pracovat s textem ve strukturované formě, protože kolem anotovaného textu mohou být ve stránce i měnící se části, které k anotovanému textu nepatří (např. reklamy). Pokud se jedná o prohlížeč jiných dokumentů (např. PDF), doplňek může dokument převést do (X)HTML a přepočítávat pozice anotací, nebo dokument linearizovat a pracovat s ním v linearizované podobě.

Mohou být i další varianty klientů do různých aplikací. Vždy je však třeba, aby byl anotovaný dokument jednoznačně identifikovatelný a bylo možné jej převést do XML a nalézt pozici anotovaného fragmentu.

Pokud se bude pracovat s jiným strukturovaným formátem než (X)HTML, bude na serveru uložena pouze linearizovaná podoba dokumentu, nebo dokument převedený do (X)HTML. Vzhledem k tomu, že vizualizace je prováděna s originálním dokumentem, na formátu uložení na serveru nezáleží. Je však důležité, aby byl převod formátů prováděn vždy stejným způsobem, na kterém se autoři doplňků pracujících s tímto formátem musejí shodnout. Pokud budou různé doplňky využívat různé převody formátů, nebude možné spolupracovat mezi různými doplňky (bude docházet k chybám synchronizace a zneplatňování anotací). Vzhledem k velkému množství různých formátů dokumentů a jejich vývoji není možné definovat převody pro všechny možné formáty, ani jeden obecný převod. Pokud se bude pracovat s linearizovanou verzí dokumentu, která bude obsahovat všechny anotovatelný text, problémy by měly být minimální.

Uživatelské rozhraní je velice důležité, protože tvorba digitální anotace je pracnější než u papírové [28] a čím je složitější, tím méně uživatelé anotují. Pokud jsou navíc anotace využity ke spolupráci, uživatelské rozhraní ovlivňuje nejenom počet, ale i typy problémů, které spolupracovníci pomocí anotací řeší [65]. Pokud tedy chceme dosáhnout toho, aby uživatelé vytvářeli velké množství anotací, které bude možné dále využívat, musí být anotování jednoduché. Současně je však nutné dbát na to, aby uživatel ve své práci nebyl příliš omezen. Zatímco v papírové podobě může uživatel vytvořit libovolnou anotaci, v elektronické formě je mu vždy vnučována určitá forma anotace a způsob anotování [28]. Strukturované anotace jsou jednou ze složitějších forem anotace, ale mají velké vyjadřovací schopnosti (minimální omezení). Pokud bude jejich vytváření pomocí navrženého uživatelského rozhraní rychlé a jednoduché, můžeme z anotací získat velké množství hodnotných informací.

Podle složitosti lze klienty rozdělit na 3 varianty:

- zjednodušený klient pro základní anotování,
- klient se základní podporou strukturovaných anotací,
- klient s pokročilou podporou strukturovaných anotací.

Nejjednodušší variantou je zjednodušený klient pro základní anotování. Tato varianta nepodporuje strukturování anotací, umožňuje tedy pouze volbu typu a zadání textového obsahu anotace. Práce s atributy není podporována, ale při editaci anotace jsou atributy v nezměněné podobě zaslány na server (je třeba je uchovávat).

Základní varianta klienta nabízí podporu strukturovaných anotací. Je možné pracovat se strukturováním anotací v plném rozsahu, ale klient „nerozumí“ typům anotací, a nemůže tedy uživateli pomoci s jejich tvorbou. Uživatelské rozhraní musí poskytovat funkcionalitu pro vytváření a prohlížení všech možných strukturovaných anotací, ale nemusí poskytovat pokročilejší vizualizaci (např. zobrazení fotografií u osob apod.).

Klient s pokročilou podporou strukturovaných anotací musí oproti základní podpoře strukturovaných anotací nabízet i pokročilejší vizualizační funkce (např. zobrazení fotografií a profilů u osob, mapy u geografického umístění apod.) a může mít vylepšenou podporu některých složitějších typů anotací (např. průvodce vytvořením úkolu).

5 Shrnutí dosavadních výsledků řešení

V této kapitole jsou uvedeny dosavadní výsledky řešení. Nejprve jsem navrhl nový formát anotace a protokol pro přenos anotací. Následně jsem navrhl jednoduchý server pro práci s anotacemi a funkcionalitu klienta. Celá kapitola je tedy věnována technickému návrhu systému, přičemž v návrhu klienta jsou obsaženy návrhy jednotlivých částí uživatelského rozhraní, které jsou pro přehlednost doplněny velkými obrázky.

5.1 Návrh formátu anotace

Navrhl jsem nový formát strukturované anotace, který je oproti jiným formátům obecnější. Využil jsem obdobný přístup jako v [9], kde jsou definovány jednoduché a komplexní atributy, ale navíc jsem přidal podporu rozšířených datových typů (viz níže). Atributy tak umožňují nejenom seskupovat anotace, ale mohou obsahovat i různé datové typy a součástí anotace tedy může být např. geografické umístění apod.

Vztahy mezi anotacemi propojenými pomocí atributů jsou popsány názvem a typem atributu, čímž se moje řešení liší od [9], kde je vztah popisován názvem a samostatnou odkazovanou anotací. Moje řešení je tak méně komplexní z hlediska zpracování a jednodušší pro uživatele. Vzhledem k tomu, že složitější vztah, pro jehož popsání nestačí název a typ odkazované anotace, lze popsat pomocí více atributů či úrovní vnoření anotací, nejsou omezeny vyjadřovací schopnosti.

Navržený formát anotace umožňuje různé způsoby anotování. Nejjednodušší je pouhé připisování jednoduchých poznámek k textu nebo značkování textu (tagging). Formát tedy může být využit k účelům, ke kterým se využívají současně dostupné formáty anotací. Navíc však poskytuje možnost strukturování anotací, a lze jej tedy využít i k mnoha dalším účelům, mezi které patří i výše zmíněná podpora workflow. Strukturování je umožněno pomocí atributů, kterými lze vyjádřit vazby na jiné anotace, nebo mohou být jiné anotace přímo obsaženy.

Vzhledem k tomu, že v případě změny v dokumentu je třeba určit, které anotace jsou nadále platné a které ne, je třeba mít možnost porovnat kopii dokumentu z okamžiku vzniku či poslední úpravy anotace s aktuálním stavem. Proto je nutné před anotováním uložit kopii anotovaného dokumentu na server a při dalších přístupech k tomuto dokumentu vyhodnotit rozdíly aktuálního stavu oproti anotované kopii a případně upravit anotace. Pozice anotovaných fragmentů jsou potom určeny v lokální kopii (stav dokumentu v okamžiku anotace či aktualizace dat).

K určení pozice anotovaného fragmentu ve strukturovaném dokumentu existují 2 přístupy:

- linearizace,
- cesta, offset a délka.

Při linearizaci je dokument nejprve linearizován a potom je pozice anotovaného fragmentu vyjádřena offsetem od začátku dokumentu a délkou či offsetem konce. Je však nutné, aby linearizace byla prováděna vždy stejným způsobem (stejný způsob průchodu DOM dokumentu). Pokud se potom změní některá z předcházejících částí v dokumentu, určení pozice je neplatné a je třeba jej upravit. Předchozí část však může být nejenom předcházející odstavec, ale také reklama v záhlaví webové stránky či jiný nesouvisející obsah. V takovém případě může být přepočítávání pozice potřebné při každém přístupu k anotovanému dokumentu, přičemž se musí vyhodnotit celý dokument a porovnat s anotovanou kopií na serveru.

Pokud je pozice určena cestou v DOM dokumentu, offsetem a délkou, není třeba vždy vyhodnocovat celý dokument. Pokud se např. změní reklama v záhlaví webové stránky, tato reklama je obvykle v uzlu DOM dokumentu, který není v cestě k anotovanému fragmentu. Celá cesta tedy při různých drobných změnách zůstává stále platná. Navíc je snazší nalézt pozici anotovaného fragmentu v případě, kdy je dokument vložen do jiného kontextu, např. jiné webové stránky (přechod na novou verzi stránek). Pokud se totiž změní prvních několik uzlů v cestě (při změně rozložení stránky), ale cesta v rámci dokumentu zůstane stejná a dostatečně dlouhá, aby bylo možné téměř jednoznačné určení, je možné anotovaný fragment nalézt. Pokud by byl fragment určen pouze offsetem a délkou, při změně obalujícího rozložení stránky už by bylo možné hledat pouze podle obsahu fragmentu, což je v případě, kdy je anotováno jedno slovo, problém (nejednoznačnost, nemožnost určit, zda nejsme v jiném dokumentu). Ukládání pozice v DOM dokumentu navíc podle [11] a [37] patří mezi robustnější způsoby určení pozice anotovaného fragmentu textu. Protože však pomocí cesty určíme pouze uzel DOM dokumentu, je třeba ještě upřesnění pozice fragmentu (např. věty, slova či znaku) pomocí offsetu a délky.

Vzhledem k výše popsaným výhodám jsem zvolil určení pozice fragmentu pomocí cesty, offsetu a délky. Pokud se následně bude pracovat s nestrukturovaným (linearizovaným) dokumentem, cesta bude ukazovat na kořen dokumentu a pozice bude určena pouze offsetem a délkou.

Formát anotace je založen na RDF/XML, kde podmětem je vždy anotace. Formát XML je vhodný zejména pro možnost popsat libovolné struktury a pro efektivní zpracování ve webových prohlížečích [42].

Součástí anotace jsou:

- typ anotace,
- datum a čas vytvoření,
- autor,
- URI anotovaného dokumentu (kopie na serveru),
- cesta k anotovanému fragmentu (XPath),
- offset anotovaného fragmentu,
- délka anotovaného fragmentu,
- textový obsah anotovaného fragmentu,
- textový obsah anotace,
- atributy.

Typy anotací jsou hierarchické a jsou tedy uspořádané do stromové struktury. Základními typy jsou nejběžnější typy anotací (poznámka, popis apod.) a základní typy entit (věc, člověk apod.). Uživatelé mohou následně přidávat nové podtypy, čímž budou vytvářet komplexní strom typů. Tímto je umožněno odlišit nejenom různé typy anotací, ale také typy anotovaných fragmentů. Typy potom lze využívat jako tagy a uživatel tedy může anotacemi pouze tagovat. Každá skupina uživatelů sdílí svůj strom typů. Globálně tedy lze na typy pohlížet jako na strom, kde skupiny uživatelů vytvářejí jednotlivé větve. V tomto stromě jsou typy identifikovány skupinou uživatelů, svojí pozicí a názvem. Pokud je typ uveden v atributu či textovém poli, je identifikován svým linearizovaným názvem či URI. Oba způsoby jsou vzájemně převoditelné, protože URI musí vždy končit skupinou uživatelů následovanou cestou v hierarchii typů, kde jednotlivé typy jsou odděleny lomítky. Linearizovaný název je cesta v hierarchii typů, kde jednotlivé typy jsou odděleny sekvencí znaků „->“. V linearizovaném názvu skupina není uvedena, protože se předpokládá její volba v jiném poli či využití parametru nastavení `DefaultGroup` (viz níže). Linearizovaný název je uživatelsky přívětivější.

Autor anotace bude identifikován svým jednoznačným identifikátorem (URI), který má na serveru přiřazen. Zobrazené jméno (přezdívka) bude v atributu `name` a v dalších volitelných attributech mohou být jiné doplňující informace.

URI anotovaného dokumentu identifikuje kopii anotovaného dokumentu, která je uložena na serveru. Při zahájení anotování nejprve dojde k synchronizaci, při které klient zašle serveru URI anotovaného dokumentu a celý obsah tohoto dokumentu. Server potom vrátí URI, který bude využit v anotacích (viz níže).

Cesta k anotovanému fragmentu jednoznačně určí oblast dokumentu (nadpis, odstavec, buňku tabulky apod.), ve které se nachází anotovaný fragment. Offset a délka určí přesnou pozici fragmentu. Textový obsah anotovaného fragmentu obsahuje samotný anotovaný text, který je třeba mít uchován pro případ, že se dokument změní. Pokud po změně fragment nebude v textu nalezen, bude označen atributem `valid` s hodnotou `false`, ale stále zůstane součástí anotace, aby informace o anotovaném textu zůstala kompletní. Anotace může náležet i k více fragmentům textu, i když v návrhu klienta z hlediska uživatele tato funkcionalita nebude obsažena, protože obvykle není potřebná (buď lze fragmenty anotovat samostatnými anotacemi, nebo vložit do atributů jiné anotace, které mohou popsat jednotlivé fragmenty i jejich vztah).

Vzhledem k tomu, že dle [30] jsou nejčastěji anotovány části vět a celé věty, obvykle nedojde k situaci, kdy by byl text vybrán přes více uzlů DOM dokumentu. Pokud však k této situaci dojde, budou součástí vybraného textu rozděleny do více fragmentů. Alternativou by bylo označení prvního a posledního uzlu, nebo označení počátečního uzlu a využití délky textu. Obě alternativní řešení se sice prakticky využívají, ale jsou méně robustní (odolná vůči změnám v dokumentu), a proto nejsou pro moje účely vhodná.

Kromě textového obsahu anotovaného fragmentu by bylo možné ukládat i kontext, který je tvořen částmi textu před a za anotovaným fragmentem. Význam kontextu při určování platnosti anotace je však malý, což bylo dokázáno v [11], a v případě potřeby je možné jej jednorázově vyhodnotit při aktualizaci anotované kopie dokumentu na serveru. Právě při této aktualizaci je totiž nutné vyhodnotit platnost anotací a ty případně upravit, přesunout na globální úroveň dokumentu či zneplatnit (vymazat). Pro úpravy, případně přesouvání na globální úroveň, se využívají tzv. algoritmy pro změny pozic anotací (repositioning algorithms či reattachement algorithms), jejichž využití je popsáno v [7], [37] a [38]. Přehledná tabulka srovnání těchto algoritmů, přesněji jejich robustnosti vůči jednotlivým změnám, je uvedena v [38].

Přesunutí anotace na globální úroveň lze nazvat také osiřením (orphaning), kdy již v dokumentu nelze nalézt anotovaný fragment, nebo se tento fragment příliš změnil a původní anotace pro nový obsah fragmentu neplatí. Podrobněji

se touto problematikou zabývá [11]. V případě, že anotace náleží k více fragmentům, může při osiření jednoho fragmentu zbytek anotace zůstat platný.

Textový obsah anotace obsahuje uživatelem zadanou textovou informaci. Význam je dán typem anotace, přičemž se může jednat o textovou poznámku, komentář, popis, doplňující informaci apod.

Atributy umožňují strukturování anotace. Každý atribut má název, typ a hodnotu. Typem může být některý ze základních datových typů z XSD (číslo, řetězec, datum apod. viz [60]), rozšiřujících datových typů (např. geografické umístění ve formátu W3C Basic Geo [10]), vnořená anotace či odkaz na anotaci. V případě vnořené či odkazované anotace je typ atributu přesněji určen typem anotace. Odkaz na anotaci umožňuje napojení na jinou anotaci, seskupování anotací apod.

V řadě existujících formátů jsou součástí anotace i speciální atributy umožňující spolupracujícím uživatelům reagovat na dané anotace, přičemž vznikají vlákna či množiny anotací [3]. V mnou navrženém formátu tyto speciální atributy nejsou potřebné, protože lze k těmto účelům pouze vyhradit jeden typ atributu, který navíc může mít podtypy dle charakteru reakce.

Ukázka strukturované anotace je v příloze A.

5.2 Návrh protokolu pro přenos anotací mezi klientem a serverem

Pro přenos anotací mezi klientem a serverem jsem navrhl nový protokol, který kromě přenosu samotných anotací umožňuje jednoduchou autentizaci, přenos nastavení, přenos anotovaného dokumentu (synchronizaci), přihlášení k odběru anotací z vybraných zdrojů, přenos typů anotací a další potřebnou komunikaci. Navíc je možné nechat si anotace nabídnout serverem. Oproti API navrženému v [18] předpokládá jednodušší formát anotace (bez explicitně uloženého popisu struktury) a zjednodušuje identifikaci v distribuovaném systému (URI místo GUID). Rovněž zaslání anotovaného textu na server a nabídnutí anotací je, narozdíl od API z [18], podporováno. V první verzi protokolu jsem nenavrhl správu řízení přístupu k anotacím, ale protokol je připraven pro toto rozšíření. Alternativou je i využití vyhrazeného typu atributu anotace a vyhodnocení oprávnění na serveru.

Protokol umožní obousměrnou asynchronní komunikaci mezi klientem a serverem. Pokud jeden uživatel zadá anotaci, server ji ihned pošle všem ostatním, kteří od daného uživatele odebírají anotace a anotují stejný dokument. Okamžitě se zasílají i aktualizace typů anotací a obsahu dokumentu. Server může kdykoliv za běhu změnit i nastavení apod.

Zprávy budou zasílány ve formátu XML a mohou být dle potřeby sdružovány (lze zaslat více zpráv v jednom XML). Zprávy lze rozdělit do následujících skupin:

- správa sezení,
- uživatelé a skupiny,
- řízení odběru anotací,
- synchronizace dokumentu,
- přenos typů anotací,
- přenos anotací,
- nabízení anotací,
- přenos nastavení,
- chyby a varování,
- prázdná odpověď.

Popis jednotlivých zpráv v těchto skupinách je umístěn v příloze B. Zjednodušený příklad komunikace mezi klientem a serverem je umístěn v příloze C.

5.3 Přenos anotací mezi servery

Protokol pro přenos anotací mezi servery navrhnu po ověření funkčnosti protokolu pro komunikaci mezi klientem a serverem. Při návrhu zohledním také potřeby zjištěné při implementaci pokročilejšího serveru pro práci s anotacemi.

Jako základ využiji protokol pro přenos anotací mezi klientem a serverem, který rozšířím o další potřebné zprávy a odstráním z něj zprávy, které v komunikaci mezi servery nemají smysl. Autentizační část protokolu bude zcela jiná.

5.4 Server

Jak bylo řečeno výše, nejprve jsem navrhl pouze jednoduchý server se základní funkcionalitou. Pokročilejší variantu navrhu po ověření konceptu.

Detailed návrh serveru zde nebudu uvádět, protože se jedná o obvyklý inženýrský návrh - je definována struktura databáze, jednotlivé třídy programu apod. Návrh je zatím v předběžné fázi a bude třeba jej upřesnit při implementaci, kdy již budou zřejmé všechny požadavky klientů na komunikaci (rozšíření oproti navrženému protokolu). Uvedu zde tedy pouze vybrané stručné základní informace:

- Serverová aplikace bude implementována v jazyce Java.
- Jako aplikační server bude využit GlassFish.
- Pro objektově relační mapování využiji EclipseLink (JPA 2.0) a pro ukládání dat databázi MySQL.
- Pro implementaci Comet využiji long polling s pomocí frameworku Grizzly.
- Kopie anotovaných dokumentů budou ukládány do databáze.
- Informace o sezeních pro jednotlivá spojení, jejichž vytváření je popsáno výše, budou udržovány v objektu na globální úrovni (na úrovni aplikace).
- Součástí objektu s informacemi o sezení budou i informace o zvolených zdrojích anotací.
- Server bude mít modulární strukturu a bude rozšiřitelný přidáváním dalších modulů, které bude možné zapínat a vypínat za běhu.

5.5 Klienti

Vzhledem k tomu, že nebudu přímo implementovat klienta, ale pouze jsem navrhl jeho funkcionalitu, v této podkapitole jsou umístěny především návrhy vzhledu obrazovek s popisy chování. Návrh provedu pouze pro textový editor, protože u ostatních programů budou jen malé rozdíly. Např. ve webovém prohlížeči, který jsem zvolil jako reprezentativní příklad prohlížeče dokumentů, anotace celého dokumentu musejí být umístěny v postranním panelu, protože zobrazení s plovoucími prvky, které je na obrázku 4, zde není možné.

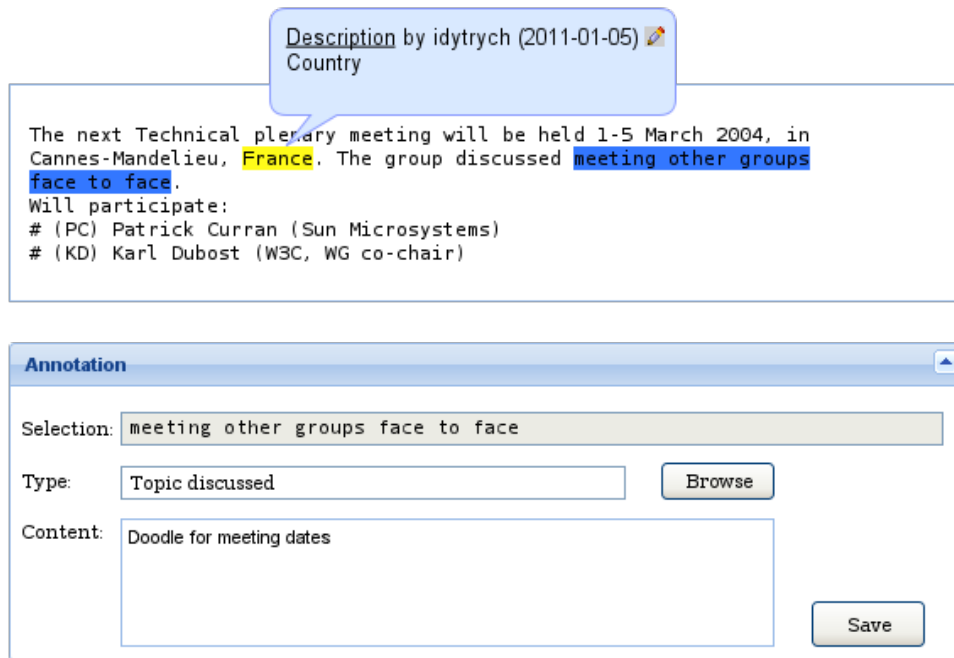
Nejjednodušší varianta rozšíření nedisponuje podporou strukturování anotací. Pokud je načtena strukturovaná anotace, je na tuto skutečnost uživatel upozorněn nějakou ikonkou u zobrazené anotace nebo jiným způsobem a strukturované informace nejsou zobrazeny. Jsou však uchovány a v případě úpravy anotace jsou v nezměněné podobě zaslány na server.

Základními nezbytnými prvky rozšíření pro zadávání anotací jsou pole pro výběr typu, pole pro zadání textového obsahu anotace (**content**) a tlačítko pro uložení (**save**).

Uživatel nejprve v textu označí anotovaný fragment, následně zadá informace o anotaci a uloží. Pro přehlednější práci uživatele je vhodné uživatelské rozhraní doplnit i o pole, ve kterém je zobrazen vybraný (**selection**), anotovaný text. Uživatel tak tento text vidí i samostatně mimo kontext, ve kterém jej označil, a může si lépe uvědomit, zda je rozsah výběru vhodně zvolený. Pole bude potřebné také u složitější varianty klienta, u které bude více výběrů (další vybrané fragmenty pro vnořené anotace).

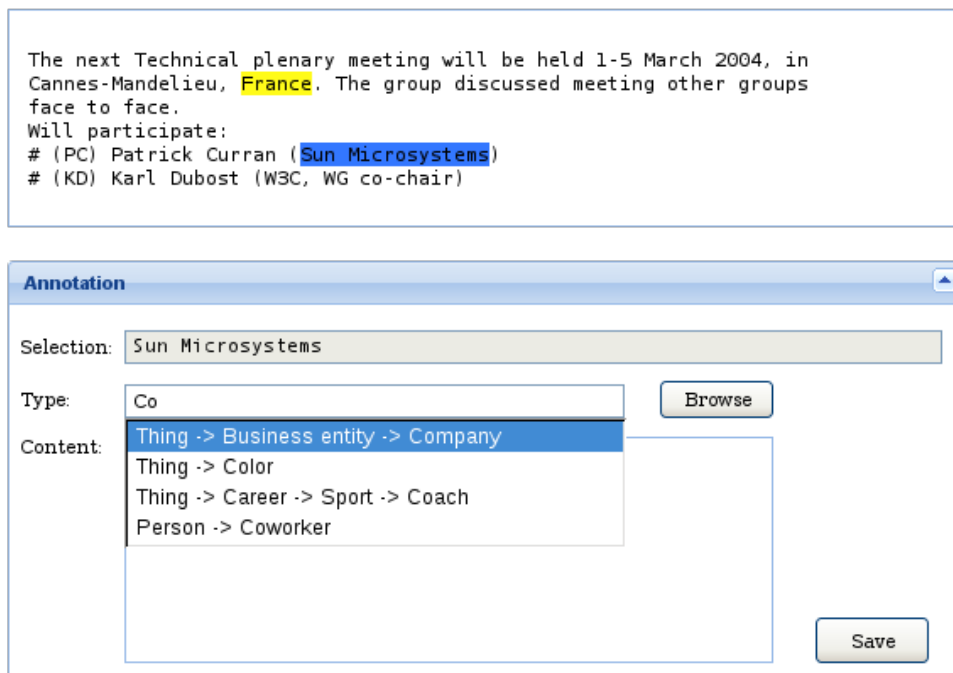
Nejjednodušší varianta klienta je na obrázku 1. Anotovaný text ve většině následujících obrázků jsem převzal z [58]. Existující anotace jsou zobrazeny zvýrazněním anotovaných fragmentů textu a v případě, že uživatel na tento fragment najede myší, je zobrazena bublina (tooltip) s informacemi o anotaci. Pokud uživatel na anotovaný fragment klikne, anotace zůstane zobrazena až do kliknutí mimo anotovaný fragment a uživatel může kliknout na ikonku pro její editaci. Při editaci je anotace vypsána do formuláře, uživatel může změnit požadované informace a změny uložit.

Vedle ikonky pro editaci může být u anotace zobrazena i ikonka pro vymazání anotace. Po kliknutí na tuto ikonku by se zobrazil potvrzovací dialog a při potvrzení by anotace byla vymazána.



Obrázek 1: Nejjednodušší varianta rozšíření

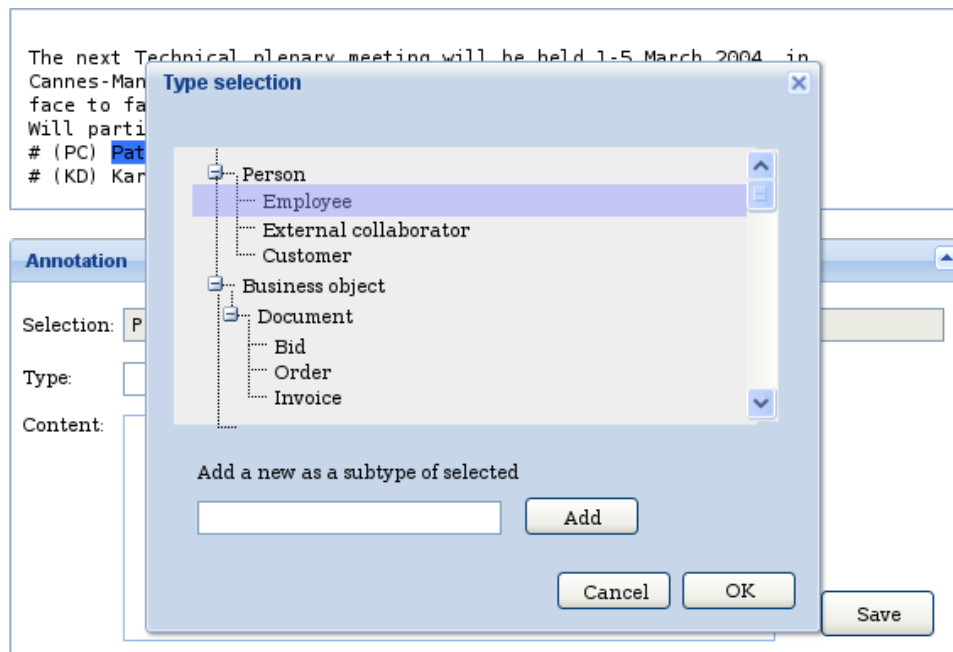
Typ anotace je možné zadat textově, ale je třeba uživateli umožnit i výběr z existujících typů. Tento může být realizován dvěma způsoby. Při prvním uživatel začne psát název typu a pod polem, ve kterém jej zadává, se mu zobrazí seznam typů, jejichž linearizovaný název obsahuje zadaný text. Tento seznam bude při psaní postupně upřesňován a uživatel si ve kterémkoliv okamžiku může ze seznamu vybrat. Zadání typu s nabízením je na obrázku 2.



Obrázek 2: Výběr typu anotace přímým zadáváním s nabízením

Jsou 2 možnosti implementace tohoto způsobu zadávání: načíst celý strom typů a podle zadaného řetězce zobrazovat nabídku, nebo při každé změně zadávaného textu načíst ze serveru typy k nabízení (filtrování typů). Druhá varianta je jednodušší, protože filtrování vyřeší server. Je ale také pomalejší. Pokud uživatel zadá neznámý typ, uloží se při uložení anotace i tento typ.

Druhým způsobem výběru typu je výběr ze stromu typů. Uživatel klikne na tlačítko vedle pole pro zadání typu a otevře se mu dialog, který je na obrázku 3. V tomto dialogu může uživatel strom typů také rozšiřovat či upravovat.



Obrázek 3: Výběr typu anotace ze stromu

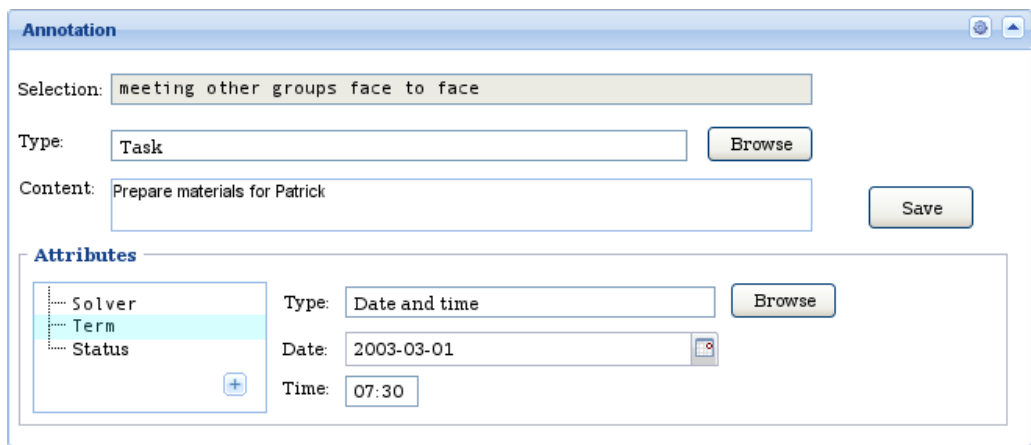
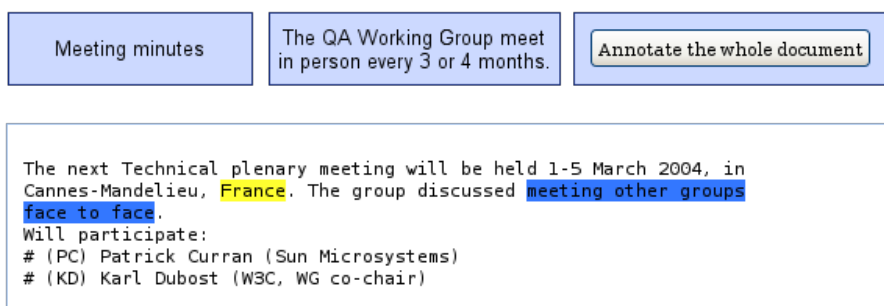
Při přidání nového typu se tento uloží na server až s anotací, aby u strukturované anotace bylo možné uložit i seznam atributů. Pokud je však typ přidán v dialogu se stromem anotací, je přidán ihned a při uložení anotace je pouze aktualizován. Takto může být v dialogu přidáno více typů, i když u nich nebudou uvedeny atributy. Další variantou by bylo rozšířit dialog se stromem typů i o zadávání seznamů atributů.

Tím, že uživatel volí typy anotací, které slouží současně jako tagy, ze stromu, je mu umožněno, aby se mohl pokusit nalézt odpovídající tag v hierarchii existujících, než zvolí vlastní termín. Takto je dosaženo snížení diversity tagů, které např. v [52] řeší pomocí seskupování tagů a nabízení souvisejících k některému z právě zadaných.

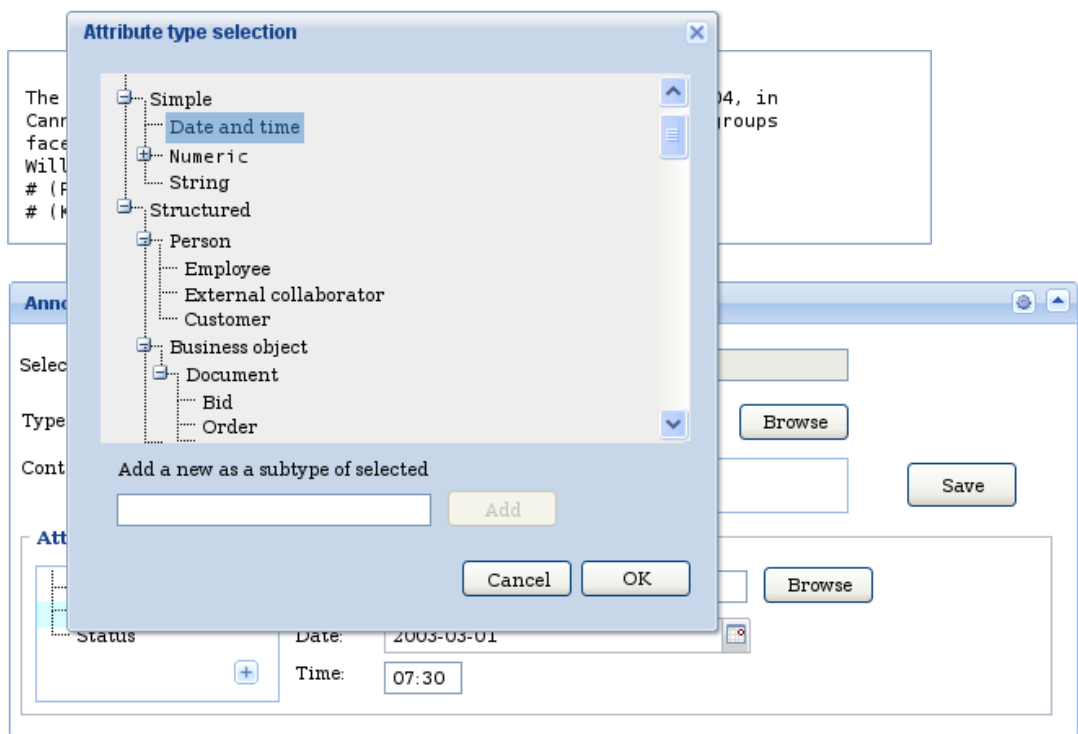
Na obrázku 4 je základní klient s podporou strukturovaných anotací. Uživatelské rozhraní je doplněno o strom atributů. Pokud je hodnotou atributu vnořená anotace, její atributy budou tvořit větve stromu vycházející z daného atributu. Takto je možné přehledně zobrazit libovolný počet úrovní vnoření. Každý atribut má název, který je zobrazen ve stromě, typ a hodnotu. Pole pro výběr typu je obdobné, jako u výběru typu anotace, pole pro zadání hodnoty jsou zobrazena dle konkrétního typu.

Anotovat lze nejenom fragmenty textu, ale také celý dokument (pozice fragmentu v anotaci není uvedena). V tomto případě musí být anotace zobrazeny v postranním panelu nebo plovoucích prvcích v záhlaví dokumentu. Aby bylo možné celý dokument vybrat (nikoliv jako jeden velký fragment), je nutné do postranního panelu či plovoucího prvku umístit i tlačítko, které celý dokument vybere. V poli s anotovaným fragmentem se potom zobrazí speciální hodnota, která tento výběr signalizuje (např. „Vybrán celý dokument“). Pole s anotovaným fragmentem by pro vyšší přehlednost mohlo změnit i barvu pozadí.

Pro výběr typu atributu lze opět využít přímé zadávání s nabízením nebo výběr ze stromu. Vzhledem k tomu, že typem atributu může být jednoduchý typ i typ anotace, je zde zobrazen stejný strom jako u výběru typu anotace, ale rozšířený o jednoduché a rozšiřující (rozšíření množiny jednoduchých typů) datové typy. Rozšířený strom je na obrázku 5. Do podstromu s jednoduchými datovými typy uživatel nemůže přidat nový typ (tlačítko pro přidání se zablokuje).

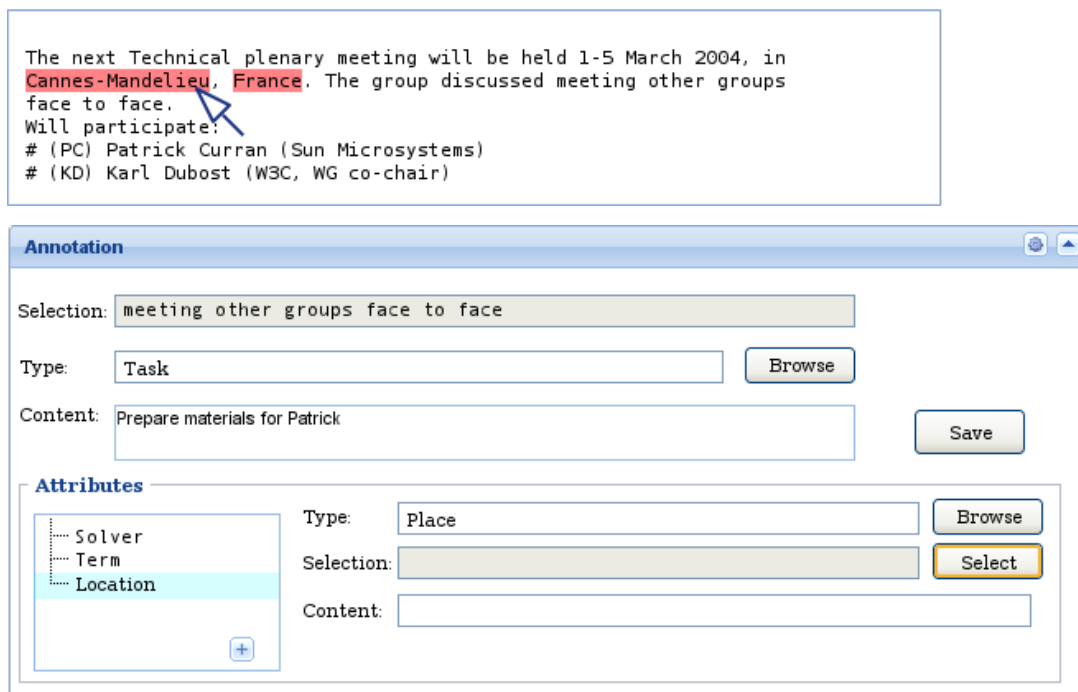


Obrázek 4: Základní klient s podporou strukturovaných anotací

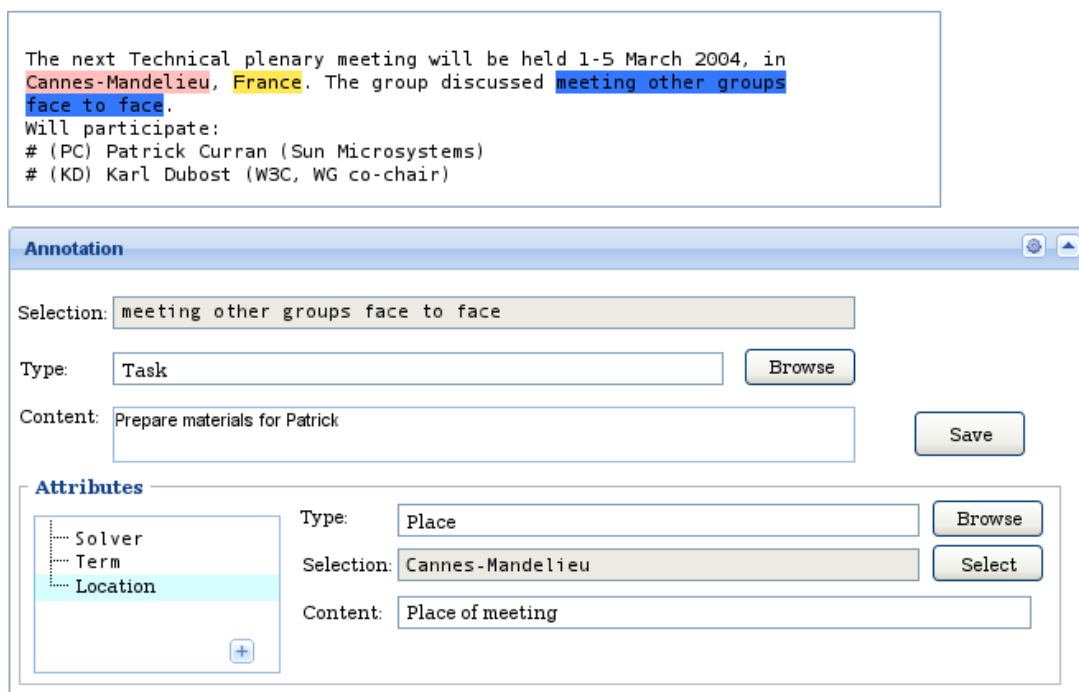


Obrázek 5: Výběr typu atributu ze stromu

Na obrázku 6 je ukázka formuláře pro zadávání vnořené anotace. Pokud uživatel klikne na tlačítko pro úpravu anotovaného fragmentu, aniž by měl zadány jiné informace o anotaci, a v dokumentu jsou anotace zvoleného typu, tyto anotace se zvýrazní. Pokud uživatel na jednu z těchto anotací klikne, místo vnořené anotace bude do atributu vložen odkaz na anotaci. Pokud je zvolen odkaz na anotaci, pole pro obsah anotace zešedne, zobrazí se atributy dané anotace a nebude možné jí přidávat další. Pokud uživatel nechce vložit odkaz, ale vnořenou anotaci, může před výběrem zadat obsah anotace nebo místo kliknutí na zvýrazněnou anotaci tažením myši vybere stejný či jiný fragment (viz obrázek 7). Po dokončení výběru se opět označí anotovaný fragment anotace nejvyšší úrovně.



Obrázek 6: Zadávání odkazu na anotaci

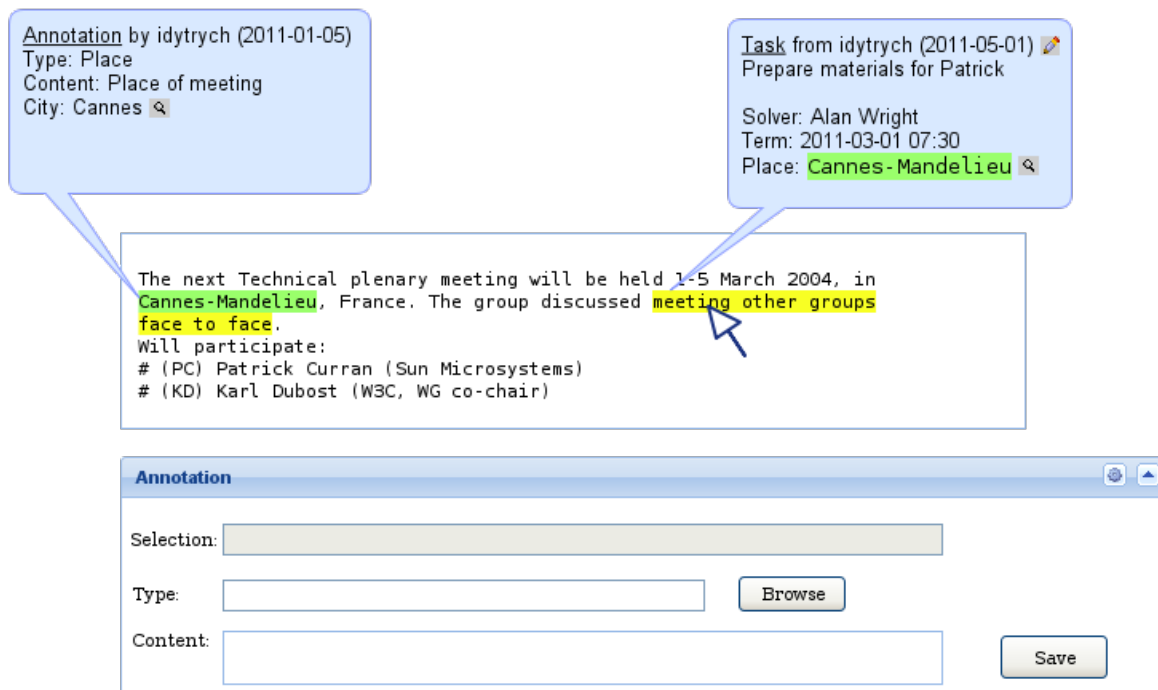


Obrázek 7: Vnořená anotace

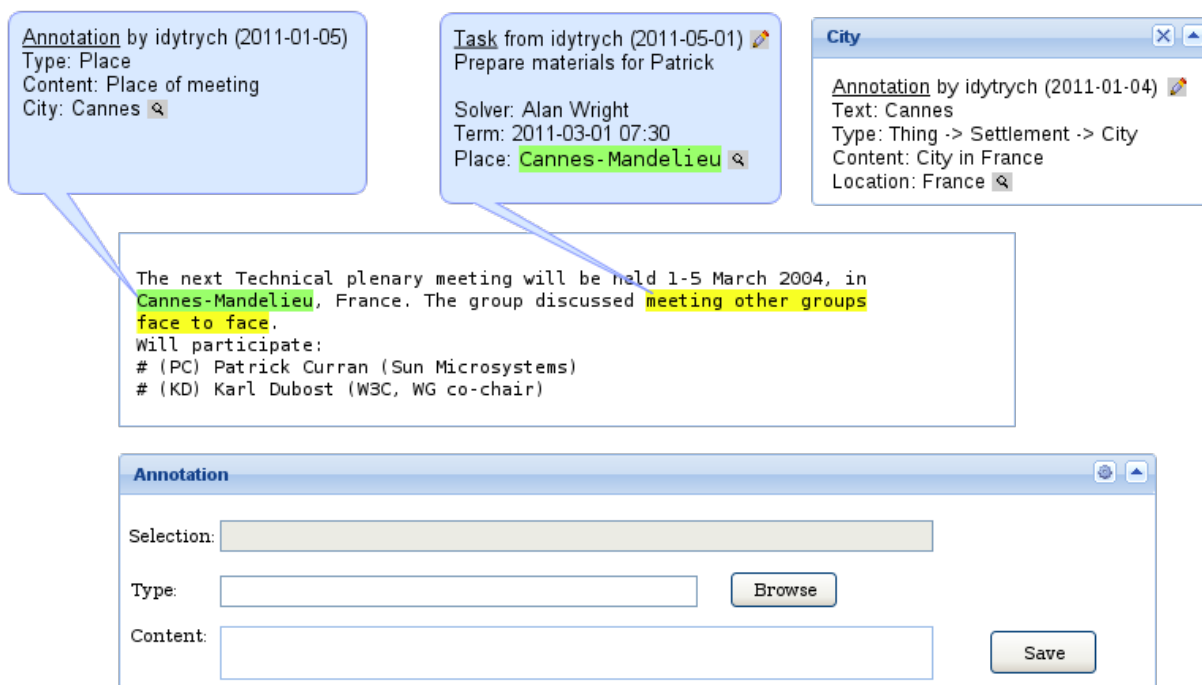
Máme-li vloženu anotaci, ve které je vnořena jiná anotace, je třeba zvolit vhodný způsob vizualizace. Pokud uživatel najede myší na anotovaný fragment, zobrazí se mu nejenom anotace k tomuto fragmentu, ale také anotace, které jsou na první úrovni vnoření (viz obrázek 8). Je-li v některé vnořené anotaci další vnořená anotace, je zobrazen pouze název atributu a část anotovaného fragmentu či textového obsahu anotace, u kterého je ikonka pro zobrazení celé anotace. Pokud uživatel na anotovaný fragment klikne, anotace zůstanou zobrazeny (jako u jednoduché anotace) a je možné na danou ikonku kliknout.

Anotace druhé a vyšší úrovně vnoření se pro zachování přehlednosti nezobrazují v bublinách, ale v oknech jako na obrázku 9.

Ikonka pro zobrazení celé anotace by pro jednotnost zobrazení měla být i u atributu s anotací první úrovně vnoření. Při kliknutí na tuto ikonku se potom i anotace první úrovně vnoření zobrazí v okně (bude tedy zobrazena dvakrát, nebo se bublinová varianta skryje).

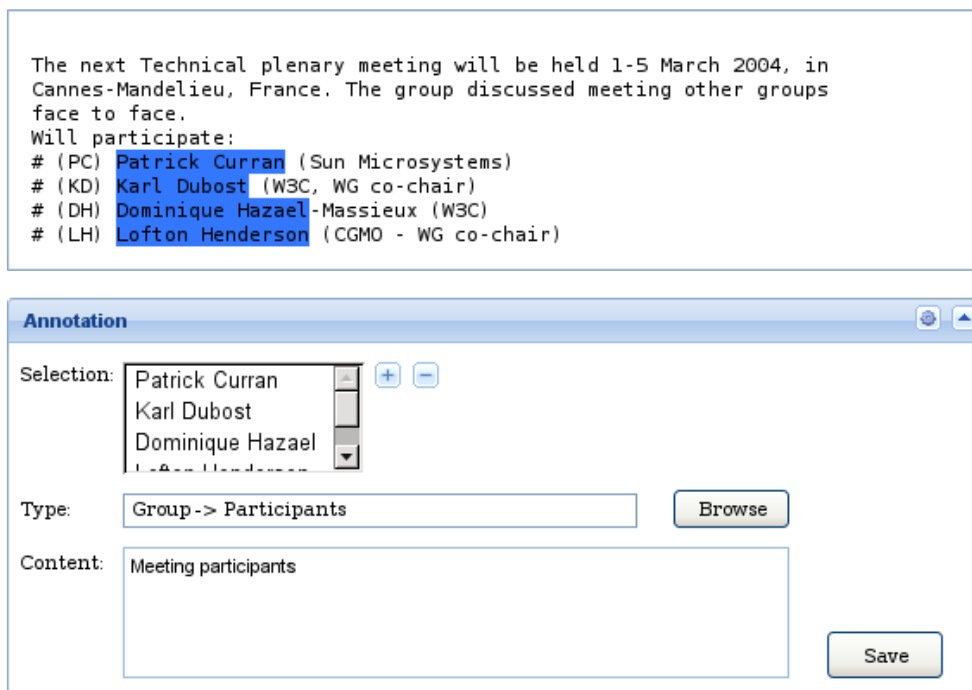


Obrázek 8: Zobrazení anotace s vnořenou anotací



Obrázek 9: Zobrazení anotace druhé a vyšší úrovně vnoření

Ve stránce lze vybrat i více fragmentů textu. V tomto případě dojde k automatickému přepnutí do režimu výběru a současného anotování více fragmentů, který je na obrázku 10. Lze tak např. vyznačit všechna města v dokumentu apod. Při uložení anotace se potom vygeneruje pro každý fragment jedna anotace, jako by byly zadávány jednotlivě. Všechny tyto anotace budou mít stejný typ, obsah i atributy.



Obrázek 10: Režim výběru a současného anotování více fragmentů

Výběr více fragmentů lze využít i při vytváření atributů. Potom bude vytvořeno více atributů (vnořených anotací) se stejným názvem, typem a obsahem (a naopak: je-li v anotaci více atributů se stejným názvem, typem a obsahem, zobrazí se způsobem, který je na obrázku 11).

Na obrázku 12 je příklad nabízení anotací. Uživatel si někde (v menu editoru či doplňku) zapne režim nabízení anotací. Doplněk požádá server o anotace a následně zobrazí nabízené anotace uživateli (odlišně od zadaných či potvrzených). Při najetí myši nad text s nabídkou se anotace zobrazí. Pokud na text uživatel klikne, anotace zůstane zobrazena a jiná se nezobrazí až do chvíle, než danou anotaci zavře (potvrdí, zvolí pro úpravu či odstraní z nabídky) nebo klikne na jiný text.

Anotace je možné nabízet nejenom k celému dokumentu, ale i k vybranému fragmentu. Uživatel si také může zvolit typ anotací, které mu budou nabízeny.

The next **Technical plenary meeting** will be held 1-5 March 2004, in Cannes-Mandelieu, France. The group discussed meeting other groups face to face.
 Will participate:
 # (PC) **Patrick Curran** (Sun Microsystems)
 # (KD) **Karl Dubost** (W3C, WG co-chair)

Annotation

Selection:

Type:

Content:

Attributes

- [-] Solver
- [-] Term
- [+] Attendees

Type:

Selection:

Content:

Obrázek 11: Vytvoření sady atributů výběrem více fragmentů

A proposal of annotation ✓ ✎ ✕

Type: Thing -> Country
 Content: Country in Europe

The next Technical plenary meeting will be held 1-5 March 2004, in Cannes-Mandelieu, **France**. The group discussed meeting other groups face to face.
 Will participate:
 # (PC) Patrick Curran (Sun Microsystems)
 # (KD) Karl Dubost (W3C, WG co-chair)

Annotation

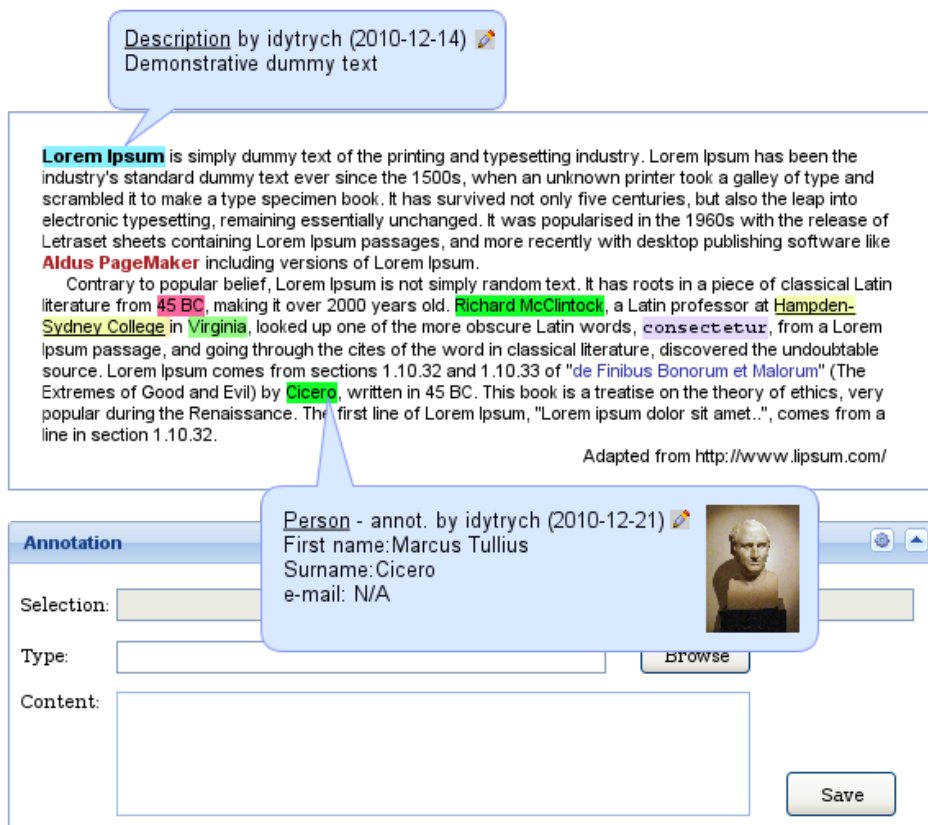
Selection:

Type:

Content:

Obrázek 12: Nabízení anotací

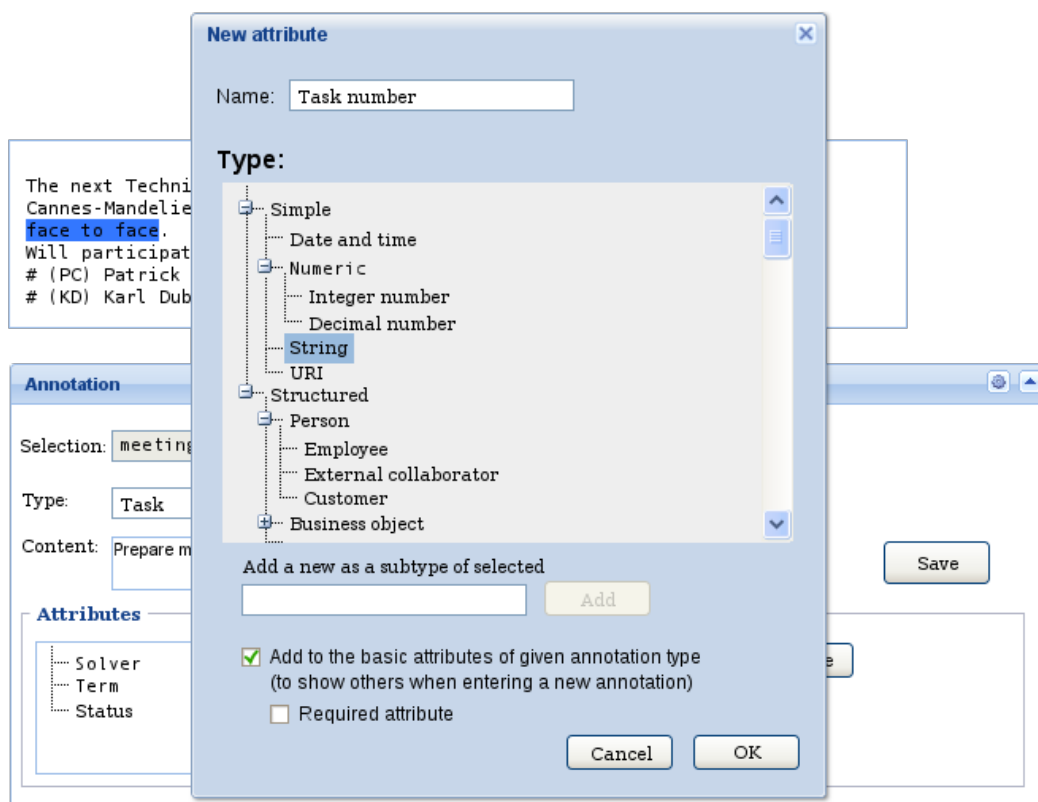
Pokud je v textu více typů anotací, je vhodné, aby byly zobrazeny různým způsobem. Uživatel by měl mít možnost volby zobrazení (barev, fontů apod.) pro jednotlivé konkrétní typy anotací a pro několik neurčených typů (např. osoby zeleně, organizace podtržené, další zobrazené typy potom červeně, modře, fialově, ...). Pokud se jedná o osobu, může se ze serveru načíst i fotografie, jako na obrázku 13.



Obrázek 13: Zobrazení více typů anotací

Na předchozích obrázcích se strukturovanými anotacemi byla ve stromě atributů zobrazena ikonka pro přidání atributu. Při kliknutí na tuto ikonku se zobrazí dialogové okno, které je na obrázku 14. V tomto okně uživatel zadá název, vybere typ a případně odškrtně zaškrťovací políčko pro přidání atributu mezi výchozí atributy daného typu anotace. Toto zaškrťovací políčko rozlišuje trvalé (zobrazované) atributy typu od tzv. adhoc (pro tento případ) atributů. Pokud je políčko zaškrtnuté, s uložením anotace dojde k aktualizaci typu anotace, při které přibude nový atribut. Ten potom server bude u tohoto typu zobrazovat všem. Pokud políčko zaškrtnuté není, atribut bude uložen pouze v této anotaci, ale typ modifikován nebude. V případě aktualizace typu anotace lze zvolit i to, zda bude nový atribut povinný (server bude vždy vyžadovat jeho vyplnění), nebo volitelný.

Obrázek dialogu pro nastavení doplňku jsem nevytvářel, protože tento závisí na konkrétním doplňku. Nastavení každého doplňku však mimo jiné musí obsahovat editovatelnou tabulku parametrů, kde každý parametr bude mít název a řetězcovou hodnotu. V této tabulce budou zobrazeny parametry nastavení načtené ze serveru, pro jejichž nastavení není k dispozici jiný formulář či prvek uživatelského rozhraní.



Obrázek 14: Přidání nového atributu

5.6 Využití anotací pro workflow

Jako ukázkou využitelnosti navrženého systému jsem zvolil využití pro workflow. K tomuto účelu je třeba stanovit minimální strukturu anotace, která bude sloužit k definici úkolu. Dále je třeba řešit podporu na straně klienta a podporu na straně serveru.

Anotace, která ponese definici úkolu, musí obsahovat informace o:

- zadavateli,
- datu zadání úkolu,
- termínu splnění úkolu,
- řešiteli či řešitelích,
- specifikaci úkolu,
- stavu rozpracování.

Informace o zadavateli úkolu a jeho zadání jsou v anotaci implicitně obsaženy. Termín lze vložit jako atribut s hodnotou primitivního datového typu nebo jako vnořenou anotaci fragmentu textu, ve kterém je termín uveden. Informace o řešitelích lze vložit jako atributy s příslušným názvem. Hodnoty těchto atributů mohou být řetězce se jmény osob či vnořené anotace fragmentů textu se jmény řešitelů. Specifikace úkolu sestává z textového popisu úkolu, který bude umístěn v textovém obsahu anotace, a z dalších atributů, které specifikují strukturované (a tím i plně či částečně strojově zpracovatelné) informace o úkolu. Informace o stavu rozpracování může být určována automaticky na serveru a v anotaci pouze přenášena v atributu. Uživatel ji tedy může, ale také nemusí měnit.

Podpora workflow na straně klienta není vyžadována. Veškerá potřebná funkcionalita je obsažena již v základní variantě klienta s podporou strukturovaných anotací. Pokud bychom na stranu klienta zabudovali přidanou podporu workflow, klient by mohl automaticky zjišťovat nové úkoly, nabízet k přečtení dokumenty se specifikacemi úkolů apod. Pokud však tato funkcionalita nebude dostupná, lze ji nahradit např. tím, že server bude úkoly a jejich změny posílat e-mailem. Při zasílání e-mailů je však nutné zvolit vhodný způsob filtrování změn, aby uživatelé nebyli zahlceni e-maily. Dle [13] je navíc vhodné, aby v e-mailech o změnách byla vždy uvedena i informace o tom, kdo danou změnu provedl.

Podpora workflow na straně serveru je nezbytnou součástí řešení. Hlavní část práce s anotacemi pro workflow musí být umístěna na serveru. Pokud by totiž server workflow nepodporoval, bylo by sice možné specifikovat úkol, ale bylo by na něj vždy nutné upozornit další osobu, která na něm má pracovat. Vše by tedy bylo řešeno manuálně. Nejdůležitější činností serveru s podporou workflow je tedy distribuce úkolů. Pokud do systému přibude nový úkol, server musí vyhodnotit, zda má přiřazeného konkrétního řešitele, a pokud nemá, řešitele přiřadit. Následně je nutné řešitele na úkol upozornit. Po tom, co řešitel svoji práci provede, je nutné upozornit další osobu, která má na úkolů pracovat, nebo zadavatele úkolu či jinou zainteresovanou osobu (v případě problému či dokončení úkolu).

Čím složitější podporu workflow server poskytne, tím více operací bude možné provádět automaticky. Pokud server poskytne dostatečnou funkcionalitu, bude možné kompletní workflow podniku nadefinovat pomocí některého z obecných jazyků pro definici workflow a server bude následně řídit veškeré operace.

6 Budoucí vývoj

V této kapitole popíši budoucí vývoj svojí disertační práce. Rozdělil jsem jej na tři oblasti, přičemž první se zabývá technickým hlediskem, tedy tím, co ještě plánuji navrhnout a implementovat. Druhá oblast se týká publikování výsledků práce a třetí spolupráce se zahraničními pracovišti a zapojení do větších výzkumných projektů.

6.1 Technický vývoj

Technický vývoj je již popsán výše, zde tedy uvedu pouze stručné shrnutí. Nejprve implementuji navržený jednoduchý server pro práci s anotacemi, jehož funkcionalitu ověřím s doplňkem do textového editoru TinyMCE, který vytvoří student v rámci mnou vedené diplomové práce. Následně navrhnu složitější verzi serveru, která bude poskytovat pokročilé nabízení anotací, přičemž k tomuto účelu využiji napojení na některý z existujících projektů pro extrakci informací z textu a automatické anotování. Server také umožní využití zásuvných modulů pro pokročilou práci s anotacemi. Jako zásuvný modul potom vytvořím modul pro workflow s anotacemi. Při ověření využitelnosti pokročilejšího serveru shromáždím požadavky na spolupráci v distribuovaném prostředí a navrhnu protokol pro komunikaci mezi servery. Následně navrhnu a implementuji výslednou variantu serveru, kterou bude možné nasadit v reálném provozu.

6.2 Plánovaná publikační činnost

V blízké době plánuji publikování výsledků svojí práce na různých seminářích a konferencích. Mezi tyto bude patřit např. ESWC (Extended Semantic Web Conference) a ISWC (International Semantic Web Conference). Na seminář v rámci konference ESWC bych chtěl zaslat svůj příspěvek už v letošním roce.

6.3 Plánovaná zahraniční spolupráce

V rámci disertační práce plánuji spolupráci s výzkumným ústavem Salzburg Research v Rakousku. V blízké době bych chtěl napojit svoji práci na projekt IKS, v rámci kterého vzniká FISE (Furtwangen IKS Semantic Engine). FISE slouží k rozšíření textového obsahu o sémantické informace, které by mohly být využity při nabízení anotací.

7 Závěr

Nejprve jsem prostudoval stav poznání v oblasti tématu práce a související oblasti. Následně jsem si vytyčil cíle práce a zvolil metody řešení. Navrhl jsem také formát anotace, protokol pro komunikaci mezi klientem a serverem, jednoduchý server pro práci s anotacemi a vzhled a chování klientů.

První z klientů je nyní implementován v rámci mnou vedené diplomové práce, která se zabývá vytvářením anotacího doplňku do textových editorů napsaných v jazyce JavaScript. Současně implementuji navržený jednoduchý server pro práci s anotacemi a připravuji se na připojení k zahraničnímu projektu IKS. Plánuji také brzké publikování na jednom z výše uvedených seminářů.

Dále budu v průběhu své disertační práce pokračovat návrhem a implementací složitější varianty serveru pro práci s anotacemi, která bude disponovat nabízením anotací se strojovým učením a možností využití zásuvných modulů pro pokročilou práci s anotacemi a následné využití anotací k různým účelům. Následně navrhnu protokol pro komunikaci mezi servery a vytvořím finální variantu serveru určenou do distribuovaného prostředí.

Literatura

- [1] Agosti, M.; Albrechtsen, H.; Ferro, N.; aj.: DiLAS: a Digital Library Annotation Service. In *Proceedings of Annotation for Collaboration – A Workshop on Annotation Models, Tools and Practices*, 2005, [Online; navštíveno 19.1.2011].
URL http://www.is.informatik.uni-duisburg.de/bib/pdf/ir/Agosti_etal:05.pdf
- [2] Agosti, M.; Bonfiglio-Dosio, G.; Ferro, N.: A historical and contemporary study on annotations to derive key features for systems design. *International Journal on Digital Libraries*, ročník 8, 2007: s. 1–19, ISSN 1432-5012, [Online; navštíveno 19.1.2011].
URL <http://www.springerlink.com/content/93j367635270641h/fulltext.pdf>
- [3] Agosti, M.; Ferro, N.: An Information Service Architecture for Annotations. In *Pre-proceedings of the 6th Thematic Workshop of the EU Network of Excellence DELOS*, 2004, s. 115–126, [Online; navštíveno 19.1.2011].
URL http://delos-old.isti.cnr.it/eventlist/dla_04_preproceedings.pdf#page=120
- [4] Agosti, M.; Ferro, N.: Annotations as Context for Searching Documents. In *Information Context: Nature, Impact, and Role, Lecture Notes in Computer Science*, ročník 3507, editace F. Crestani; I. Ruthven, Springer Berlin / Heidelberg, 2005, s. 155–170, [Online; navštíveno 19.1.2011].
URL <http://www.springerlink.com/content/he1rc9v2np17a6v3/fulltext.pdf>
- [5] Agosti, M.; Ferro, N.: A System Architecture as a Support to a Flexible Annotation Service. In *Peer-to-Peer, Grid, and Service-Oriented in Digital Library Architectures, Lecture Notes in Computer Science*, ročník 3664, editace C. Türker; M. Agosti; H.-J. Schek, Springer Berlin / Heidelberg, 2005, s. 147–166, [Online; navštíveno 19.1.2011].
URL <http://www.springerlink.com/content/90xvqnnwajaktd7l/fulltext.pdf>
- [6] Agosti, M.; Ferro, N.: Search Strategies for Finding Annotations and Annotated Documents: The FAST Service. In *Flexible Query Answering Systems, Lecture Notes in Computer Science*, ročník 4027, editace H. Larsen; G. Pasi; D. Ortiz-Arroyo; T. Andreasen; H. Christiansen, Springer Berlin / Heidelberg, 2006, s. 270–281, [Online; navštíveno 19.1.2011].
URL <http://www.springerlink.com/content/p701258v346k80n2/fulltext.pdf>
- [7] Agosti, M.; Ferro, N.: A formal model of annotations of digital content. *ACM Trans. Inf. Syst.*, ročník 26, Listopad 2007, ISSN 1046-8188, [Online; navštíveno 17.1.2011].
URL http://portal.acm.org/ft_gateway.cfm?id=1292594&type=pdf&CFID=6442261&CFTOKEN=24665493
- [8] Bottoni, P.; Levialdi, S.; Rizzo, P.: An Analysis and Case Study of Digital Annotation. In *Databases in Networked Information Systems, Lecture Notes in Computer Science*, ročník 2822, editace N. Bianchi-Berthouze, Springer Berlin / Heidelberg, 2003, s. 216–231, [Online; navštíveno 19.1.2011].
URL <http://www.springerlink.com/content/u1cu2mx7cexjf01h/fulltext.pdf>
- [9] Bremer, M.; Gertz, M.: Web Data Indexing Through External Semantic-carrying Annotations. In *Eleventh International Workshop on Research Issues in Data Engineering on Document Management for Data Intensive Business and Scientific Applications*, Washington, DC, USA: IEEE Computer Society, 2001, ISBN 0-7695-0957-6, s. 69–76, [Online; navštíveno 17.1.2011].
URL <http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=00916493>
- [10] Brickley, D.: Basic Geo (WGS84 lat/long) Vocabulary. W3C Semantic Web Interest Group, 2004, [Online; navštíveno 2.1.2011].
URL <http://www.w3.org/2003/01/geo/>
- [11] Brush, A. J. B.; Barger, D.; Gupta, A.; aj.: Robust annotation positioning in digital documents. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '01, New York, NY, USA: ACM, 2001, ISBN 1-58113-327-8, s. 285–292, [Online; navštíveno 17.1.2011].
URL http://portal.acm.org/ft_gateway.cfm?id=365117&type=pdf&CFID=6442261&CFTOKEN=24665493
- [12] Bry, F.; Kotowski, J.: A Social Vision of Knowledge Representation and Reasoning. In *SOFSEM 2010: Theory and Practice of Computer Science, Lecture Notes in Computer Science*, ročník 5901, editace J. van Leeuwen; A. Muscholl; D. Peleg; J. Pokorný; B. Rumpe, Springer Berlin / Heidelberg, 2010, s. 235–246.

- [13] Cadiz, J. J.; Gupta, A.; Grudin, J.: Using Web annotations for asynchronous collaboration around documents. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, CSCW '00, New York, NY, USA: ACM, 2000, ISBN 1-58113-222-0, s. 309–318, [Online; navštíveno 19.1.2011].
URL http://portal.acm.org/ft_gateway.cfm?id=359002&type=pdf&CFID=6215385&CFTOKEN=18093147
- [14] Constantopoulos, P.; Doerr, M.; Theodoridou, M.; aj.: On Information Organization in Annotation Systems. In *Intuitive Human Interfaces for Organizing and Accessing Intellectual Assets, Lecture Notes in Computer Science*, ročník 3359, editace G. Grieser; Y. Tanaka, Springer Berlin / Heidelberg, 2005, s. 189–200, [Online; navštíveno 19.1.2011].
URL <http://www.springerlink.com/content/7gh22qnha8uu77a8/fulltext.pdf>
- [15] Crocodoc. Crocodoc, 2010, [Online; navštíveno 29.12.2010].
URL <http://crocodoc.com/>
- [16] Désilets, A.; Gonzalez, L.; Paquet, S.; aj.: Translation the Wiki Way. In *Proceedings of the 2006 international symposium on Wikis*, WikiSym '06, New York, NY, USA: ACM, Srpen 2006, ISBN 1-59593-413-8, s. 19–32, [Online; navštíveno 17.1.2011].
URL <http://www.wikisym.org/ws2006/proceedings/p19.pdf>
- [17] What is screen scraping? Deskperience, 2011, [Online; navštíveno 6.1.2011].
URL http://wiki.screen-scraping.deskperience.com/wiki/What_is_screen_scraping%3F
- [18] Feigenbaum, L.: Standardize annotations with Web services. IBM, Duben 2004, [Online; navštíveno 17. 12. 2010].
URL <http://www.ibm.com/developerworks/webservices/library/ws-annotation.html>
- [19] Flickr. Yahoo! Inc., 2011, [Online; navštíveno 2.1.2011].
URL <http://www.flickr.com/>
- [20] Garrett, J. J.: Ajax: A New Approach to Web Applications. Adaptive Path Inc., Únor 2005, [Online; navštíveno 28.12.2010].
URL <http://www.adaptivepath.com/ideas/essays/archives/000385.php>
- [21] Google Wave. Google, 2010, [Online; navštíveno 28. 12. 2010].
URL <http://wave.google.com/about.html>
- [22] Google Wave Protocol. Google, 2010, [Online; navštíveno 28. 12. 2010].
URL <http://www.waveprotocol.org/>
- [23] FISE. IKS Project, 2010, [Online; navštíveno 6.1.2011].
URL <http://wiki.iks-project.eu/index.php/FISE>
- [24] Kahan, J.; Koivunen, M.-R.: Annotea: an open RDF infrastructure for shared Web annotations. In *Proceedings of the 10th international conference on World Wide Web*, WWW '01, New York, NY, USA: ACM, 2001, ISBN 1-58113-348-0, s. 623–632, [Online; navštíveno 19.1.2011].
URL http://portal.acm.org/ft_gateway.cfm?id=372166&type=pdf&CFID=6211761&CFTOKEN=37218305
- [25] Koivunen, M.-R.; Swick, R. R.: Metadata Based Annotation Infrastructure offers Flexibility and Extensibility for Collaborative Applications and Beyond. In *K-CAP 2001 Workshop October 21, 2001, Victoria B.C., Canada*, Calgary, Alberta Canada T2N 1N4: Department of Computer Science, University of Calgary, Říjen 2001, [Online; navštíveno 19.1.2011].
URL http://semannot2001.aifb.uni-karlsruhe.de/papers/1_annotea.pdf
- [26] LaLiberte, D.; Braverman, A.: A protocol for scalable group and public annotations. *Comput. Netw. ISDN Syst.*, ročník 27, Duben 1995: s. 911–918, ISSN 0169-7552, [Online; navštíveno 19.1.2011].
URL http://www.sciencedirect.com/science?_ob=MIimg&_imagekey=B6TYT-3YGTSP5-1X-1&_cdi=5627&_user=640830&_pii=0169755295000172&_origin=search&_coverDate=04%2F30%2F1995&_sk=999729993&view=c&wchp=dGLzV1b-zSkzV&md5=b3bd6af4ba2b8af53be56024dcf9841e&ie=/sdarticle.pdf
- [27] ISO 639-2 Language Code List - Codes for the representation of names of languages. Library of Congress, Říjen 2010, [Online; navštíveno 2.1.2011].
URL http://www.loc.gov/standards/iso639-2/php/code_list.php

- [28] Marshall, C. C.: Annotation: from paper books to the digital library. In *Proceedings of the second ACM international conference on Digital libraries*, DL '97, New York, NY, USA: ACM, 1997, ISBN 0-89791-868-1, s. 131–140, [Online; navštíveno 19.1.2011].
URL http://portal.acm.org/ft_gateway.cfm?id=263806&type=pdf&CFID=6215385&CFTOKEN=18093147
- [29] Marshall, C. C.: Toward an ecology of hypertext annotation. In *Proceedings of the ninth ACM conference on Hypertext and hypermedia : links, objects, time and space—structure in hypermedia systems: links, objects, time and space—structure in hypermedia systems*, HYPERTEXT '98, New York, NY, USA: ACM, 1998, ISBN 0-89791-972-6, s. 40–49, [Online; navštíveno 19.1.2011].
URL http://portal.acm.org/ft_gateway.cfm?id=276632&type=pdf&CFID=6215385&CFTOKEN=18093147
- [30] Marshall, C. C.; Brush, A. J. B.: Exploring the relationship between personal and public annotations. In *Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries*, JCDL '04, New York, NY, USA: ACM, 2004, ISBN 1-58113-832-6, s. 349–357, [Online; navštíveno 19.1.2011].
URL http://portal.acm.org/ft_gateway.cfm?id=996432&type=pdf&CFID=6215385&CFTOKEN=18093147
- [31] MediaWiki. nadace WIKIMEDIA, 2010, [Online; navštíveno 28.12.2010].
URL <http://www.mediawiki.org/wiki/MediaWiki>
- [32] TinyMCE. Moxiecode Systems AB, 2011, [Online; navštíveno 2.1.2011].
URL <http://tinymce.moxiecode.com/>
- [33] Annozilla (Annotea on Mozilla). Mozdev Community Organization, 2010, [Online; navštíveno 29.12.2010].
URL <http://annozilla.mozdev.org/index.html>
- [34] Neuwirth, C. M.; Kaufer, D. S.; Chandhok, R.; aj.: Computer support for distributed collaborative writing: defining parameters of interaction. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, CSCW '94, New York, NY, USA: ACM, 1994, ISBN 0-89791-689-1, s. 145–152, [Online; navštíveno 19.1.2011].
URL http://portal.acm.org/ft_gateway.cfm?id=192893&type=pdf&CFID=6215385&CFTOKEN=18093147
- [35] Niranatlamphong, W.; Choochaiwattana, W.; Spring, M.: A conceptual framework for digital annotation system on WWW. In *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on*, Srpen 2009, s. 27–31, [Online; navštíveno 17.1.2011].
URL <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&number=5234462>
- [36] Phelps, T.; Wilensky, R.: Multivalent Annotations. In *Research and Advanced Technology for Digital Libraries, Lecture Notes in Computer Science*, ročník 1324, editace C. Peters; C. Thanos, Springer Berlin / Heidelberg, 1997, s. 287–303, [Online; navštíveno 19.1.2011].
URL <http://www.springerlink.com/content/d87v173g7671j20p/fulltext.pdf>
- [37] Phelps, T. A.; Wilensky, R.: Robust Hyperlinks and Locations. *D-Lib Magazine*, ročník 6, Červenec 2000, ISSN 1082-9873, [Online; navštíveno 19.1.2011].
URL <http://www.dlib.org/dlib/july00/wilensky/07wilensky.html>
- [38] Phelps, T. A.; Wilensky, R.: Robust intra-document locations. In *Proceedings of the 9th international World Wide Web conference on Computer networks : the international journal of computer and telecommunications netourking*, Amsterdam, The Netherlands, The Netherlands: North-Holland Publishing Co., Květen 2000, s. 105–118, [Online; navštíveno 19.1.2011].
URL <http://www9.org/w9cdrom/312/312.html>
- [39] Quint, V.: Amaya. W3C, 2010, [Online; navštíveno 19.1.2011].
URL <http://www.w3.org/Amaya/>
- [40] da Rocha, T.; Willrich, R.; Fileto, R.; aj.: Supporting Collaborative Learning Activities with a Digital Library and Annotations. In *Education and Technology for a Better World, IFIP Advances in Information and Communication Technology*, ročník 302, editace A. Tatnall; A. Jones, Springer Boston, 2009, s. 349–358, [Online; navštíveno 19.1.2011].
URL <http://www.springerlink.com/content/gug76w637102x821/fulltext.pdf>
- [41] Russell, A.: Comet: Low Latency Data for the Browser. In *Infrequently Noted*, Březen 2006, [Online; navštíveno 28.12.2010].
URL <http://infrequently.org/2006/03/comet-low-latency-data-for-the-browser/>

- [42] Sannomiya, T.; Amagasa, T.; Yoshikawa, M.; aj.: A framework for sharing personal annotations on web resources using XML. In *Proceedings of the workshop on Information technology for virtual enterprises*, ITVE '01, Washington, DC, USA: IEEE Computer Society, 2001, ISBN 0-7695-0960-6, s. 40–48, [Online; navštíveno 19.1.2011].
URL http://portal.acm.org/ft_gateway.cfm?id=545625&type=pdf&CFID=6206112&CFTOKEN=33710793
- [43] Schaffert, S.: IkeWiki: A Semantic Wiki for Collaborative Knowledge Management. In *Enabling Technologies: Infrastructure for Collaborative Enterprises, 2006. WETICE '06. 15th IEEE International Workshops on*, Červen 2006, ISSN 1524-4547, s. 388–396, [Online; navštíveno 17.1.2011].
URL http://portal.acm.org/ft_gateway.cfm?id=1135863&type=pdf&CFID=6442261&CFTOKEN=24665493
- [44] Schaffert, S.: Semantic social software – semantically enabled social software or socially enabled semantic web. In *Semantics 2006*, 2006, [Online; navštíveno 17.1.2011].
URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.107.8638&rep=rep1&type=pdf>
- [45] Schaffert, S.; Eder, J.; Grünwald, S.; aj.: KiWi – A Platform for Semantic Social Software (Demonstration). In *The Semantic Web: Research and Applications, Lecture Notes in Computer Science*, ročník 5554, editace L. Aroyo; P. Traverso; F. Ciravegna; P. Cimiano; T. Heath; E. Hyvönen; R. Mizoguchi; E. Oren; M. Sabou; E. Simperl, Springer Berlin / Heidelberg, 2009, s. 888–892, [Online; navštíveno 17.1.2011].
URL <http://www.springerlink.com/content/351683094p8m1862/fulltext.pdf>
- [46] Schaffert, S.; Gruber, A.; Westenthaler, R.: A Semantic Wiki for Collaborative Knowledge Formation. In *Proceedings of SEMANTICS 2005 Conference*, Vienna, Austria, 2006, [Online; navštíveno 6.1.2011].
URL http://wworld.salzburgresearch.at/research/gfx/SemWikiForCollKnowForm_20060120.pdf
- [47] Schickler, M. A.; Mazer, M. S.; Brooks, C.: Pan-browser support for annotations and other meta-information on the World Wide Web. In *Proceedings of the fifth international World Wide Web conference on Computer networks and ISDN systems*, Amsterdam, The Netherlands, The Netherlands: Elsevier Science Publishers B. V., 1996, s. 1063–1074, [Online; navštíveno 19.1.2011].
URL http://www.sciencedirect.com/science?_ob=MIimg&_imagekey=B6TYT-3XY8PKJ-H-1&_cdi=5627&_user=640830&_pii=0169755296000608&_origin=search&_coverDate=05%2F31%2F1996&_sk=999719992&view=c&wchp=dGLbVzz-zSkzk&md5=cf6aecc25031c591511c241053ef72bf&ie=/sdarticle.pdf
- [48] ScraperWiki. ScraperWiki Ltd, 2011, [Online; navštíveno 6.1.2011].
URL <http://scraperwiki.com/>
- [49] Semantic MediaWiki. In *Semantic MediaWiki*, semantic-mediawiki.org, 2010, [Online; navštíveno 28.12.2010].
URL http://semantic-mediawiki.org/wiki/Semantic_MediaWiki
- [50] SharedCopy. SharedCopy, 2010, [Online; navštíveno 29.12.2010].
URL <http://sharedcopy.com/>
- [51] ShiftSpace. ShiftSpace, 2010, [Online; navštíveno 29.12.2010].
URL <http://www.shiftspace.org/>
- [52] Specia, L.; Motta, E.: Integrating Folksonomies with the Semantic Web. In *The Semantic Web: Research and Applications, Lecture Notes in Computer Science*, ročník 4519, editace E. Franconi; M. Kifer; W. May, Springer Berlin / Heidelberg, 2007, s. 624–639, [Online; navštíveno 17.1.2011].
URL <http://www.springerlink.com/content/413285327hj53234/fulltext.pdf>
- [53] Stickis. Stickis, 2010, [Online; navštíveno 14.12.2010].
URL <http://stickis.com/>
- [54] Twitter. Twitter, 2011, [Online; navštíveno 2.1.2011].
URL <http://twitter.com/>
- [55] Vaishar, A.: *Folksonomie*. Diplomová práce, ÚK ÚZ FF MU v Brně, 2007, [Online; navštíveno 28. 12. 2010].
URL http://is.muni.cz/th/4157/ff_b/Folksonomie.pdf
- [56] Völkel, M.; Krötzsch, M.; Vrandečić, D.; aj.: Semantic Wikipedia. In *Proceedings of the 15th international conference on World Wide Web*, WWW '06, New York, NY, USA: ACM, 2006, ISBN 1-59593-323-9, s. 585–594, [Online; navštíveno 17.1.2011].
URL http://portal.acm.org/ft_gateway.cfm?id=1135863&type=pdf&CFID=6442261&CFTOKEN=24665493

- [57] Völkel, M.; Haller, H.: Conceptual Data Structures (CDS) – Towards an Ontology for Semi-Formal Articulation of Personal Knowledge. In *14th International Conference on Conceptual Structures, Aalborg University, Denmark, 16.07.06*, 2006, [Online; navštíveno 17. 12. 2010].
URL <http://www.xam.de/2006/04-cds.pdf>
- [58] Minutes of QAWG/IG f2f meeting, 21-23 October 2003, Boulder, Colorado, USA. Quality Assurance at W3C, 2003, [Online; navštíveno 5.1.2011].
URL <http://www.w3.org/QA/2003/10/f2f-minutes>
- [59] Annotea Project. W3C, 2010, [Online; navštíveno 29.12.2010].
URL <http://www.w3.org/2001/Annotea/>
- [60] XML Schema Tutorial. Refsnes Data, 2011, [Online; navštíveno 2.1.2011].
URL <http://www.w3schools.com/Schema/default.asp>
- [61] Weng, C.; Gennari, J. H.: Asynchronous collaborative writing through annotations. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work, CSCW '04*, New York, NY, USA: ACM, 2004, ISBN 1-58113-810-5, s. 578–581, [Online; navštíveno 19.1.2011].
URL http://portal.acm.org/ft_gateway.cfm?id=1031705&type=pdf&CFID=6206112&CFTOKEN=33710793
- [62] Semantic wiki. In *Wikipedia – The Free Encyclopedia*, nadace WIKIMEDIA, 2010, [Online; navštíveno 28.12.2010].
URL http://en.wikipedia.org/wiki/Semantic_wiki
- [63] Data scraping. In *Wikipedia – The Free Encyclopedia*, nadace WIKIMEDIA, 2011, [Online; navštíveno 6.1.2011].
URL http://en.wikipedia.org/wiki/Data_scraping
- [64] Wiki. In *Wikipedie – Otevřená encyklopedie*, nadace WIKIMEDIA, 2010, [Online; navštíveno 28.12.2010].
URL <http://cs.wikipedia.org/wiki/Wiki>
- [65] Wojahn, P. G.; Neuwirth, C. M.; Bullock, B.: Effects of interfaces for annotation on communication in a collaborative task. In *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '98*, New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 1998, ISBN 0-201-30987-4, s. 456–463, [Online; navštíveno 19.1.2011].
URL http://portal.acm.org/ft_gateway.cfm?id=274706&type=pdf&CFID=6215385&CFTOKEN=18093147
- [66] Wright, J.; Andriesian, D.; Wright, R.; aj.: MyStickies. MyStickies, 2010, [Online; navštíveno 29.12.2010].
URL <http://www.mystickies.com/>
- [67] Protocol Stack. XMPP Standards Foundation, 2010, [Online; navštíveno 28.12.2010].
URL <http://xmpp.org/xmpp-protocols/xmpp-core/>
- [68] Zheng, Q.: *Structured Annotations to Support Collaborative Writing Workflow*. Diplomová práce, The Faculty of Graduate Studies (Computer Science), The University of British Columbia, Prosinec 2005, [Online; navštíveno 17. 12. 2010].
URL https://circle.ubc.ca/bitstream/handle/2429/17720/ubc_2006-0141.pdf?sequence=1
- [69] Zheng, Q.; Booth, K.; McGrenere, J.: Co-Authoring with Structured Annotations. Department of Computer Science, University of British Columbia, 2006, [Online; navštíveno 17. 12. 2010].
URL <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.126.2558&rep=rep1&type=pdf>
- [70] Zhou, D.; Bian, J.; Zheng, S.; aj.: Exploring social annotations for information retrieval. In *Proceeding of the 17th international conference on World Wide Web, WWW '08*, New York, NY, USA: ACM, 2008, ISBN 978-1-60558-085-2, s. 715–724, [Online; navštíveno 19.1.2011].
URL http://portal.acm.org/ft_gateway.cfm?id=1367594&type=pdf&CFID=6215385&CFTOKEN=18093147

Přílohy

Seznam příloh

Seznam příloh	34
A Ukázka strukturované anotace	35
B Protokol pro přenos anotací mezi klientem a serverem	36
B.1 Správa sezení	36
B.2 Uživatelé a skupiny	36
B.3 Řízení odběru anotací	37
B.4 Synchronizace dokumentu	38
B.5 Přenos typů anotací	39
B.6 Přenos anotací	39
B.7 Nabízení anotací	40
B.8 Přenos nastavení	41
B.9 Chyby a varování	41
B.10 Potvrzení bez doplňujících dat	44
C Zjednodušený příklad komunikace mezi klientem a serverem	45

A Ukázka strukturované anotace

```
<rdf:Description rdf:about="http://example.com/annotations/123456">
  <rdf:type rdf:resource="http://example.com/types/g01/annotation/task"/>
  <a:dateTime rdf:value="2011-01-01T:20:00:00Z" />
  <a:author id="http://example.com/authors/123456" name="Jaroslav Dytrych"
    address="idytrych@fit.vutbr.cz"/>
  <a:source rdf:resource="http://example.com/documents/getDoc?id=123456"/>
  <a:fragment>
    <a:path>/html/body/div[@id='container']/div[@id='main']/div[@id='post1']/DIV[2]/p[1]</a:path>
    <a:offset>22</a:offset>
    <a:length>32</a:length>
    <a:annotatedText>Fakulta informačních technologií</a:annotatedText>
  </a:fragment>
  <a:content>
    <![CDATA[
      ...
    ]]>
  </a:content>
  <a:attribute name="place" type="geoPoint">
    <geo:Point> <geo:lat>55.701</geo:lat> <geo:long>12.552</geo:long> </geo:Point>
  </a:attribute>
  <a:attribute name="date" type="nestedAnnotation">
    <rdf:Description rdf:about="http://example.com/annotations/123457">
      <rdf:type rdf:resource="http://example.com/types/g01/annotation/description"/>
      <a:dateTime rdf:value="2011-01-01T:20:00:00Z" />
      <a:author id="http://example.com/authors/123456" name="Jaroslav Dytrych"
        address="idytrych@fit.vutbr.cz"/>
      <a:source rdf:resource="http://example.com/documents/getDoc?id=123456"/>
      <a:fragment>
        <a:path>/html/body/div[@id='container']/div[@id='main']/div[@id='post1']/p[1]</a:path>
        <a:offset>92</a:offset>
        <a:length>14</a:length>
        <a:annotatedText>14. ledna 2011</a:annotatedText>
      </a:fragment>
      <a:content>
        <![CDATA[
          ...
        ]]>
      </a:content>
      <a:attribute name="date" type="DateTime" rdf:value="2011-01-14T:00:00:00Z"/>
    </rdf:Description>
  </a:attribute>
  <a:attribute name="reason" type="annotationLink" uri="http://example.com/annotations/1234567"/>
</rdf:Description>
```

B Protokol pro přenos anotací mezi klientem a serverem

B.1 Správa sezení

Správa sezení zahrnuje dohodu na verzi protokolu, přihlášení a odhlášení uživatele.

Nejprve klient zahájí spojení tak, že zašle na server zprávu `connect`, kde v atributu uvede nejvyšší verzi protokolu, kterou může využít:

```
<connect protocolVersion="1.0"/>
```

Server odpoví chybou nebo následující zprávou:

```
<connected protocolVersion="1.0" sessionID=""/>
```

V atributu `protocolVersion` server uvede verzi protokolu, kterou bude komunikovat. Server by měl komunikovat stejnou verzí jako klient, nebo nejnižší verzí, která je zpětně kompatibilní s verzí klienta. Pokud klient nabídne novější verzi než server, server použije nejnovější verzi, kterou zná. Pokud klient zjistí, že jeho verze není zpětně kompatibilní, musí přepnout na verzi serveru (případně jinou s ní kompatibilní), nebo se odpojit. Pokud verze protokolu není podporována a server vrátí chybovou zprávu, klient se může pokusit o spojení se starší verzí protokolu.

Vzhledem k tomu, že novější verze protokolu může být zpětně kompatibilní, klient a server mohou implementovat různé verze protokolu. Pokud server či klient umožňuje využít novou funkcionalitu, protistrana se starší verzí protokolu ji nevyužije a zaslané elementy či atributy navíc bude ignorovat. Pokud nová verze protokolu nebude zpětně kompatibilní, server či klient, který ji implementuje, musí spojení s danou kombinací verzí protokolu odmítnout. Ukončení spojení je signalizováno následující zprávou:

```
<disconnect/>
```

Ve zprávě `connected` server zašle také id sezení (atribut `sessionID`), které bude klient zasílat se všemi následujícími zprávami v atributu `id` elementu `session`:

```
<session id=""/>
```

Přihlášení je realizováno následující zprávou:

```
<login user="" password=""/> ,
```

kde `user` je uživatelské jméno nebo e-mail uživatele a `password` je heslo uživatele. V případě úspěšného přihlášení server na tuto zprávu odpoví seznamem parametrů nastavení a následující zprávou:

```
<logged id="" name=""/> ,
```

kde `id` je identifikátor uživatele a `name` jeho zobrazované jméno. V případě neúspěšného přihlášení server odpoví chybovou zprávou.

Odhlášení bude prováděno zprávou `<logout/>`.

Přihlášení lze provést současně se zahájením spojení a dohodou na verzi protokolu, odhlášení společně s jeho ukončením.

B.2 Uživatelé a skupiny

Pro získání informací o profilech uživatelů zašle klient zprávu:

```
<queryPersons filter="" withGroups=""/>
```

Ve filtru lze pro výběr pole využít klíčové slovo `id`, `e-mail` či `name` následované dvojtečkou. Při filtrování dle více polí budou filtry odděleny středníkem. Pokud bude uveden i volitelný atribut `withGroups` s hodnotou `true`, budou v informacích o profilech zahrnuty i informace o členství ve skupinách. Server na tuto zprávu odpoví zprávou:

```
<persons>
  <person id="" login="" name="" email="" photoURI=""/>
</persons>
```

Atribut `name` obsahuje celé jméno uživatele, `photoURI` slouží k získání URI fotografie, která bude u uživatele zobrazena. Pokud byly požadovány informace o skupinách, bude každá značka `person` obsahovat i značku `userGroups` (viz níže). Značka `person` by mohla obsahovat i další značky a textový obsah s informacemi o profilu uživatele.

Pro získání informací o skupinách uživatelů zašle klient zprávu:

```
<queryUserGroups filter="" withPersons=""/>
```

Ve filtru lze využít URI skupiny nebo její název. Pokud bude uveden i volitelný atribut `withPersons` s hodnotou `true`, budou v informacích o skupinách zahrnuty i informace o jejich členech. V názvu lze využít zástupné symboly „*“ (libovolný počet libovolných znaků). Server na tuto zprávu odpoví:

```
<userGroups>
  <group uri="" name=""/>
</userGroups>
```

Pokud byly požadovány informace o členech skupin, v každé značce `group` bude obsažena i značka `persons` (viz výše).

Pro přihlášení ke skupině uživatelů klient zašle následující zprávu:

```
<join group=""/>
```

kde atribut `group` obsahuje URI skupiny. K odhlášení uživatele ze skupiny slouží zpráva:

```
<leave group=""/>
```

B.3 Řízení odběru anotací

Klient může přijímat pouze anotace zvolených typů ze zvolených zdrojů. Zdrojem může být jiný uživatel nebo URI, který identifikuje anotační server, skupinu uživatelů či jiný obecný zdroj.

Klient se k odběru anotací přihlašuje následující zprávu:

```
<subscribe>
  <source type="" user=""/>
  <source type="" uri=""/>
  <source type=""/>
  <source user=""/>
  <source uri=""/>
</subscribe>
```

Elementů `source` může být libovolné nenulové množství a mohou mít kombinace parametrů, které jsou uvedeny výše. Parametr `user` identifikuje uživatele, `uri` obecný zdroj anotací. Parametr `type` udává typ anotací, přičemž s typem jsou automaticky vybrány i všechny podtypy. V typu může být využit i zástupný symbol „*“, který nahrazuje libovolný počet libovolných znaků.

Pokud není uveden typ, budou přijímány všechny typy anotací (dle skupin, ve kterých se uživatel nachází) z daného zdroje. Pokud není uveden zdroj, budou přijímány všechny anotace daného typu.

K odhlášení může klient využít zprávu `unsubscribe`:

```
<unsubscribe>
  <source type="" user=""/>
  <source type="" uri=""/>
  <source type=""/>
  <source user=""/>
  <source uri=""/>
</unsubscribe>
```

Pro odhlášení platí stejná pravidla jako pro přihlášení (libovolný počet elementů `source`, atd.). Klient automaticky oddebírá všechny anotace od svého přihlášeného uživatele, pokud se od jejich odběru explicitně neodhlásí.

B.4 Synchronizace dokumentu

Synchronizace dokumentu je proces, při kterém server získá kopii aktuální verze anotovaného dokumentu. Pokud server tento dokument získá poprvé, uloží si jej a vrátí klientovi adresu uložené verze, kterou bude klient využívat v anotacích. Pokud má server dokument uložený, porovná novou verzi s uloženou verzí, a pokud se shodují, zašle klientovi adresu uložené verze. Pokud se dokumenty shodují částečně, server vyhodnotí změny. Pokud změny neovlivní žádné anotace, dokument se transparentně aktualizuje (pro klienta stejně jako shoda dokumentů). Pokud by změny ovlivnily některé anotace, server zašle klientovi varování, že některé anotace musely být aktualizovány, a synchronizaci dokončí. V případě zásadnějších změn či zneplatnění anotací dojde k chybě synchronizace a uživatel se musí rozhodnout, zda synchronizaci dokončí a zneplatní tak některé či všechny anotace (vymaže či přesune na úroveň celého dokumentu), nebo nedokončí a bude anotovat uloženou (starší) verzi dokumentu či jiný dokument.

Protože klient může být i jednoduchý textový editor, který nepracuje se strukturovaným textem, server musí podporovat i linearizaci dokumentu. V tomto případě klient dokument linearizuje na prostý text a zašle jej serveru v linearizované podobě. Pokud má server strukturovanou podobu, linearizuje ji a porovná se zaslou. Pokud se linearizované verze neshodují, dojde k chybě synchronizace. Pokud se shodují, synchronizace bude dokončena a server bude každou následně zaslou anotaci upravovat pro strukturovanou verzi dokumentu.

Syntaxe:

```
<synchronize resource="http://example.com/documents/doc1.txt" linearize="false" overwrite="false">
  <content>
    <![CDATA[
      ...
    ]]>
  </content>
</synchronize>
```

Klient pošle serveru kopii anotovaného dokumentu a adresu, ze které pochází. Parametr **resource** udává umístění zdroje (např. URI anotované webové stránky). Element content obsahuje obsah daného dokumentu.

Nepovinný parametr **linearized** udává, zda klient pracuje s linearizovanou verzí dokumentu. Výchozí hodnota je false.

Nepovinný parametr **overwrite** umožňuje vynutit synchronizaci v případě, kdy je dokument s daným URI na serveru uložen, ale jeho obsah se neshoduje. Server v tomto případě musí nahradit uložený dokument a upravit (přesunout na úroveň celého dokumentu a doplnit textový obsah o informaci o změně) či vymazat všechny anotace. Klient by tento atribut neměl využít při prvním pokusu o synchronizaci. Jeho použití musí být schváleno uživatelem. Pokud se obsah textu shoduje, server atribut ignoruje. Výchozí hodnota je false.

Při úspěšné synchronizaci server odpoví zprávou:

```
<synchronized resource=""/>
```

Atribut **resource** obsahuje URI kopie anotovaného dokumentu, která je uložena na serveru. Tento URI musí klient využívat v anotacích.

Pokud v průběhu práce dojde k situaci, kdy se obsah anotovaného fragmentu neshoduje s obsahem, který server nalezne na dané pozici v dokumentu, dojde k tzv. rozsynchronizování. V tomto případě server zašle chybovou zprávu a zprávu **<resynchronize/>** (element bez obsahu a atributů), čímž požádá o resynchronizaci.

Klient provádí resynchronizaci zasláním obsahu v následující zprávě:

```
<resynchronize>
  <content>
    <![CDATA[
      ...
    ]]>
  </content>
</resynchronize>
```

Po resynchronizaci je vždy nutné znovu načíst všechny anotace. Server je tedy automaticky zašle v odpovědi. Pokud klient provádí modifikace dokumentu, musí každou změnu zaslat na server:

```
<textModification path="" offset="" length="">
  <![CDATA[
    Nový obsah vybraného fragmentu ...
  ]]>
</textModification>
```

Atribut `path` udává XPath uzlu DOM dokumentu, ve kterém byla změna provedena. Atributy `offset` a `length` udávají offset a délku změněného fragmentu. V těle elementu `<textModification/>` je potom uveden nový obsah fragmentu. Pokud je vložen nový fragment, délka původního fragmentu je nulová. Pokud je fragment vymazán, element je prázdný. Server následně tuto zprávu rozešle všem klientům pracujícím se stejným dokumentem.

B.5 Přenos typů anotací

Přenos typů anotací probíhá obousměrně. Pokud je přidán, upraven či vymazán typ, klient tuto změnu okamžitě zašle serveru a ten ji rozešle všem ostatním klientům, kterých se týká. Klient však nemusí udržovat kompletní strom typů, ale může mít načtenou pouze jeho část. V tomto případě může server buď zasílat všechny změny, nebo může udržovat informaci o tom, které části stromu má klient načtené, a zasílat pouze informace o změnách v těchto částech. Server musí vždy udržovat kompletní strom typů dané skupiny uživatelů (všechny změny jsou mu zasílány).

Klient o část stromu typů žádá zprávou:

```
<queryTypes filter=""/>
```

Atribut `filter` umožňuje získání určitého podstromu typů. Jedná se o linearizovaný název či URI typu se zástupnými symboly „*“ (libovolný počet libovolných znaků). Filtr tedy umožňuje vybrat podstrom nebo množinu typů, jejichž název či URI obsahuje daný text.

Pokud klient zažádá o strom typů, server na to odpoví zprávou s přidáním typů (viz níže). Pokud filtru nevyhovuje žádný typ, server zašle prázdný seznam typů.

Přenos typů anotací je prováděn následující zprávou:

```
<types>
  <add>
    <type name="" ancestor="" uri="" group="">
      <attribute name="" type="" required=""/>
    </type>
  </add>
  <change/>
  <remove/>
</types>
```

Element `types` obsahuje:

- element `add`, pokud byl přidán typ,
- element `change`, pokud byl upraven typ,
- element `remove`, pokud byl vymazán typ.

V každém ze tří výše uvedených elementů může být obsažen libovolný počet elementů `type`. Atributy tohoto elementu jsou `name` (název typu), `ancestor` (URI rodičovského typu), `uri` (URI typu) a `group` (URI skupiny uživatelů, které typ patří). Pokud je URI rodičovského typu prázdný, jedná se o základní typ. URI jednoznačně identifikuje typy. Pokud není uvedena skupina, určí se z URI typu, podle rodičovského typu nebo podle výchozí skupiny uživatele.

U každého typu mohou být definovány i výchozí atributy. Tyto jsou potom obsaženy v elementech `attribute`. Každý atribut má název (`name`) a typ (`type`). Pokud se jedná o jednoduchý datový typ, je uveden název tohoto typu. Pokud se jedná o vnořenou anotaci či odkaz na anotaci, je uveden očekávaný typ této anotace (informace o vnoření či odkazování zde není potřebná, protože obě varianty jsou zde významově ekvivalentní). Pokud je atribut povinný, má element `attribute` i atribut `required` s hodnotou `true`.

Název typu nelze upravit jinak, než smazáním starého typu a přidáním nového.

B.6 Přenos anotací

Přenos anotací je prováděn obdobně jako přenos typů:

```
<annotations>
  <add>
    <annotation/>
  </add>
  <change/>
  <remove/>
</annotations>
```

Každé přidání, úprava či vymazání anotace jsou ihned zaslány na server. V elementu `annotations` jsou dle provedené operace obsaženy elementy `add` (přidané) `change` (upravené) a `remove` (vymazané). V každém z těchto elementů může být obsažen libovolný počet elementů `annotation`.

Server každou změnu ihned zašle klientům, kterých se týká (u kterých dotčená anotace patří mezi odebírané). Pokud je přidána anotace, je tato anotace vždy zaslána i klientovi, který ji přidal, aby získal přidělený identifikátor anotace.

Po synchronizaci či resynchronizaci dokumentu jsou klientovi jako přidané automaticky zaslány všechny anotace, které patří k tomuto dokumentu a vyhovují klientem definovaným požadavkům na odběr anotací (zdroje a typy).

Pokud klient potřebuje znovu načíst některou anotaci (např. po neúspěšném pokusu o editaci) či všechny anotace, může serveru zaslat jednu z následujících dvou variant zprávy:

```
<reload uri="http://example.com/annotations/123456"/>
<reload all="true"/>
```

Atribut `uri` u první varianty zprávy udává URI požadované anotace, atribut `all` u druhé varianty udává, že mají být znovu zaslány všechny anotace.

B.7 Nabízení anotací

Server může klientovi nabídnout automaticky vygenerované anotace k danému dokumentu či jeho části. Klient o nabídku anotací zažádá následovně:

```
<suggestAnnotations path="" offset="" length="" type=""/>
```

Atributy `path`, `offset` a `length` udávají cestu (XPath), `offset` a délku fragmentu dokumentu, ke kterému mají být nabídnuty anotace. Pokud je uvedena pouze cesta, budou nabídnuty anotace k celému uzlu DOM dokumentu. Pokud není uveden žádný z těchto atributů, budou nabídnuty anotace k celému dokumentu. Volitelný atribut `type` udává požadovaný typ nabízených anotací. S tímto typem budou současně nabízeny i všechny jeho podtypy. Server odpoví nabídkou anotací:

```
<suggestions>
  <annotation tmpId="" confidence=""/>
</suggestions>
```

V elementu `suggestions` může být libovolný počet anotací (elementů `annotation`). Atribut `confidence` udává odhadnutou míru jistoty anotace v procentech. Hodnota může být využita klientem při automatickém přijímání a odmítání anotací. Anotace v nabídce nemají trvalý identifikátor, ale pouze dočasný (`tmpId`). Dočasný identifikátor slouží k informování serveru o manipulaci s nabídkami a vyřazení anotace z nabídky při její aktualizaci.

Pokud klient chce anotaci potvrdit (at' už akcí uživatele či automaticky na základě uživatelského nastavení), vrátí ji serveru mezi přidávanými anotacemi (ve zprávě `annotations`), přičemž dané anotaci (elementu `annotation`) přidá atributy `confirmed` a `tmpId`. Atribut `tmpId` udává dočasný identifikátor potvrzené anotace. Atribut `confirmed` udává způsob potvrzení a může nabývat následujících hodnot:

- `manually` - uživatelem potvrzená anotace,
- `manuallyEdited` - uživatel anotaci editoval a uložil,
- `automatically` - automaticky potvrzená anotace (dle nastavení doplňku),
- `automaticallyEdited` - automaticky potvrzená anotace s provedením automatických úprav.

Pokud dojde ke změně dokumentu, může být potřeba upravit nabídku anotací. V tomto případě server okamžitě zašle aktualizaci nabídky anotací. Pro jednodušší implementaci klienta není podporována úprava nabídnutých anotací. Pokud se některá anotace změní, je odstraněna a server nabídne novou verzi. V elementu `suggestions` tedy může být i libovolný počet elementů `delete`:

```
<suggestions>
  <annotation tmpId="" confidence=""/>
  <delete tmpId=""/>
</suggestions>
```


Pokud klient nemá anotaci s daným dočasným identifikátorem, element `delete` ignoruje. Pokud uživatel (či klient, dle nastavení) některou nabídku odmítne, klient zašle na server zprávu:

```
<refusedSuggestions>
  <suggestion tmpId="" method=""/>
</refusedSuggestions>
```

Atribut `method` udává způsob odmítnutí a může mít jednu z následujících hodnot:

- `manually` - odmítnutí uživatelem,
- `automatically` - automatické odmítnutí na základě nastavení.

Pokud klient nechce přijímat další aktualizace nabídky anotací, zažádá server o nabídky anotací k fragmentu dokumentu s nulovou délkou.

B.8 Přenos nastavení

Nastavení je seznam položek, které mají název a řetězcovou hodnotu. Nastavení lze rozdělit na nastavení serveru a nastavení klienta s tím, že nastavení klienta budou mít prefix „`Client`“ (např.: „`ClientAnnotationTypeColor:Animal->People->Employee`“ s hodnotou „`green`“). Při zobrazení uživateli se potom některá (známá) nastavení zpracují a zobrazí ve formulářích a ostatní se vypíší v tabulce pro ostatní nastavení, kde je uživatel může měnit.

Konkrétní položky nastavení závisí na implementaci serveru a klienta. Aby nedocházelo k problémům při využití více různých klientů jedním uživatelem, měly by být názvy položek nastavení prefigovány i typem a názvem klienta (např. „`ClientFFExtAnnotFox`“ bude prefix pro rozšíření Firefoxu nazvané `AnnotFox`) nebo by měly být takové, aby byl jejich význam zcela zřejmý (např. „`ClientDefaultAnnotationType`“ pro výchozí typ anotace).

Vzhledem k tomu, že je předpokládán malý počet položek a malá frekvence přenášení, vždy je přenášen kompletní seznam položek nastavení. Neuvedení položky tedy povede k jejímu odstranění (je-li možné). Pokud položku není možné odstranit (např. nastavení serveru), bude nastavena na výchozí hodnotu.

Nastavení se přenáší zprávou:

```
<settings>
  <param name="" value=""/>
</settings>
```

Elementů `param`, které tvoří jednotlivé položky, může být libovolný (i nulový) počet. Atribut `name` obsahuje název položky, atribut `value` řetězcovou hodnotu položky.

B.9 Chyby a varování

Chybové zprávy slouží k informování klienta o chybě. Chybová zpráva obsahuje číslo chyby (`number`) a její textový obsah (v elementu `message`). Může obsahovat i doplňující informace, které se týkají konkrétní chyby. Při chybě oprávnění při přístupu k anotacím bude obsažena informace o tom, ke kterým zdrojům byl odepřen přístup. Při nezdařené operaci s existující anotací (úprava či mazání) musí chybová zpráva obsahovat informaci o tom, které anotace se týká (kterou anotací je třeba znovu načíst). Při problému s atributy musí obsahovat i informaci o tom, kterých atributů se týká.

Textový obsah chybových zpráv bude lokalizován do jazyka nastaveného parametrem „`ServerLanguage`“, který bude mít hodnoty dle ISO 639-2 [27] (varianty pro bibliografické účely). Pokud tento parametr nebude nastaven, zprávy budou v angličtině.

Syntaxe:

```
<error number="1">
  <message>
    <![CDATA[
      Neplatné přihlašovací jméno nebo heslo.
    ]]>
  </message>
</error>
```

```

<error number="2">
  <accessDenied user=""/>
  <accessDenied uri=""/>
  <accessDenied type=""/>
  <accessDenied type="" user=""/>
  <accessDenied type="" uri=""/>
  <message>
    <![CDATA[
      Nemáte oprávnění k prohlížení zvolených anotací.
    ]]>
  </message>
</error>
<error number="3">
  <message>
    <![CDATA[
      Nemáte oprávnění přidávat anotace.
    ]]>
  </message>
</error>
<error number="4">
  <reload uri="http://example.com/annotations/123456"/>
  <message>
    <![CDATA[
      Editace dané anotace není povolena.
    ]]>
  </message>
</error>
<error number="5">
  <reload uri="http://example.com/annotations/123456"/>
  <message>
    <![CDATA[
      Vymazání dané anotace není povoleno.
    ]]>
  </message>
</error>
<error number="6">
  <reload uri="http://example.com/annotations/123456"/>
  <message>
    <![CDATA[
      Anotace tohoto typu musí obsahovat následující atributy: ...
    ]]>
  </message>
  <attribute name="" type=""/>
  <attribute name="" type=""/>
</error>
<error number="7">
  <reload uri="http://example.com/annotations/123456"/>
  <message>
    <![CDATA[
      Hodnota atributu ... musí být v rozsahu ...
    ]]>
  </message>
  <attribute name="" type=""/>
</error>

```

Čísla chyb jsou:

- 0 Nepodporovaná verze protokolu.
- 1 Chybné přihlašovací údaje.
- 2 Nedostatečná oprávnění k požadovaným anotacím.
- 3 Nelze přidávat anotace - přístup je pouze pro čtení.
- 4 Editace zvolené anotace není povolena.
- 5 Mazání anotace není povoleno.
- 6 Chybí povinné atributy.
- 7 Nedovolená hodnota atributu.
- 8 Chybné určení textu pro návrhy anotací.
- 9 Chyba synchronizace (s daným URI je asociován rozdílný obsah).
- 10 Nucená synchronizace není povolena.
- 11 Rozsynchronizování (neshoda anotovaného textu s textem na dané pozici).
- 12 Editace daného typu není povolena.
- 13 Mazání daného typu není povoleno.
- 14 Nelze přidávat typy anotací.
- 15 Neznámý typ atributu.
- 16 Chyba v přidávaném typu atributu.
- 17 Atributy přidávaného typu jsou chybné. Tyto atributy byly vynechány.
- 18 Chyba v upravovaném typu anotace - změny nelze uložit.
- 19 Neznámý typ anotace
- 20 Změna názvu typu či nadtypu není možná.
- 21 Chyba v nastavení. Změny nelze uložit.
- 22 Pokus o synchronizaci bez adresy dokumentu.
- 23 Pokus o synchronizaci bez obsahu dokumentu.
- 24 Chybná adresa zdroje anotovaného fragmentu.
- 25 Chybný anotovaný fragment. Anotace bude uložena bez tohoto fragmentu.
- 26 Chybný atribut anotace. Atribut nelze uložit.
- 27 Chybná hodnota v atributu rozšířeného typu.
- 28 Chyba v informacích o způsobu potvrzení nabídnuté anotace.
- 29 Upravovaná anotace nebyla nalezena. Změny nelze uložit.
- 30 Mazaná anotace nebyla nalezena. Anotaci nelze vymazat.
- 31 Vaše přihlášení vypršelo.
- 32 Chybná zpráva. Pravděpodobně chyba klienta nebo nekompatibilní protokol.
- 33 Chyba v modulu serveru.
- 34 Požadovaná anotace nebyla nalezena.
- 35 Chybná cesta v anotovaném fragmentu. Fragment nebude uložen.
- 36 Chyba v anotovaném dokumentu.
- 37 Chybný offset nebo délka v anotovaném fragmentu. Fragment nebude uložen.
- 38 V upravené anotaci jsou chyby. Změny nelze uložit.
- 39 Modifikaci textu není možné aplikovat na dokument.
- 40 Neznámý typ anotací. Nabízení anotací není možné.
- 41 Chybný formát data v anotaci. Datum bylo upraveno na aktuální.
- 42 Chybný formát data v atributu. Atribut byl vynechán.
- 43 Bez synchronizace dokumentu nelze manipulovat s anotacemi.
- 44 Chyba v informacích o autorovi anotace.
- 45 Neznámá skupina uživatelů.
- 46 Neznámá skupina uživatelů v typu anotace. Skupina bude nastavena dle definovaných pravidel.
- 47 Mazání využitých typů anotací není povoleno. Nejprve je nutno vymazat anotace tohoto typu.
- 48 Interní chyba serveru způsobila neúspěch při ukládání dat. Žádná data nebyla uložena.
- 49 Pokus o vytvoření duplicitního typu anotace (shoda URI).
- 50 Chybný popis modifikace textu.

100 Neznámá chyba.

Pokud dojde k zneplatnění anotace či k jejímu přesunu na globální úroveň, aniž by došlo k chybě, může být potřeba varovat uživatele. V tomto případě server zašle zprávu s varováním. Klient by měl varování zobrazit uživateli a to buď přímo u dotčené anotace, nebo pomocí nějakého postranního panelu či dialogového okna.

Zpráva s varováním má následující syntaxi:

```
<warning number="1" annotation="http://example.com/annotations/123456">
  <![CDATA[
    Anotace byla zneplatněna editací textu.
  ]]>
</warning>
```

Každé varování má číslo (atribut `number`) a textový obsah pro zobrazení uživateli. Pokud se varování týká konkrétní anotace, element `warning` má i atribut `annotation`, který obsahuje URI dané anotace.

Čísla varování jsou:

- 1 Anotace zneplatněna.
 - 2 Anotace přesunuta na globální úroveň.
 - 3 Anotace automaticky aktualizována.
 - 4 Některé anotované fragmenty v anotaci byly zneplatněny.
- 100 Jiné varování.

B.10 Potvrzení bez doplňujících dat

Pokud je zaslána zpráva, která vyžaduje provedení nějaké operace, je třeba na ni odpovědět, aby druhá komunikující strana měla potvrzeno přijetí zprávy, případně úspěšné provedení operace, kdy nejsou vrácena data. Pokud není zaslána chybová zpráva, předpokládá se úspěšné provedení operace. Odpověď tedy může obsahovat pouze informace, které jsou důsledkem provedené operace, či jiné užitečné informace. V některých případech však nejsou k dispozici žádné užitečné informace k zaslání a je potřeba, aby server či klient potvrdil úspěšnost operace. V tomto případě pošle následující zprávu:

```
<ok/>
```

C Zjednodušený příklad komunikace mezi klientem a serverem

Klient (zahájení komunikace, uvedení verze protokolu, autentizace):

```
<connect/>  
<login/>
```

Server (potvrzení, uvedení verze protokolu, nastavení parametrů):

```
<connected/>  
<logged/>  
<settings/>
```

Klient (volba zdrojů anotací a anotovaného textu, požadavek na seznam typů):

```
<session/>  
<subscribe/>  
<synchronize/>  
<queryTypes/>
```

Server (adresa zasynchronizovaného zdroje (kopie na serveru), požadované anotace a typy):

```
<synchronized/>  
<annotations/>  
<types><add/></types>
```

Klient (požadavek na přidání typu anotace):

```
<session/>  
<types><add/></types>
```

Server:

```
<ok/>
```

Klient (zašle vloženou anotaci):

```
<session/>  
<annotations/>
```

Server (zašle zpět vloženou anotaci s přiděleným identifikátorem):

```
<annotations/>
```

Klient (zašle upravenou anotaci):

```
<session/>  
<annotations/>
```

Server (vrátí chybu):

```
<errors/>
```

Klient (požádá o původní obsah anotace):

```
<session/>  
<reload/>
```

Server (vrátí požadovanou anotaci):

```
<annotations/>
```

Klient (přidání anotace):

```
<session/>  
<annotations/>
```

Server (rozsynchronizování):

```
<errors/>  
<resynchronize/>
```

Klient (opakování synchronizace):

```
<session/>  
<resynchronize/>
```

Server (všechny anotace pro daný text ze zvolených zdrojů):

```
<annotations/>
```

Klient (požadavek na změnu nastavení):

```
<session/>  
<settings/>
```

Server (aktuální nastavení):

```
<settings/>
```

Klient (požadavek na získání nabízených anotací):

```
<session/>  
<suggestAnnotations/>
```

Server (nabídka anotací):

```
<suggestions/>
```

Klient (přechod na jiný anotovaný text):

```
<session/>  
<synchronize/>
```

Server:

```
<synchronized/>  
<annotations/>
```

Klient (odhlášení, ukončení spojení):

```
<session/>  
<logout/>  
<disconnect/>
```

Server:

```
<ok/>
```