

Sémantická anotace textu

Jaroslav Dytrych

Fakulta informačních technologií Vysokého učení technického v Brně
Božetěchova 1/2. 612 66 Brno – Královo Pole
dytrych@fit.vutbr.cz



24 března 2017

Garuda Indonesia (GA)

AdamC said Dec 2014:

5 stars

Best South east Asian Airline

The only thing I can say is being American Size, I wish the seat would be bigger, But the First class ain't that expensive either... If you compare price wise, availability, Service wise. Garuda is the best. I had nightmares (Plural) with air Asia at baggage drop. Singapore airline is good but a little expensive. Now I don't even questions myself, it's always GARUDA airline. Me and my wife which is native from Indonesia, travel often around the world from Montreal to Jakarta, Surabaya. Denpasar. I have never been disappointed. Service is above average, agents are respectful.]

File ▾ Edit ▾ View ▾ Format ▾ Tools ▾

Annotations <>

Garuda Indonesia (GA)

AdamC said Dec 2014:

5 stars

Best South east Asian Airline

The only thing I can say is being American Size, I wish the seat would be bigger, But the **First class** ain't that expensive either... If you compare **price** wise, availability, **Service** wise. **Garuda** is the best. I had **nightmares (Plural)** with air Asia at **baggage drop**. Singapore airline is good but a little expensive. Now I don't even questions myself, it's always GARUDA airline. Me and my wife which is native from Indonesia, **travel** often around the world from Montreal to Jakarta, Surabaya, Denpasar. I have never been disappointed. **Service** is above average. **agents** are respectful.

Annotations (1) ✕

Surabaya (Location)

Name Surabaya

Visual_representation



Description

Surabaya is Indonesia's second-largest city with a population of over 3.1 million, and the capital of the province of East Java. It is located on the northern shore of eastern Java at the mouth of the Mas River and Java Sea (the Madura Strait). It is one of the most important ports in Indonesia.

- Primárním cílem je ukázat, že propojení **strukturování znalostí s tagováním**, podpořeným **pokročilými anotačními nástroji**, zjednodušuje proces **získávání a strukturování znalostí** a na základě **experimentů** vyhodnotit a **kvantifikovat přínosy** tohoto postupu.

- Srovnání uživatelských rozhraní anotačních nástrojů
- Vytváření strukturovaných anotací
- Anotování kdekoliv a kdykoliv
- Zvýšení kvality vytvářených anotací
- Formalizace znalostí na základě anotování textu

- V první fázi srovnání ~ 65 nástrojů z obecnějších hledisek
 - granularita anotací, atributy, možnosti strukturování, ...
- Ve druhé fázi detailní srovnání pokročilé funkcionality 6 nejvhodnějších nástrojů
 - vztahy mezi anotacemi, překrývání fragmentů, nabízení anotací, ...
- Ve 3. fázi experimenty s vybranou skupinou uživatelů

- Univerzální modulární nástroj se širokými možnostmi konfigurace
- Pro praktické využití i jako platforma pro ověřování nových konceptů v oblasti anotování
- Pokročilá funkcionality:
 - podpora překrývání fragmentů (i částečného)
 - práce s nabídkami anotací a jejich alternativami
 - řízení odběrů anotací
 - ...
- Popis publikován v článku *Advanced Features of Collaborative Semantic Annotators – The 4A System*

As he had in **Luncheon on the Grass**, **Manet** again paraphrased a respected work by a Renaissance artist in the painting **Olympia** (1863), a nude portrayed in a style reminiscent of early studio photographs, but whose pose was based on **Titian's Venus of Urbino** (1538). The painting is also reminiscent of **Francisco Goya's** painting **The**

Annotations (1)

Venus of Urbino (Art->Artwork)

Visual_representation



Author

Titian (artist->Coreference_to_artist)

Italian painter, the most important member of the 16th-century Venetian school

Visual_representation



Name Titian

Description

Tiziano Vecelli or Tiziano Vecellio

Document Annotations

(Influence)

Influencer

Titian (artist->Coreference_to_artist)

Italian painter, the most important member of the 16th-century Venetian school

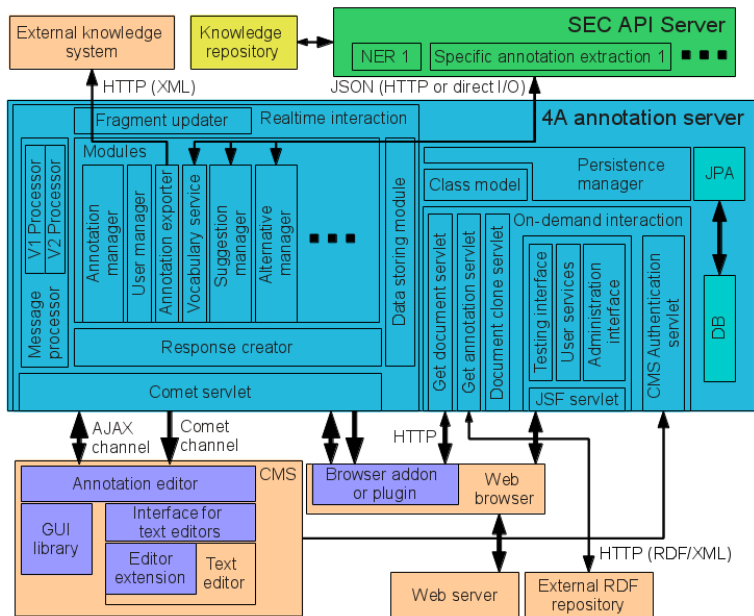
Influencee

Manet (artist)

French painter. He was one of the first 19th-century artists... (1832 - 1883)

Visual_representation





- Annotea (v roce 2000)
- 4A (vývoj od r. 2009, 2011 využit v 1. verzi systému)
- Web Annotation
 - 2010 Annotation Ontology
 - 2011 Open Annotation Beta
 - 2012 Open Annotation Extensions – 1. verze podporující potřebné **strukturování anotací**
 - 2013 Web Annotation Community Draft
 - 2014 **Working draft** – přechod 4A na Open Annotation
 - 2016 Candidate Recommendation

- Obvykle využití REST API nebo proprietárního protokolu
- V r. 2004 Lee Feigenbaum ve firmě IBM navrhl API
 - Vyžaduje využití knihoven (je definované API, nikoliv protokol)
 - Nepodporuje nabízení anotací
 - Neřeší synchronizaci dokumentu
- Protokol 4A
 - v roce 2011 verze 1.0
 - v roce 2012 verze 1.1 (opravy chyb a drobná vylepšení)
 - v r. 2013 nasazena v projektu DECIPHER
 - v roce 2014 verze 2.0 (Open Annotation, více klientů na stránce, vylepšení odběrů a nabízení anotací, ...), která byla formalizována J. Macháčkem dle poskytnutých podkladů

- Správa sezení
- Skupiny uživatelů
- Synchronizace dokumentu
- Typy anotací
- Odběry anotací
- Nabízení anotací
- Kontrolovaný slovník
- ...

- Zjištění vlivu množství zobrazovaných informací a dalších aspektů uživatelského rozhraní na kvalitu vytvářených anotací
 - Sledování pracovních postupů uživatelů
- Optimalizace
- Kvantifikování přínosů pokročilých funkcí:
 - Sémantické filtrování
 - Alternativní nabídky anotací

- 1 Úvodní experimenty
 - Vyzkoušení první verze nástroje a ověření základních konceptů
- 2 Praktické nasazení v projektu Decipher
 - Vyhodnocení z perspektivy uživatelů a jejich pohledu na jednotlivé prvky GUR – využití pro tvorbu 2. verze systému
 - Testy vedli projektoví partneři, nejsou k dispozici kvantitativní data
- 3 První série experimentů
 - 3.1 Jak jednotlivé prvky uživatelského rozhraní ovlivňují kvalitu výsledků a rychlost anotování?
 - 3.2 V jakém rozsahu množství zobrazovaných informací ovlivňuje výsledky zjednodušování entit?
 - 3.3 Je pro uživatele výhodou znalost potenciálních alternativních anotací a poskytnutí míry důvěryhodnosti jednotlivých nabídek anotací?
 - 3.4 Jaké výhody do procesu anotování přináší koncept sémantického filtrování?
- 4 Druhá série experimentů
 - 4.1 Sémantické filtrování a optimalizace množství zobrazovaných informací
 - 4.2 Přínosy alternativních nabídek anotací
- 5 Kolaborativní vytváření ontologie

- Všechny experimenty byly zaměřeny především na pokročilé anotační úlohy, jako je hierarchické anotování komplexních vztahů v textu a entit s mnohoznačnými jmény a tomu odpovídajícími mnoha potenciálními položkami ve znalostní bázi

- Výsledky:
 - Při anotování vztahů uživatelé preferují strukturované anotace
 - Definuje-li jeden uživatel šablonu pro nový typ anotace, ostatní ji využívají, aniž by dělali větší změny
 - Nabízení anotací zrychluje anotování
- Publikace *Towards New Scholarly Communication: A Case Study of the 4A Framework*

- Popis celé série experimentů publikován v článku *Interaction Patterns in Computer-assisted Semantic Annotation of Text – An Empirical Evaluation*
- První experiment zaměřen na srovnání anotačních nástrojů s různými GUR a vzory interakce
 - 4A, RDFaCE a Gate Developer
 - Každý uživatel strávil 20 minut seznámením s nástrojem a max. 40 minut anotováním každým z nástrojů
 - Instrukce požadovaly, aby se uživatelé snažili dosáhnout **maximální možné přesnosti, ale čas na dokončení celého úkolu byl omezený**. Intuitivně tedy volili vhodný kompromis mezi rychlostí a kvalitou

Larkin was born in Liverpool, England, in 1876 to northern Irish immigrants, James Larkin and Mary Ann McNulty. He and his fa

Edit Annotation

⊖ (Event)

▶ ◇ start_time: 1876 (Date)

⊖ ◇ agent: Larkin (Person->

◇ Name: James Larkin

◇ Visual_representation

◇ Description: James L

◇ URI_of_entity: http:/

▶ ◇ location: Liverpool, Engl

▶ ◇ activity: born (Activity)

◇ end_time: -- no value --

Attribute Type

Person

Value Type

Person->Politician

Selection

Larkin

Content

☐ Nest Values

☒ Include attribute in the annotation type

☐ Required attribute

☐ Apply changes on the type attribute

+ Attribute

Attribute

+ Value

Value

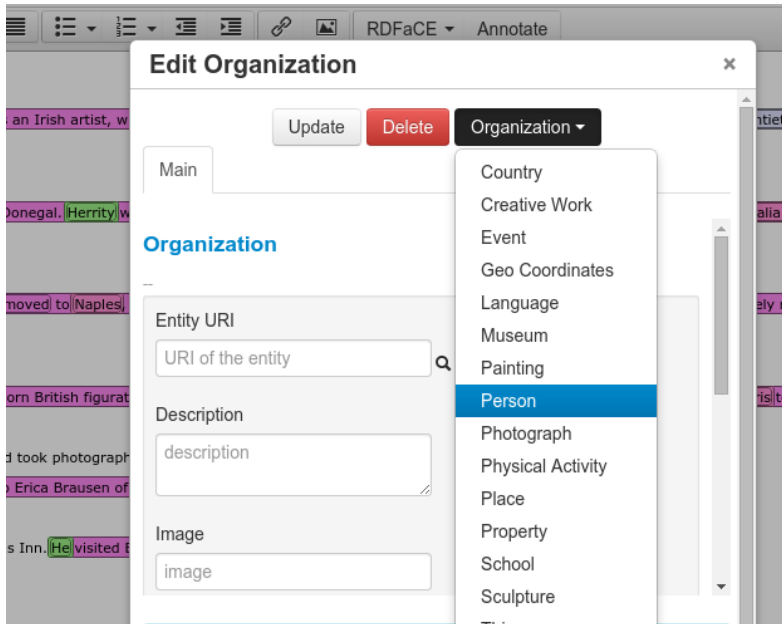
Clear

KB ref.

☐ Stay open

✕ Cancel

Save



GATE Developer 8.0 build 4825 <@knot09>

File Options Tools Help

GATE

- Applications
- Language Resources
 - idytrych.xml_00012
 - smrz.xml_00008
- Date
- Activity
- Person
- Location
- Event
- Processing Resources
- Datstores

Messages idytrych.xml_00...

Annotation Sets Annotations List Annotations Stack Co-reference Editor Text

Francis Joseph O'Meara -
Francis Joseph O'Meara (March 30, 1853 - 1888) was an Irish artist. The son

☒ Activity

Person

description		X
name	Francis Joseph O'Meara	X
uri	http://en.wikipedia.org/wiki/Frank_O%27Meara	X
visual_representation		X
		X

Open Search & Annotate tool

befriended.

- Norah McGuinness -
Norah McGuinness (7 November 1901, County Londonderry, Ireland - 22 November 1980, County Dublin) was an Irish painter and illustrator. She married the editor Geoffrey Phibbs, but they divorced in 1929. The same year she travelled to Paris to work at the studio of André Lhote. From there she moved to London where she was a member Lucy Wertheim's 'Twenties Group' and of the avant-garde London Group, from 1937-39 she lived in New York.

MimeType text/plain
docNewLineType LF
gate.SourceURL file://

Nástroj	Nesprávné hodnoty	chybějící hodnoty	čas na událost
GATE	9.4 %	8.3 %	135 s
RDFaCE	8.7 %	8.8 %	193 s
4A	6.2 %	5.6 %	116 s

- Přiřazení atributů k události pouhým faktem, že fragment označující událost překrývá fragment s hodnotou atributu, vede k nejasným či nekonzistentním výsledkům
- Vizuální odlišení nabídek anotací generovaných systémem od anotací zkontrolovaných či vytvořených uživatelem vede na lepší a konzistentnější anotace

2. experiment zaměřen na srovnání 3 nastavení pohledu pro zjednoznačňování, která se lišila zobrazenými atributy (6 testerů, 30 min. na 40 textů)

1. Pro každou nabídku zobrazen rozsáhlý seznam atributů a jejich hodnot
2. Pouze typ entity a URL (jako např. v RDFaCE)
3. Speciální zjednoznačňovací atribut (určen z popisů dostupných alternativ), typ entity a URL

Bloomsbury

Some time ago Jenny Woolf contacted me to draw my attention to a book by
Thomas Burke (1886-1945) entitled Living in Bloomsbury. London: G. Allen &

Suggestions (1)

Thomas Burke (Person)



Irish physician, lawyer, and statesman who lived in Hillsborough,
 North Carolina (1747 - 1783)

URI_of_entity

[http://en.wikipedia.org/wiki/Thomas_Burke_\(North_Carolina\)](http://en.wikipedia.org/wiki/Thomas_Burke_(North_Carolina))

nastavení	průměrný čas	množství chyb	kliknutí na URL
detailní informace	33.92 s	6.2 %	1.3 %
pouze typ a URL	37.26 s	27.9 %	41.7 %
zjednoznačňovací atribut	32.98 s	2.1 %	1.5 %

3. experiment zaměřen na srovnání 2 nastavení zobrazení alternativních nabídek anotací:

- 1 Nejlepší alternativa se zobrazením zjednotňovacího atributu
- 2 Pět nejpravděpodobnějších nabídek a míra důvěry
 - Dvě datové sady po 15 úryvcích textu zmiňujících vybrané entity, kde 3 z nich nebyly pokryty referenčními zdroji, tedy žádnou z alternativ (poměr/počet nepokrytých uživatelé neznali)
 - První nabídka byla správná v 9 z 15 případů v každé sadě

nastavení	průměrný čas	množství chyb	kliknutí na URL
jedna nabídka	41.2 s	14.4 %	3.3 %
všechny alternativy	42.9 s	12.2 %	2.2 %

- Výsledky nejsou závislé na kvalitě procesu generování nabídek anotací v pozadí a na četnosti případů, kdy zmíněná entita není pokryta referenčními zdroji
- Zobrazení všech alternativ se ukazuje jako vhodnější, chceme-li v daném čase preferovat kvalitu vytvářených anotací před kvantitou

4. experiment zaměřen na nabízení anotací a sémantické filtrování

- Dvě sady textů, v každé 6 událostí
- V první sadě měli uživatelé identifikovat atributy událostí manuálně, ve druhé sadě bylo využito automatické předanotování a sémantické filtrování

vyplněné atributy	nesprávné hodnoty	chybějící hodnoty	čas na událost
manuálně	11.7 %	6.6 %	303.5 s
s nabízením	4.5 %	3.4 %	109.3 s

- Manuální proces bez předanotování je zdoluhavý. Výsledky také obsahovaly mnoho chyb
- Z výsledků není jasné, v jakém rozsahu zvýraznění preferovaných typů anotací v textu napomáhá k vyšší kvalitě a rychlosti anotování

- Popis druhé série experimentů bude publikován v kapitole *Advanced User Interfaces for Semantic Annotation of Complex Relations in Text* knihy s nejlepšími články z ICAART 2017 ze série LNCS

1. experiment zaměřen na sémantické filtrování a optimalizaci množství zobrazovaných informací

- 14 uživatelů, 20 úryvků o uměleckých dílech a vztazích vzájemného ovlivnění umělců, včetně ovlivnění mezi těmito díly
- 2 konfigurace systému:
 - zvýrazněny anotace příslušných typů dle právě vyplňovaného atributu
 - sémantické filtrování vypnuto

sémantické filtrování	nesprávné hodnoty	chybějící hodnoty	čas na vztah
vypnuté	6.9 %	5.7 %	41.4 s
zapnuté	5.1 %	4.2 %	35.1 s

- Zapnuté sémantické filtrování vedlo k vyšší kvalitě výsledků. Relativní snížení obou typů chyb překročilo 25 %. Proces anotování byl také o 15 % rychlejší
- Dotazníky vyplňované po skončení experimentu ukázaly, že 11 ze 14 uživatelů je toho názoru, že funkce sémantického filtrování výrazně zlepšuje jejich komfort při anotování

2. experiment byl zaměřen na přínosy alternativních nabídek anotací

- Srovnány dvě možnosti prezentace seznamu alternativ uživateli:
 - vypsané pouze zjednoznačňovací kontexty a umožněno rozbalení kompletního seznamu atributů
 - všechny atributy ihned rozbalené a uživatel měl možnost je pro každou sbalit

Alan Young

Alan Young was born in Kirkcaldy but played his entire professional career in

Suggestions (5) Refuse all 5 suggestions

- Alan Young (Person)
English-born, naturalised United States actor, comedian and voice artist (b. 1919) **C: 56 %**
- Alan Young (Person)
lawyer, associate professor of law at Osgoode Hall Law School (b. 1933)
- Alan Young (Person)
Australian footballer, former Australian rules footballer who played with Collingwood
- Alan Young (Person)
Scottish footballer who played as a forward in the Football League (b. 1955)

Visual_representation



Name Alan Young

Description

nastavení	čas na alternativu	množství chyb	kliknutí na sbalení /rozbalení
sbalené	11.2 s	12.1 %	1.1 %
rozbalené	15.1 s	11.4 %	5.3 %

- Stručný pohled uživatelům umožnil rychleji procházet alternativy a zvolit jednu z nich.
- Všichni kromě jediného uživatele zde jako preferovaný označili právě tento pohled.

- Protokol pro přenos anotací a formát anotace 4A
 - Dosud neexistoval žádný otevřený univerzální protokol pro přenos anotací
 - Použitelnost a univerzálnost prokázána při nasazení v projektu DECIPHER. Změna z 4A na Open Annotation nevyžadovala zásah do základu protokolu
 - Do budoucna by na základě formátu 4A mohlo být vytvořeno rozšíření formátu Web Annotation, které by rozšířilo jeho možnosti pro využití pro komplexní anotační úlohy
- Anotační systém
 - Univerzální anotační nástroj, který lze snadno nasadit do CMS a využít pro hierarchickou anotaci komplexních událostí
 - Platforma umožňující výzkum v oblasti anotování
 - Využitelnost pro ověřování nových anotačních konceptů byla demonstrována experimenty

- Prokázán výrazný vliv uživatelského rozhraní na rychlost a kvalitu anotování
- Prokázána správnost konceptů optimalizujících množství zobrazovaných informací, pro dosažení vyšší rychlosti anotování i kvality vytvářených anotací
- Prokázán přínos nového konceptu alternativních nabídek anotací a konceptu sémantického filtrování
- Zahájeno zkoumání výhod realizovaného anotačního nástroje pro kolaborativní rozšiřování ontologií

- Nasazení anotačního systému v aktuálních projektech
- Příprava protokolu pro přenos anotací pro spolupráci s W3C
- Detailnější experimenty vycházející z vyhodnocení výsledků z jednotlivých kroků realizovaných v této práci
- Příprava nové verze exportu ontologií
- Experimenty na paralelní kolaborativní práci v různých prostředích a vyhodnocení příslušných vzorů interakce na týmových úlohách.

Time elapsed: 33 ms

[Text](#) | [Context](#) | [A](#) | [Link](#) artists including [Gerard Dou](#) , [Jan Steen](#) and [Gabriel Metsu](#) . References External lin
[Text](#) | [Context](#) | [A](#) | [Link](#) [Golden Age](#) genre painter in [the style](#) of [Gabriel Metsu](#) . Biography Biography.
[Text](#) | [Context](#) | [A](#) | [Link](#) founded the
[Text](#) | [Context](#) | [A](#) | [Link](#) painting a d
[Text](#) | [Context](#) | [A](#) | [Link](#) (c. [1628 - 16](#)
[Text](#) | [Context](#) | [A](#) | [Link](#) In [1665](#) he t
[Text](#) | [Context](#) | [A](#) | [Link](#) [Jan ! van !](#)
[Text](#) | [Context](#) | [A](#) | [Link](#) Haarlem [165](#)
[Text](#) | [Context](#) | [A](#) | [Link](#) [Oman](#). Hea
[Text](#) | [Context](#) | [A](#) | [Link](#) , [Pieter ! d](#)
[Text](#) | [Context](#) | [A](#) | [Link](#) , Volume I (J
[Text](#) | [Context](#) | [A](#) | [Link](#) This painting
[Text](#) | [Context](#) | [A](#) | [Link](#) after him, su
[Text](#) | [Context](#) | [A](#) | [Link](#) [1675](#); [Histor](#)
[Text](#) | [Context](#) | [A](#) | [Link](#) [Hans de Jo](#)
[Text](#) | [Context](#) | [A](#) | [Link](#) corset blue"
[Text](#) | [Context](#) | [A](#) | [Link](#) [Pieter ! de](#)

AnnotationWindow

File ▾ Edit ▾ View ▾ Format ▾ Tools ▾

Annotations <>

The Empty Glass The Empty Glass [The Empty Glass](#) (c. [1652](#)) is an oil on canvas painting by example of Dutch Golden Age painting and is now in the collection of the [Museum Boijmans](#) documented as a work by [Gabriel Metsu](#) by [Hofstede de Groot](#) in [1908](#) who wrote: "204b. A Officer. Sm. by 14 inches Heemskerck website in t

Suggestions (1)

Dutch (Art->Art_period_movement) painting of the Dutch Golden Age, a period in Dutch...

Name Dutch_Golden_Age_painting **C: 39%**

Visual_representation



Description

The Dutch Golden Age painting is the painting

1 2 [Next](#)

Otázky recenzentů

Jak náročné by bylo integrovat 4A, formou editorového doplňku, s některým běžným CMS, např. WordPressem?

- Integrace zahrnuje tyto úkony:
 - Přidání příslušných skriptů do stránky
 - Přidání zásuvného modulu do textového editoru
 - Přidání komunikace s anotačním serverem ohledně autentizačních informací v reakci na přihlášení a odhlášení uživatele
 - Úprava proxy pro dané prostředí
 - Případný formulář v konfiguraci pro volbu adresy anotačního serveru (má-li být možná i přes web)
 - Vytvoření metadat modulu
- Pro Drupal cca 500 řádků kódu
- Pro vývojáře, který zná daný CMS, max. 1 den práce

V textu není jasně demonstrováno, jak se v případě obohacování ontologie začlenění nová entita do její struktury. Jak se např. rozliší, jestli chtěl uživatel přidat třídu nebo instanci? Může být přes anotační rozhraní přidána i vlastnost? Prosím o vysvětlení a ukázání konkrétních příkladů. Návazně prosím i o stručné vyjádření, jaké typy logického odvozování jsou ontologií (resp. jejími částmi) vzniklými přes anotační rozhraní) podporovány.

- Primárně se **nejedná o nástroj pro editaci ontologií**
- Anotační nástroj poskytne částečné informace (nadtyp, vlastnosti, některá omezení, ...) a umožní zobrazit příklad použití, začlenění se dokončí ve specializovaném nástroji
- Zjednodušení – **vždy je anotována instance** (např. konkrétní člověk) a **typ anotace vždy představuje třídu** (jeho **atributy pak vlastnosti**)
- Nejprve je naimportována ontologie ze znalostního nástroje, následně je rozšiřována v anotačním nástroji, exportována a nově přidané části jsou dodefinovány ve specializovaném nástroji
- Odvozování bylo mimo zkoumanou oblast (nebylo testováno)

Spíše formálně zaměřená otázka: jaký je podíl dalších spolupracovníků na jednotlivých výsledcích v disertaci prezentovaných?

- Článek na SePublica 50 % doc. Smrž
- Článek na FLAIRS 50 % doc. Smrž
- Článek na ICAART 20 % doc. Smrž
- Článek do LNCS 20 % doc. Smrž
- Tabulkové srovnání anotačních nástrojů jsem dělal sám, doc. Smrž mi pak posílal odkazy na další nástroje
- Návrh architektury anotačního systému a rozložení všech podstatných částí jeho uživatelského rozhraní jsem udělal sám a později zapracovával připomínky doc. Smrže, projektových partnerů apod. (podíl těžko měřitelný)
- Návrh 1. verze protokolu pro přenos anotací a formátu 4A jsem připravil sám, protokol verze 2 na základě mých podkladů formalizoval J. Macháček

- Produkty dle IS FIT, např.:
 - Server for sharing annotations (2010, 1. vykázaný produkt)
50 % doc. Smrž (připomínkování, testování, ...) – zahrnut podíl na pomoc v počátcích práce
 - Annotation system for Drupal 22 % doc. Smrž, 56 % mnou vedení studenti (1–6 % / studenta, Ing. Cudrák 16 % – implementace klienta)
- Při měření experimentů mi pomáhali studenti (sledování uživatele, počítání akcí, zapisování apod.) – podíl těžko měřitelný
- Datové sady jsem připravil sám s využitím dat projektů kolegů ve výzkumné skupině (např. předzpracovaná stažená Wikipedie v HTML a seznam stránek s výskyty textově shodných odkazů). S některými mechanickými činnostmi mi později pomáhali studenti
- NER pro generování nabídek anotací je mimo moji práci (práce Ing. Otrusiny)

V práci zmiňujete možnost předání Vámi navrženého protokolu kolegům z W3 konsorcia. Jak jste pokročil v této záležitosti?

- Další postup je v plánu na léto 2017, kdy bych rád navázal na svoji předchozí e-mailovou komunikaci s dr. Paolem Ciccarese. Je potřeba připravit:
 - online demo
 - detailní popisy k některým částem dokumentace
 - dokumentaci obecné architektury systému
 - obecnější popis protokolu (nyní obsahuje i popis konkrétního použití Open Annotation)

Má Váš systém, případně jeho části, další perspektivu využití mimo dokončený evropský projekt?

- V plánu je nasazení v projektu Centrálního portálu českých knihoven (v projektu NAKI II CPK)
- Momentálně s několika studenty pracuji na 3. verzi systému
 - Přechod na novější technologie (novější knihovny, aplikační server apod.)
 - Lokalizace 2. verze klienta do češtiny a odstranění drobných nedostatků zjištěných při testování
 - Optimalizace serveru

Můžete se případně srovnat/vymezit se vůči systému ANNOTA.

- Annota (<http://annota.fiit.stuba.sk/>)
 - Zaměřen na vyhledávání relevantních dokumentů. Nástroj primárně určen pro vytváření záložek
 - Platforma pro výzkum metod pro automatickou organizaci kolekcí dokumentů, navigaci ve velkém množství informací a vyhledávání dle obsahu
- 4A
 - Zaměřen na anotování textu, včetně vytváření strukturovaných anotací
 - Platforma pro výzkum uživatelských rozhraní pro anotování a postupů anotování

- Annota (<http://annota.fiit.stuba.sk/>)
 - Dokumentu lze přiřadit textovou poznámku a tagy. V rámci dokumentu pak lze označit fragment a anotovat textovou poznámkou
 - Anotace nelze strukturovat (atributy ani odkazy mezi anotacemi). Anotace jsou provázané pouze tagy a vztahy k dokumentům a uživatelům
 - Pro (celý) dokument se automaticky extrahují metadata (nadpis, autor, název, relevance dokumentu pro daný vyhledávaný termín, ...). Výsledek nelze editovat
- 4A
 - K celému dokumentu nebo k libovolnému fragmentu lze vytvořit anotaci s tagem (typem), textovou poznámkou a atributy
 - Anotace lze strukturovat pomocí atributů, které mohou odkazovat i na jiné anotace, nebo je mohou přímo obsahovat
 - Pro celý dokument i pro jednotlivé fragmenty v dokumentu se nabízejí anotace rozpoznané nástroji pro vyhledávání pojmenovaných entit v textu. Anotace lze potvrdit, odmítnout či editovat a potvrdit

- Annota (<http://annota.fiit.stuba.sk/>)
 - Neumožňuje filtrovat anotace
 - Výběr anotovaného fragmentu obsahem, kontextem, pořadím výskytu a počáteční a koncovou stránkou (pouze PDF)
 - ...
- 4A
 - Anotace lze filtrovat podle typu, autora či skupiny uživatelů
 - Pro výběr anotovaného fragmentu používá XPointer či XPath, offset, délku a obsah
 - ...

A formal model of annotations of digital content ¹

Annota ¹

Read later: ☐ Favorite: ☐

Tags:

² Insert tag

Notes:

³

⁴ Add shared note.

⁵ Highlights

This page has no annotations yet.

A formal model of annotations of digital content 20

Private ☐

² This is my first highlight with note

 ³

Převzato z <http://annota.fiit.stuba.sk/>

- (PC) Patrick Curran (Sun Microsystems)
 - (KD) 
 - (DH) Jiří Trhlik 2012-06-19 13:54
 - (LH) Person
- The following
- (DD) Email = patrik@email.com
 - (DM) David Marston (DMV) [GUEST]

Test Case Description Language

The group had a closer look to the proposal by Dave Marston for a Test Description Language. It seemed worth developing and having in our "toolbox", but apart from Dave, the the WG lacks resources to develop it.

Anotace - tester00

Vybraný text: Patrick Curran

Anotace:

Typ: Person Vybrat

Atributy:

- Age
- Email
- Home
- Name
- User

Typ: String Vybrat

Text: patrik@email.com

Změnit nastavení atributu

Dashboard Content Structure Appearance People Modules Configuration Reports Help

Add content Find content

Tahiti

Gauguin left France again on 13 July 1895, never to return. His time away, particularly in Tahiti and Hiva Oa Island, was the subject of much interest both to his alleged sexual exploits.

Annotate

Attributes

☐ Annotation

Activity

Object

End_time

Agent

Location

Start_time

P12F.occurred_in_the_pres

P20B.was_purpose_of

P11F.had_participant

Paul Signac

Signac loved sailing and began his boat at Tahiti.

Path:

Disable rich-text

Text format

- Allowed
- <figcaption>
- Web page
- Lines are

Selection:

*** NOTHING ***

Select

Annotate whole document

1/1

<

>

Add

Delete

Type:

E1.CRM_Entity->E77.Persistent_Item

Browse

☒ Add to the basic attributes of given annotation type

☐ Required attribute

Add

Delete

Vyjádření k dalším otázkám v posudcích

- Výběr existujících řešení v kapitole 3.3
 - Uvedeny různé typy nástrojů pro popsání šířky spektra dostupných možností
 - Nástroj MAT není podporovaný (je neperspektivní) a byl proto vyřazen z praktických experimentů. Ze stejného důvodu jsem se rozhodl mu nevěnovat prostor v textu práce
 - Nástroj Pundit v pokročilém srovnání nedopadl dobře. Potřebná funkcionality je sice přítomna, ale má přílišná omezení. Jeho bližší popis by dle mého názoru nebyl přínosný
- Výběr nástrojů pro srovnání:
 - V 1. kole nebyl prováděn v podstatě žádný výběr – všechny nalezené srovnatelné nástroje byly srovnány dle stanovených kritérií
 - Ve 2. kole byly nástroje ze srovnání z 1. kola seřazeny podle rozsahu implementace potřebné funkcionality. Následně bylo vybráno 5 nejvhodnějších, protože 6. nástroj na seznamu bylo alternativní GUR k 5. a 7. již neměl jednu ze základních potřebných funkcí
 - Ve 3. kole byly vybrány 2 nejvhodnější, aby uživatelé testovali 3 nástroje (4. měl navíc výrazná omezení)

- Experimenty v projektu DECIPHER byly naplánovány, avšak vzhledem k časové tísní v projektu nebyly provedeny dle původního plánu. Vedení experimentů, které vzhledem k uživatelům proběhly v zahraničí, se ujali projektoví partneři a nesesbírali či neposkytli potřebná data pro detailnější vyhodnocení

- Otazníky v tabulkách v přílohách značí informace, které z časových důvodů nebyly zjištěny či detailně zkoumány. Typicky se jedná o:
 - Proprietární komunikační protokoly nad HTTP, jejichž detaily by bylo nutné zjistit analýzou zdrojových textů programu
 - Proprietární formáty anotací a popisy fragmentů (resp. popisy vnitřních struktur programu využívaných pro ukládání anotací)
- V tabulce s pokročilým srovnáním u funkcí, které z určitých důvodů nebyly testovány:
 - Např. Domeo umožňuje využití zásuvných modulů, pomocí kterých by teoreticky mohlo být možné doplnit vnoření anotací

- Instrukce byly testerům předány verbálně (velmi krátké školení). Textový popis nebyl připraven
- V dotaznících byla i řada textových odpovědí. Některé části jsem vzhledem k výsledkům vyhodnocení nepovažoval za natolik přínosné, abych je zahrnul do textu práce. Příklady otázek z dotazníku jsou na následujícím snímku ...

- Je počet současně zobrazených alternativ vhodný (je to přehledné?) (1–10)?
- Jak moc bylo zjednodušování náročné (1–10)?
- Co by mohlo zrychlit zjednodušování?
- Jak moc Vám při anotování pomáhalo sémantické filtrování (1–10)?
- Jak moc Vás sémantické filtrování urychlilo (1–10)?
- Je anotování vztahů prostřednictvím atributů intuitivní (1–10)?
- Co pro Vás v GUR nástroje bylo nejvíce intuitivní?
- Co pro Vás v GUR nástroje bylo nejméně intuitivní (nelogické použití)?
- Co se Vám v GUR zdálo zbytečné?
- Co Vám v GUR chybělo?
- Jaké informace Vám chyběly v nabídkách anotací?
- Které informace byly v nabídkách anotací zbytečné?
- ...

- Pojem podtyp atributu neměl být zaveden (doslovně se v textu nevyskytuje). Myšlenkou daného vyjádření byla možnost nastavení typu hodnoty atributu na podtyp typu samotného atributu
- URI typu anotace obsahuje identifikátor skupiny uživatelů, neboť každá skupina sdílí svůj strom typů anotací. Při transformaci na linearizovaný název se vynechává, protože volba skupiny je v konfiguraci klienta
- Příklady použití nejsou zařazeny do ontologie, ale jsou uloženy v anotačním systému a znalostnímu inženýrovi umožňují lepší pochopení daného prvku ontologie a tím i jeho správné začlenění
- Komplexita importované ontologie by měla odpovídat profilu OWL 2 QL (mimo hlavní zaměření práce – zatím netestováno)

- Vertikalizace
 - Z korpusu CommonCrawl je extrahován prostý text (boilerplate removal – modifikovaný nástroj jusText), který je tokenizován a uložen do vertikálního formátu (1 token na řádek)

Time elapsed: 33 ms

[Text](#) | [Context](#) | [A](#) | [Link](#) artists including [Gerard Dou](#) , [Jan Steen](#) and [Gabriel Metsu](#) . References External lin
[Text](#) | [Context](#) | [A](#) | [Link](#) [Golden Age](#) genre painter in [the style](#) of [Gabriel Metsu](#) . Biography Biography.
[Text](#) | [Context](#) | [A](#) | [Link](#) founded the
[Text](#) | [Context](#) | [A](#) | [Link](#) painting a d
[Text](#) | [Context](#) | [A](#) | [Link](#) (c. [1628 - 16](#)
[Text](#) | [Context](#) | [A](#) | [Link](#) In [1665](#) he t
[Text](#) | [Context](#) | [A](#) | [Link](#) [Jan ! van !](#)
[Text](#) | [Context](#) | [A](#) | [Link](#) Haarlem [165](#)
[Text](#) | [Context](#) | [A](#) | [Link](#) [Oman](#). Hea
[Text](#) | [Context](#) | [A](#) | [Link](#) , [Pieter ! d](#)
[Text](#) | [Context](#) | [A](#) | [Link](#) , Volume I (J
[Text](#) | [Context](#) | [A](#) | [Link](#) This painting
[Text](#) | [Context](#) | [A](#) | [Link](#) after him, su
[Text](#) | [Context](#) | [A](#) | [Link](#) [1675](#); [Histor](#)
[Text](#) | [Context](#) | [A](#) | [Link](#) [Hans de Jo](#)
[Text](#) | [Context](#) | [A](#) | [Link](#) corset blue"
[Text](#) | [Context](#) | [A](#) | [Link](#) [Pieter ! de](#)

AnnotationWindow

File ▾ Edit ▾ View ▾ Format ▾ Tools ▾

Annotations <>

The Empty Glass The Empty Glass [The Empty Glass](#) (c. [1652](#)) is an oil on canvas painting by example of Dutch Golden Age painting and is now in the collection of the [Museum Boijmans](#) documented as a work by [Gabriel Metsu](#) by [Hofstede de Groot](#) in [1908](#) who wrote: "204b, A Officer. Sm. by 14 inches Heemskerck website in t

Suggestions (1)

Dutch (Art->Art_period_movement) painting of the Dutch Golden Age, a period in Dutch...

Name Dutch_Golden_Age_painting **C: 39%**

Visual_representation



Description

The Dutch Golden Age painting is the painting

1 2 [Next](#)

- Poslední experiment se zabýval kolaborativním rozšiřováním ontologie pro aspektově orientovanou analýzu sentimentu
- V 1. kroku příprava, ve 2. modifikace ontologie, aby vyjadřovala společné porozumění dané oblasti

nástroj	průměrný čas	
	individuálního kroku	kolaborativního kroku
WebProtégé	18.6 min.	15.3 min.
4A	19.1 min.	12.6 min.

- Dostupnost reálných dat anotovaných navrženou ontologií zrychlila řešení identifikovaných konfliktů a redukovala potřebu vysvětlovat a diskutovat kategorizaci každého jednotlivého aspektu