

Технологии Semantic Web в публикации памятников письменности

Алексей Варфоломеев,
Александр Иванов

Петрозаводск, 5 сентября 2012 г.

Литература и Интернет-ресурсы

- ACE 6.6 Construction Rules. http://attempto.ifi.uzh.ch/site/docs/ace_constructionrules.html
- Attempto Tools. <http://attempto.ifi.uzh.ch/site/tools/>
- Charters Encoding Initiative (CEI). <http://www.cei.lmu.de>
- Greengrass, M., Hughes, L., eds. *The Virtual Representation of the Past*. [S.I.]: Ashgate, 2008 (reprinted in 2010).
- Shotton, D., Portwin, K., Klyne, G. & A. Miles. Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article. *PLoS. Computational Biology*, 2009 5 (4). e1000361. <http://www.ploscompbiol.org/article/info%3Adoi%2F10.1371%2Fjournal.pcbi.1000361>
- Schreibmann, S., Siemens, R., Unsworth, J., eds. *A Companion to Digital Humanities*. [S. I.]: Blackwell Publishing. 2004.
- Text Encoding Initiative (TEI). <http://www.tei-c.org>
- de Waard, A. From Proteins to Fairytales: Directions in Semantic Publishing. *IEEE Intelligent Systems*. 2010. Vol. 25, Issue 2. Pp.83 - 88.

Публикации в Интернете

1. Базы данных сканированных источников (изображений) с метайнформацией о памятнике письменности

Codices Electronici Ecclesiae Coloniensis. <http://www.ceec.uni-koeln.de>

Codices Electronici Sangallenses. <http://www.cesg.unifr.ch>

Публикации в Интернете

2. Электронные тексты источников (транскрипции) в виде HTML-страниц, полнотекстовых реляционных баз данных или *XML*-документов

Манускрипт. Древние славянские памятники. <http://manuscripts.ru>

Model Editions Partnership (MEP). <http://adh.sc.edu>

Публикации в Интернете

3. Электронные тексты и исходные изображения памятников письменности

Древнерусские берестяные грамоты. <http://gramoty.ru>

Technique pour l'Historien En Ligne: Etudes, Manuels, Exercices (THELEME). <http://theleme.enc.sorbonne.fr/dossiers/>

Среда для совместной работы с текстами памятников письменности

- TextGrid (TextGrid Project).
<http://www.textgrid.de>
- Monasterium Project.
<http://monasterium.net>
- Проект «Источник»
<http://istochnik.karelia.ru>

Развитие Web

Статические документы



**Динамическое создание и редактирование контента
(CMS, Wiki, блоги)**



Социальные сети, Communities of practice



Semantic Web

Семантические слои в публикациях научных статей (Anita de Waard)

- термины и концепты, выделенные в статье и связанные с соответствующими тезаурусами, словарями, онтологиями
- «факты о тексте» (метаданные), представленные в форме триплетов (субъект-предикат-объект)
- формализованная репрезентация «научного дискурса»

Преимущества семантических публикаций памятников письменности

- эффективные средства для поиска и связывания информации
- базы знаний, порождающие новые знания (или гипотезы) с помощью процедур автоматического вывода

Полифункциональный прототип семантической публикации комплекса древнерусских актов

Субкомплекс смоленско-рижских актов XIII в. из
комплекса «Moscovitica-Ruthenica» (Латвийский
государственный исторический архив)

- LNA LVVA, 673 (Внешний архив рижского магистрата), 4, Kasten 18, № 1-268; Kasten 19, № 1-269; Kasten 20, № 1-227.
- LNA LVVA, 8 (Внутренний – «секретный» - архив рижского магистрата), 3, Capsula A, № 14-18, 41, 72; Capsula B, № 42; Capsula C, № 1-11, 23, 27, 34, 43.
- LNA LVVA, 8, 4, № 6-58)
- LNA LVVA, 214 (Архив коллегии ландратов Лифляндии), 6, № 114-116, 140.

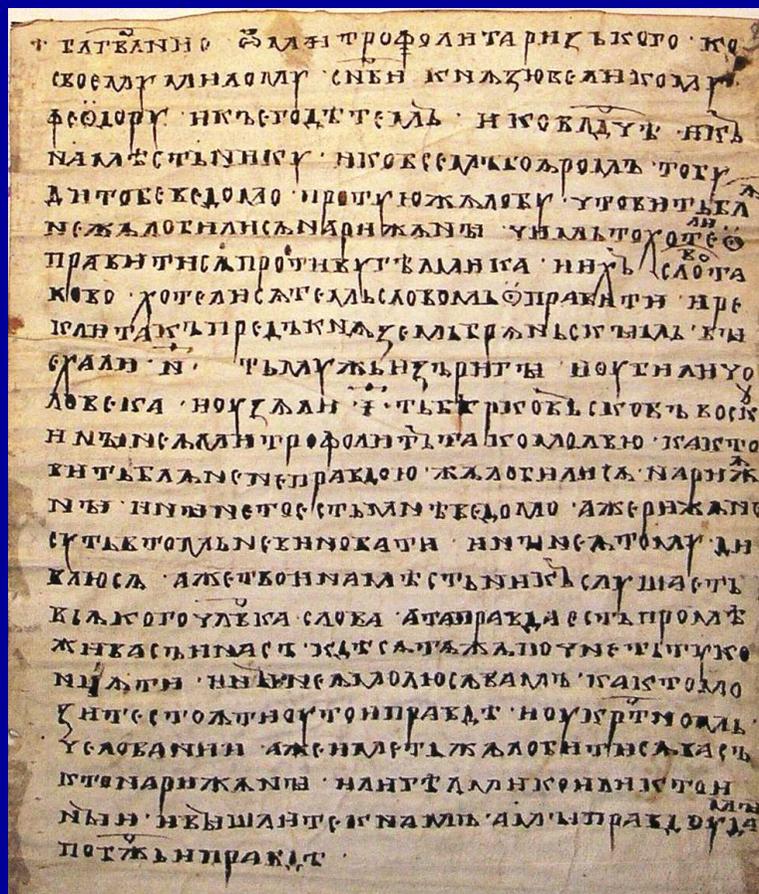
№ 6

Грамота архиепископа рижского Иоганна II смоленскому князю Федору Ростиславичу с обвинением жителей города Витебска в необоснованной жалобе на рижан

Рига. [1284-1297 гг., очевидно, ближе к 1285–1287 гг.]

Концепт (отпуск?). Пергамен, 16,5 × 23,3 см

LNA LVVA., 673, 4, K. 18, № 8, л. 3,



№ 4

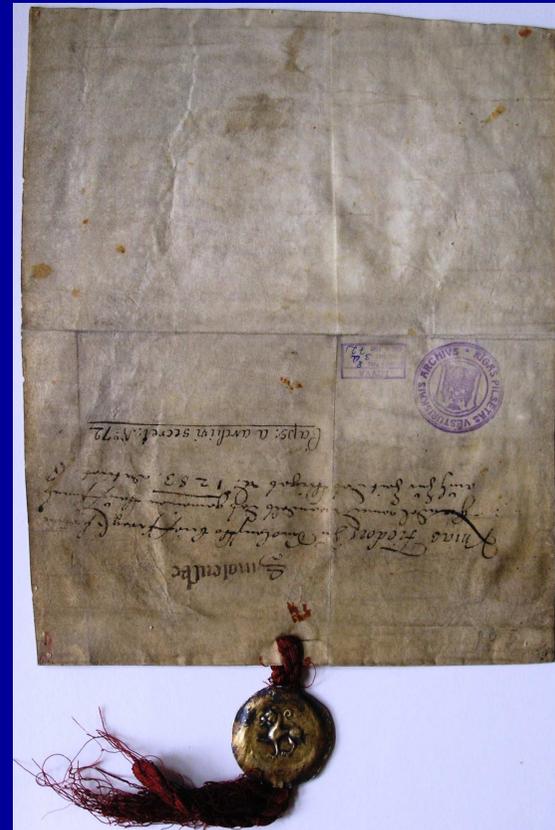
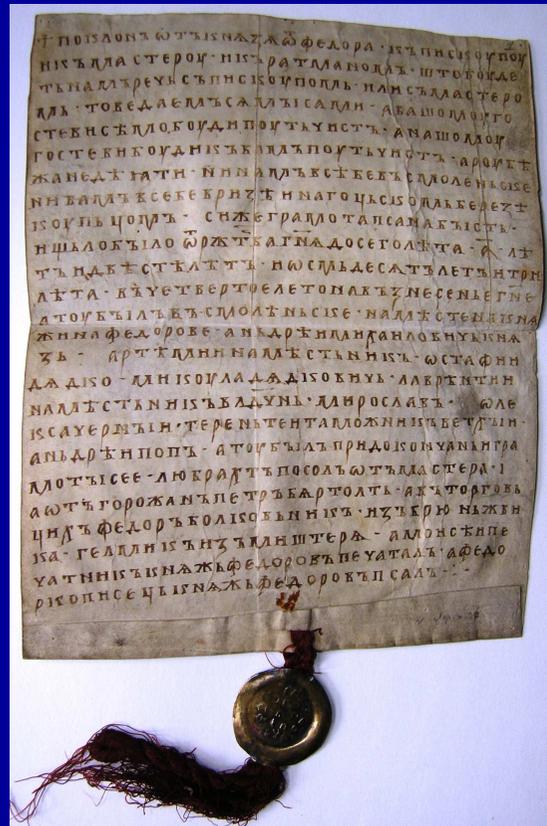
Грамота о предоставлении чистого пути немецким и русским купцам, написанная от имени смоленского князя Федора Ростиславича архиепископу рижскому Иоганну I, магистру

Ливонского ордена Виллекину и ратманам города Риги

Смоленск. 1284 г. [мая 18].

Подлинник. Пергамен, 20,0 × 24,5 см

LNA LVVA., 8, 3, Caps. A, № 72, л. 1

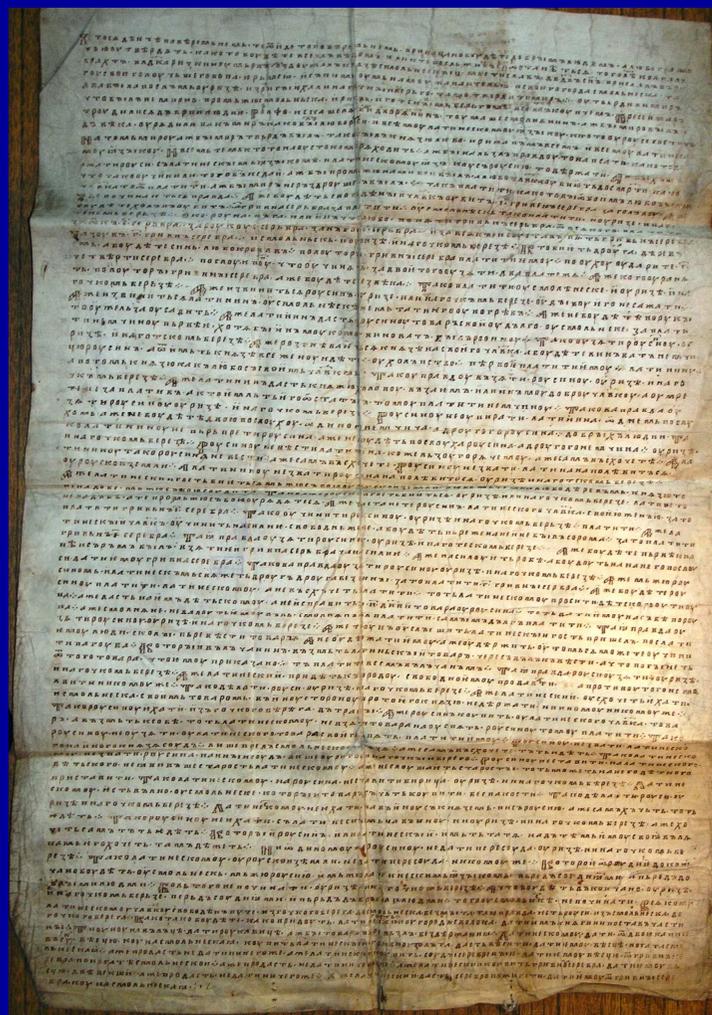


№ 3а

Списком А Готландской редакции договора 1229 г., заключенного смоленским князем Мстиславом Давидовичем с Ригой, Готландом и немецкими городами

[Готланд]. 1229 г.

Подлинник. Пергамен, 53,5 × 77,8 см
LNA LVVA., 8, 3, Caps. A, №.14, л. 1



№ 5

Грамота смоленского князя Федора Ростиславича по судному делу о немецком «колоколе» (гире)

Смоленск. 1284 г.

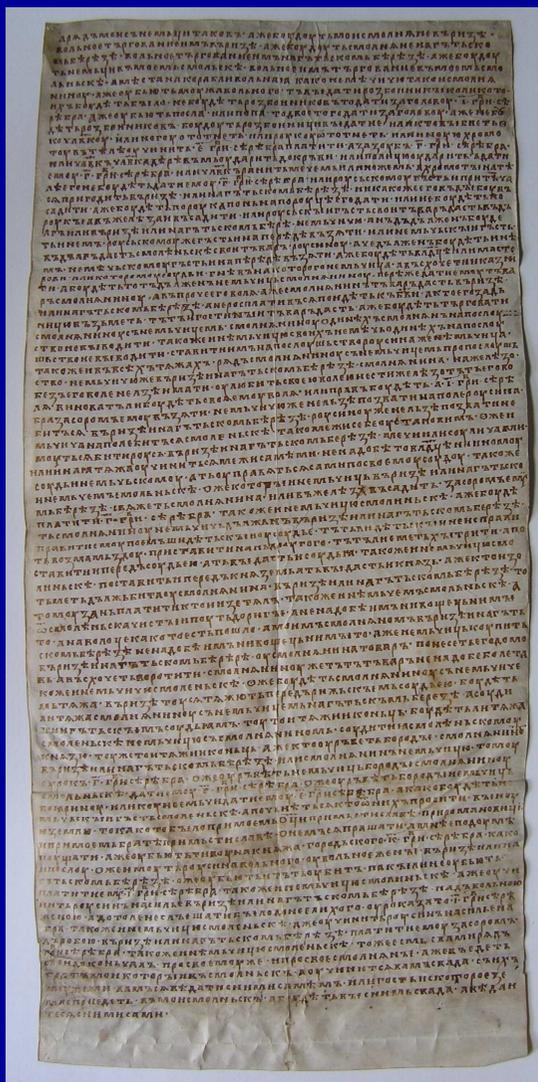
Подлинник. Пергамен, 13,0 × 18,3 см
LNA LVVA, 673, 4, К. 18, № 6, л. 3



Условия к договору Смоленска с Ригой и Готландом (так называемый

«договор неизвестного смоленского князя»)

Подлинник. Пергамен, 25,05 × 53,30 см
LNA LVVA, 8, 3, Caps. A, № 16, л. 1



Смоленск. [1219 – 1228/29 гг.]

1-й шаг

Дипломатическая транскрипция (№ 6)

`<tenor>`
† блѣвлне · ѿ мнтроф олнта рнзького · ко `<lb n="1"/>`
своему мнлому · сѣвн князю велнкому `<lb n="2"/>`
феѡдору · н къ его дѣтемъ · н ко влѣкъ н къ `<lb n="3"/>`
намѣстъннку · н ко всемъ бодромъ · то бу `<lb n="4"/>`
дн тебе ведомо · про тую жалобу · что внтѣбл`<super>ѧ</super>`
`<lb n="5"/>`
не жалобнлнса на рнжаны · чнмъ то

«Семантика» символов и способ их представления

- TEI:

```
<charDecl>
  <glyph xml:id="n1">
    <glyphName>SMALL CAPITAL CIRILLIC N LIKE LATIN N</glyphName>
    <mapping type="standardized">н</mapping>
  </glyph>
</charDecl>

.....

<g ref="#n1">N</g>
```

- Правила автоматической замены символов:
IF “N” THEN “н”
- Разработка специального Web-шрифта (например, SVG-font, в котором “N” употребляется вместо кириллической “н”)

2-й шаг

Создание семантического слоя с информацией об исторических «объектах» (CEI Markup Scheme)

```
<charter id="6">
<chDesc> ... </chDesc>
<tenor>
<protocol>
<invocatio> † </invocatio> <salutatio> бл҃гвлинне · ѿ </salutatio> <intitulatio>
<issuer> <roleName id="Archbishop"> мнтрополнта </roleName> рнзького · </issuer>
</intitulatio> <inscriptio> ко <lb n="1"/> своему мнлому · снвн <addressee>
<roleName id="Grand Prince"> кназю велнкому </roleName> <lb n="2"/> <persName
id="Fedor Rostislavich Prince of Smolensk"> феѡдору · </persName> </addressee> н къ
его дѣтемъ · н ко <roleName id="bishop of Smolensk"> влѣкъ </roleName> н къ <lb
n="3"/> <recipient> <roleName id="governor of Smolensk"> намѣстьннку · </roleName>
</recipient> н ко всемъ бодромъ · </inscriptio>
</protocol>
```

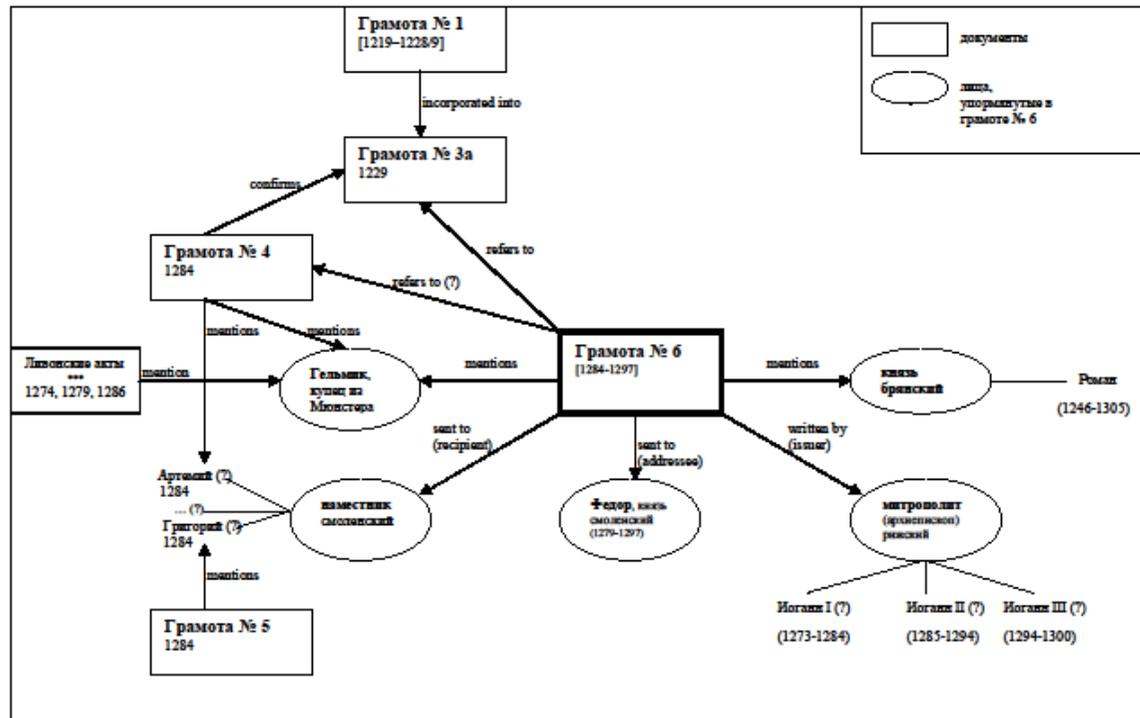
```
<context>
то су<lb p="4"/>дн тобѣ ведомо · про тую жалобу · что вѣтъблѣ<lb p="5"/>не жалобилнса на рижаны · унмь то хотѣли ѿ<lb
p="6"/>правнтнса протнву <persName id="Helmich merchant from Munster"> гѣлмнка </persName> · н нхъ слѣбѣ та<lb
p="7"/>ково · хотелн са темь словомь ѿправнтн · н ре<lb p="8"/>кн такъ предъ <roleName id="Prince of Briansk">
кнзземь брянскымь </roleName> вы<lb p="9"/>ехалн · ѿ · ть мужь нзъ <placeName> рнги </placeName> н ѿубилн ю<lb
p="10"/>ловека · н оузалн · ꙗ · ть <term id="unit of weight">бѣрковъсковъ </term> воскъ <lb p="11"/> н ныне а
<roleName id="Archbishop"> мнтрофолнтъ </roleName> тако молвю · как то <lb p="12"/> вѣтъблане неправдою ·
жалобилнса · на рнж<lb p="13"/>ны · н ныне то есть мнѣ ведомо · аже рижане <lb p="14"/> суть в томь невнноватн · н
ныне а тому · дн<lb p="15"/>влюса · аже твон <roleName id="governor of Smolensk"> намѣстъннкъ </roleName> слушаеть
<lb p="16"/> всакого улѣка · слова · а <refToDoc> та правда есть промѣ<lb p="17"/> жн васъ н насъ · кдѣ са тажа
поунеть ту ко <lb p="18"/> нцатн · </refToDoc> н ныне а молюса вамъ · как то мо<lb p="19"/>знте столтн оу
<refToDoc> тон правдѣ · н оу крѣтномь <lb p="20"/> человаѣннн · · </refToDoc> аже нметъ жалобнтнса васъ <lb p="21"/>
кто на рижаны · нлн <persName id="Helmich merchant from Munster"> гѣлмнко </persName> нлн кто н<lb p="22"/>нын · н
вы шлнте к намъ · а мы правдою даѣмъ <lb p="23"/> по бжѣн правдѣ · <lb p="24"/>
</context>
</tenor>
</charter>
```

3-й шаг
(вос)создание (дизайн)
семантической сети

Специализированные исторические онтологии (примеры)

- **The CIDOC** (le Comité International pour la Documentation) **Conceptual Reference Model**. <http://www.cidoc-crm.org/>
- Franklin D. Roosevelt Presidential Library, **Pearl Harbor Project**. (Ide, N. & D. Woolner. Historical Ontologies. In: Ahmad, K., Brewster, C. & M. Stevenson (eds.). *Words and Intelligence*. Vol. II: *Essays in Honor of Yorick Wilks*. Dordrecht: Springer, 2007, pp. 137–152.)
- **CultureSampo Project**, Finland. (Ahonen, E. & E. Hyvönen. Publishing Historical Texts on the Semantic Web – A Case Study. In: *Proceedings of the Third IEEE International Conference on Semantic Computing (ICSC2009)*. Berkeley (CA), 2009, pp. 167–173.)

Представление субкомплекса источников в виде семантической сети



4-й шаг

создание онтологии для описания
семантической сети и проблема выбора
«подходящей» технологии

RDF (Resource Descriptive Framework),
OWL (Ontological Web Language)

vs

Semantic Wiki-systems (SMW)

Передача содержания грамоты №6 на английском языке на сайте histdocs.referata.com

Be6 histdocs.referata.com/wiki/Charter_6   Искать в Google

[page](#) [discussion](#) [edit](#) [history](#) [delete](#) [move](#) [protect](#) [watch](#) [refresh](#) [Avarf](#) [my talk](#) [site settings](#) [admin links](#) [my preferences](#) [my watchlist](#) [my contributions](#) [log out](#)

Charter 6

Missive of archbishop of Riga Johann II von Vechten (?) to Fedor Rostislavich, prince of Smolensk, blaming inhabitants of Vitebsk for unjustified complaint against Rigans

Riga. [1284-1297, presumably 1285-1287]. Duplicate (draft document?). Parchment, 16,5 x 23,3 cm. Translated from Old Russian into English. Latvian National Archives - Latvian State Historical Archives, Riga (Latvijas Nacionlais arhivs - Latvijas Valsts vestures arhivs). Record group No. 673. Inventory No. 4. Kasten No. 18. File No. 8. Page 3 recto.

Metropolitan's of Riga benediction over our beloved sun, **Grand Prince Fedor**, and over his children, and over **the bishop of Smolensk**, and over the governor of Smolensk, and over all boyars. You should know about that complaint, which the inhabitants of lodged against Rigans in order to set themselves right with Helmich. And their words were as follows - they wanted to justify themselves with those words - and they said to prince of Briansk that 50 men had ridden out of Riga, and killed a man, and taken 10 berkovets of wax. And now I, metropolitan, say that those inhabitants of Vytebsk unjustly complained on Rigans. And now I know that Rigans are not guilty of that. And now I am surprised that the governor listen to anyone. But there is an agreement between you and us: if a controversy arouses, it should be settled between you and us. And now I pray that you would observe that agreement and oath: if anyone complains to you about Rigans, or Helmich, or somebody else, you should send a missive to us, but we will render justice in accordance with divine justice.

Facts about Charter 6  RDF feed 

- Mentions** [Helmich](#) +  and [Prince of Briansk](#) + 
- Probably refers to** [Charter 4](#) + 
- Refers to** [Charter 3a](#) + 
- Sent to** [Fedor Rostislavich, prince of Smolensk](#) + , [Bishop of Smolensk](#) +  and [Governor of Smolensk](#) + 
- Written by** [Archbishop of Riga](#) + 

Разметка текста грамоты №6 в SMW (Semantic MediaWiki)

```
[[wrtitten_by::archbishop of Riga|Metropolitan's of Riga]]
benediction over our beloved sun, [[sent_to::Fedor Rostislavich,
prince of Smolensk|Grand Prince Fedor]], and over his children,
and over [[sent_to::bishop of Smolensk|the bishop of Smolensk]],
and over [[sent_to::governor of Smolensk|the governor of
Smolensk]], and over all boyars. You should know about that
complaint, which the inhabitants of lodged against Rigans in
order to set themselves right with [[mentions::Helmich]]. And
their words were as follows - they wanted to justify themselves
with those words - and they said to [[mentions::prince of
Briansk]] that 50 men had ridden out of Riga, and killed a man,
and taken 10 berkovets of wax. And now I, metropolitan, say that
those inhabitants of Vytebsk unjustly complained on Rigans. And
now I know that Rigans are not guilty of that. And now I am
surprised that the governor listen to anyone. But there is
[[refers_to::Charter 3a|an agreement between you and us]]: if a
controversy arouses, it should be settled between you and us.
And now I pray that you would [[probably_refers_to::Charter
4|observe that agreement and oath]]: if anyone complains to you
about Rigans, or [[Helmich]], or somebody else, you should send
a missive to us, but we will render justice in accordance with
divine justice.
```

Helmich

Helmich is a merchant from Münster. He is mentioned in Charter 6 and Charter 4.

Facts about Helmich ⓘ

- Is a Merchant + 🔍
- Is from Münster + 🔍
- Is mentioned in Charter 6 + 🔍 and Charter 4 + 🔍

Charter 4

This charter confirms Charter 3a. It mentions Helmich and Artemii. It is probably referred in Charter 6.

Facts about Charter 4 ⓘ

- Confirms Charter 3a + 🔍
- Is probably referred in Charter 6 + 🔍
- Mentions Helmich + 🔍 and Artemii + 🔍

5-й шаг

ПОТЕНЦИАЛЬНЫЕ ВОЗМОЖНОСТИ РЕПРЕЗЕНТАЦИИ СОДЕРЖАНИЯ ТЕКСТОВ НА ОСНОВЕ КОНТРОЛИРУЕМЫХ ЕСТЕСТВЕННЫХ ЯЗЫКОВ

- CELT (Controlled English to Logic Translation)
- PENG (Processable English)
- Common Logic Controlled English
- Computer-Processable Language (CPL)
- Formalized English
- Attempto Controlled English (ACE)

*Workshop on Controlled Natural Language, Marettimo Island, Italy,
September 13-15, 2010*

Спасибо за внимание!