

Krzysztof Nowak
Pracownia Łaciny Średniowiecznej
Instytut Języka Polskiego PAN

Fontes Mediae et Infimae Latinitatis Polonorum

**Z prac nad korpusem polskiej łaciny
średniowiecznej**

Quis?

- Zespół: Pracownia Łaciny Średniowiecznej IJP PAN
 - dr hab. Michał Rzepiela, prof. IJP PAN – kierownik projektu
 - mgr Aleksandra Kulbicka
 - dr Anna Ledzińska
 - dr Agnieszka Maciąg
 - mgr Krzysztof Nowak – koncepcja projektu
 - dr Krzysztof Pawłowski
- Finansowanie: I edycja NPRH
- Czas: 2012-2016

Quid?

- Cechy korpusu
 - historyczny
 - czas: ok. 1000 – ok. 1550 r.
 - zasięg terytorialny: ziemie polskie X – poł. XVI w.
 - ogólny i zrównoważony (czy reprezentatywny?)
 - rozmiar: 5 mln wyrazów
- Pozostałe kryteria doboru tekstów
 - relacja do rzeczywistości (teksty fikcyjne / historyczne)
 - brak tekstów mówionych
 - rejestr językowy
 - charakter publiczności (teksty naukowe / popularne)
 - popularność w kulturze średniowiecza

Cui?

- Interfejs przeszukiwania
 - przeszukiwanie wg metadanych
 - konkordancje, KWIC, tablice kolokacji
- Adaptacja istniejących rozwiązań
 - koncentracja na danych
 - oszczędność środków
 - współpraca z innymi projektami
 - testowani kandydaci
 - Philologic (zarzucony)
 - XTF (oparty o Lucene)
 - TXM (desktop i TXM Web)
- Forma udostępnienia
 - darmowy dostęp
 - narzędzia wyszukiwania + pliki do pobrania (txt, TEI XML): sourceforge i inne repozytoria
- Adresat
 - językoznawcy, zwł. leksykografowie
 - literaturoznawcy i historycy
 - nauczyciele łaciny średniowiecznej

Quaestiones

- Precyzja datowania tekstów
- Proweniencja tekstów
 - podstawa ustalenia?
- Gatunkowe zróżnicowanie
- Dysproporcje
 - teksty wydane i ich cechy pochodne
 - centra intelektualne
 - przewaga tekstów XIV-XVI w.
- Prawa autorskie do edycji
- Cytowanie i uznawalność dzieł elektronicznych

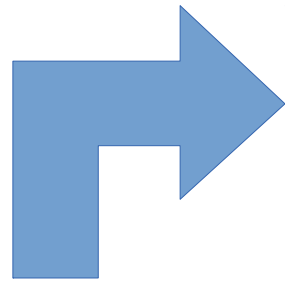
Helsinki Corpus of English Texts

Old English	Middle English	Early Modern English
Law	Law	Law
Document	Document	
Handbk. Astronomy	Handbk. Astronomy	
Handbk. Medicine	Handbk. Medicine	
	Handbk. Other	Handbk. Other
Science Astronomy		
	Science Medicine	Science Medicine
		Science Other

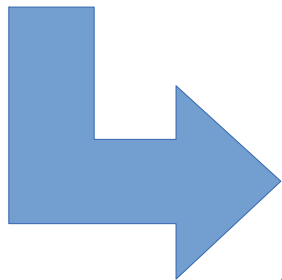
<i>statutory (STA)</i>	law, document*
<i>secular instruction (IS)</i>	handbook, science (astronomy*, medicine*), philosophy*, educational treatise*
<i>religious instruction (IR)</i>	religious treatise*, homily, rule, preface*, sermon
<i>expository (EX)</i>	science (astronomy*, medicine*, other), educational treatise*
<i>nonimaginative narration (NN)</i>	history, biography (saint's life, autobiography, other), religious treatise*, Middle English secular lyric*, travelogue*, diary
<i>imaginative narration (NI)</i>	fiction, romance, travelogue*, geography
*) not all representatives of the type of text in question belong to one and the same prototypical category	

Quomodo?

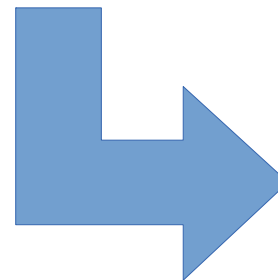
Akwizycja
tekstów



Korekta
ręczna
Strukturyzacja



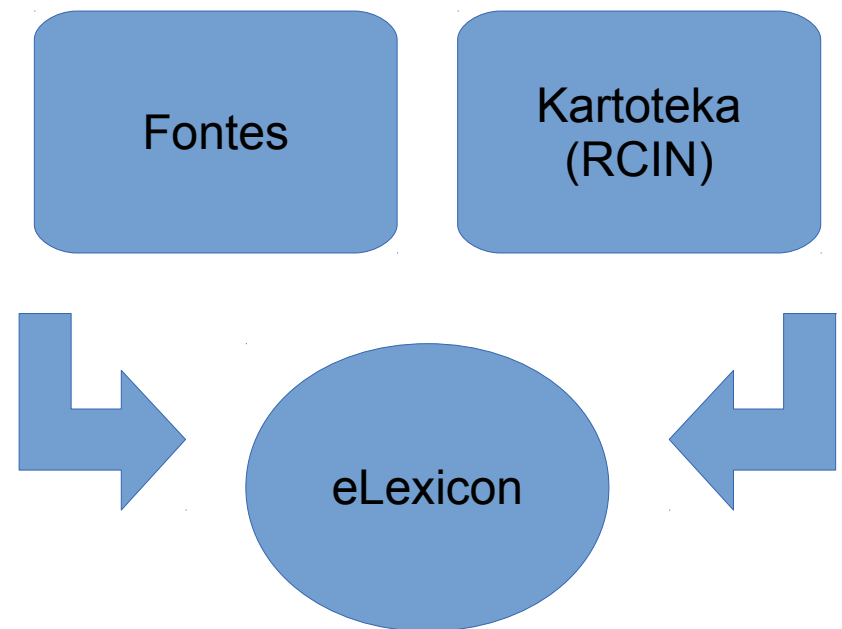
Korekta
automatyczna
Przetwarzanie
XSLT



„TElżacja”
Eksport

Quid nunc?

- Lemmatyzacja i anotacja PoS (glossaria.eu)
- Integracja z innymi narzędziami
- Korpus łaciny europejskiej



Plus?

Strona projektów Pracowni

<http://scriptores.pl>