





### Publications

[1] MJ. Meurs, C. Murphy, I. Morgenstern, G. Butler, J. Powlowski, A. Tsang, and R. Witte, "Semantic Text Mining Support for Lignocellulose Research" BMC MIDM, Vol.12 Suppl.1, 2012, to appear [2] H. Cunningham et al., "Text Processing with GATE (Version 6)" University of Sheffield, Department of Computer Science 2011

# TEXT MINING ASSISTANTS IN

Bahar Sateli<sup>1</sup>, Caitlin Murphy<sup>2</sup>, René Witte<sup>1</sup>, M 1.Semantic Software Lab. & 2.Centre For St

[3] R. Witte, T. Gitzinger, "Semantic Assistants – User-Centric Natural Language Processing Services for Desktop Clients" ASWC 2008, Bangkog, Thailand. Springer LNCS 5367, pp. 360–374 2009

ructural and Functional Genomics	
	Application
IediaWiki engine ki environment lied in the wiki content	<ul> <li>Templating mech</li> <li>Semantic metada</li> <li>Using the Semantic</li> </ul>
Sysop my talk my preferences my watchlist my contributions log out watch refresh	Content cellobiohydrolase E
roduced by Magnaporthe oryzae. [edit]	
oorthe grisea based on the complete genome sequence. Here, we demonstrate GH-6 family cellobiohydrolase (MoCel6A). In addition, the effect of cellobiose on stidine tag was overexpressed in M. oryzae and purified by affinity acid-swollen cellulose (PSC), β-glucan, and cellooligosaccharide derivatives A tandemly aligned cellulose binding domain (CBD) at the N terminus caused (tic domain only) showed decreased activity on cellulose. MoCel6A hydrolysis of was not inhibited by exogenously adding cellobiose up to 438 mM, which, rather,	Semantic En
(MoCel7A) was severely inhibited by more than 29 mM cellobiose. Furthermore, and Trichoderma reesei cellobiohydrolase (TrCel6A), which were prepared in used as a substrate in the presence of 292 mM cellobiose at pH 4.5 and pH 6.0, slightly increased hydrolysis at pH 4.5, and hydrolysis was severely inhibited at ctivities by cellobiose is dependent on the reaction mixture pH.	<ul> <li>Retrieve literatur</li> <li>Using SMW inlin</li> </ul>
1       Free PMC Article B         Privacy policy       About GenWiki       Disclaimers         Image: Second Secon	{{#ask: [[hasType::En   ?Enzyme = Enzyme   format = table   headers = plain   default = No pages   mainlabel = Page Na }}
oncordia.ca/GenWiki/index.php/PMID:_20709852_Abstra	Property:Enzyr
Clear	Minimit Page Name PMID: 20709852
terface	
	Evaluation
sistants Framework [3] .ns	<ul> <li>Methodology:</li> <li>Corpus of 15 doc</li> <li>Manually pre-fill</li> <li>Used mycoMINE</li> <li>Keeping the track</li> </ul>
	Average Curation
	Abstract
NLP Service Connector       Wiki       Ontologies	1 min
Service Invocation	
Service Information Language Service Descriptions	Conclusion
Semantic Assistants	GenWiki proves the inside a wiki is in need to <b>annotate</b> domain literature.

# Acknowledgments

Funding: Genome Canada & Génome Québec. Collaboration: Sherry Wu.



#### chanism to present the NLP results to users lata embedded in each page ntic MediaWiki (SMW) markup

Туре	Start	End	Features
Enzyme	103	120	<ul> <li>enzyme_alias: cellobiohydrolase</li> <li>BRENDA_SystematicName: oligoxyloglucan reducing-end cellobiohydrolase</li> <li>BRENDA_ECNumber: 3.2.1.150</li> <li>abbreviation_alias: -</li> <li>google_search: http://www.google.com /search?q=cellobiohydrolase &amp;</li> <li>BRENDA_RecommendedName: oligoxyloglucan reducing-end-specific cellobiohydrolase</li> </ul>

## ntity Retrieval

are that contain certain type of entities ine queries

Enzyme]] ne Entities Found

es found! Name

#### yme

M Enzyme Entitles Found
Cellobiohydrolase
Cellulases
endoglucanases
β-glucosidases
Invitrogen
DNA polymerase

ocuments

lled GeneWiki with full-text papers IE to automatically extract relevant entities ck of time

n Time: No Semantic Support vs. GenWiki

t Selection	
-------------	--

election	Full Paper	
GenWiki	No Support	Ge
20 sec.	37.5 min.	30.0
67% ↓		2(

GenWiki 30.6 min. 20% ↓

that seamless integration of NLP capabilities indeed efficient, in terms of the time curators retrieve entities of interest from the te and

