Personalized Semantic Assistance for the Curation of Biochemical Literature

Fedor Bakalov\textsuperscript{1}, Marie-Jean Meurs\textsuperscript{2}, Birgitta König-Ries\textsuperscript{1}, Bahar Sateli\textsuperscript{2}, René Witte\textsuperscript{2}, Greg Butler\textsuperscript{2} and Adrian Tsang\textsuperscript{2}

\textsuperscript{1} Friedrich Schiller University of Jena, Germany

\textsuperscript{2} Concordia University, Montréal, QC, Canada
**problem:** CURATION OF BIOCHEMICAL LITERATURE

finding relevant knowledge → of critical interest
- conducting research
- designing experiments

mining the literature → not that simple
- ever growing amount of scientific publications
- multiple repositories

We are looking for needles in a huge haystack...
context: CROSS-DISCIPLINARY COLLABORATION

presented work

▷ computer scientists
  user modeling
  natural language processing (NLP)
  software engineering

▷ biologists
  genomics
  molecular biology

users

▷ omics researchers

▷ biocurators
approach: WEB PLATFORM

- web platform supporting researchers
  - navigating literature
  - curating relevant papers

- single-point of access to abstracts
  - publications harvested from multiple DBs

- NLP services to process the literature
  - various NLP pipelines

- personalized view of the harvested papers
  - editable according to the user’s interests
System Architecture - Overview

- Domain Modeling Service (DMS)
- Domain Ontology
- Resource Management Service (RMS)
- Resource Metadata
- User Modeling Service (UMS)
- User Model
- Personalization Service (PS)
- Personalization Rules
- Personalizable Content Portlet
- Personalizable NLP-Results Portlet
- Introspective Views
- Client Side Abstraction Layer
- Semantic Assistants Framework

Domain Modeling
Resource Management
User Modeling
Personalization

GUI / Portal
Logic
Data / Rules

Introduction
Architecture
User Interface
Conclusion
System Architecture I

Domain Model Unit

- storing, accessing, managing the domain model
- domain model represented as an ontology (OWL)

Resource Management Unit

- harvesting literature
- annotating papers ↔ Semantic Assistants ↔ NLP pipelines
System Architecture II

User Modeling Unit

- storing information about users’ interests
- providing models, mechanisms, interfaces
- managing information about users required for adaptation

Personalization Unit

- storing personalization rules
- providing mechanisms for performing personalization
Cellulose percentage

Release behaviour of propranolol HCl from hydrophilic matrix tablets containing psyllium powder in combination with hydrophilic polymers. - 2011-11-28
Mohammad R. Sheibani- Shadrad, Kofi Asare-Adoo, Kokali Azizian, Davoud Hassanzadeh, Ali Nokhodchi

The objective of this study was to investigate the release behaviour of propranolol hydrochloride from psyllium matrices in the presence of hydrophilic polymers. The dissolution test was carried out at pH 1.2 and pH 6.8. Binary mixtures of psyllium and hydroxypropyl methylcellulose (HPMC) showed that an increase in the percentage of HPMC in the binary mixtures caused a significant decrease in the release rate of propranolol. Psyllium-HPMC matrices produced lower drug release compared to when the alginate was the matrix former alone. When sodium carboxy methyl cellulose (NaCMC) was incorporated into the psyllium, the results showed that matrices containing the ratio of psyllium-NaCMC in the 1:1 ratio are able to slow down the drug release significantly. As compared to matrices made from psyllium or NaCMC as a single agent, in-situ forming matrices showed a sigmoidal release rate. The double-layered tablets showed that the presence of HPMC in the outer shell of an inner formulation of psyllium alone had the greatest effect of protecting the inner core and thus producing the lowest drug release (Q10: 38%, MDT: 93 min). A significant decrease in the value of n in Q = k(t)n from 0.70 to 0.51 as the psyllium content was increased from 50 to 150 mg suggests that the presence of psyllium in HPMC matrices affected the release mechanism. Psyllium powder had the ability to combine with other hydrophilic polymers to produce controlled release profiles. Care and consideration should be taken when formulating hydrophilic matrices in different combinations.

Expression of a library of fungal β-glucosidases in Saccharomyces cerevisiae for the development of a biomass fermenting strain. - 2012-01-05
Caroline Wilde, Nicholas D. Gold, Nancy Bawa, José Humberto M. Tambor, Lina Mougharbel, Reginald Storms, Vincent J. J. Martin

Converting cellulose biomass to ethanol involves the enzymatic hydrolysis of cellulose and the fermentation of the resulting glucose. The yeast Saccharomyces cerevisiae is naturally ethanologenic, but lacks the enzymes necessary to degrade cellulose to glucose. Towards the goal of engineering S. cerevisiae for hydrolysis of and ethanol production from cellulose, 35 fungal β-glucosidases (BGL) from the BGL1 and BGL5 families were screened for their ability to be functionally expressed and displayed on the cell surface. Activity assays revealed that the BGL families had different substrate specificities, with only the BGL1s displaying activity on their natural substrate, cellulose. However, growth on cellulose showed no correlation between the specific growth rates, the final cell titers, and the level of BGL1 activity that was expressed. One of the BGLs that expressed the highest levels of cellulase activity, Aspergillus niger BGL1 (Antig-Bgl10), was then used for further studies directed at developing an efficient cellulose-fermenting strain. Expressing Antig-Bgl101 from a plasmid yielded higher ethanol levels when secreted into the medium rather than anchored to the cell surface. In contrast, ethanol yields from anchored and secreted Antig-Bgl101 were comparable when integrated on the chromosome. Flow cytometry analysis revealed that chromosomal integration of Antig-Bgl10 resulted in a higher percentage of the cell population that displayed the enzyme but with overall lower expression levels.
Query Portlet

- list of user search queries
- used by the portal to retrieve publications
- allows users to add, edit, delete queries
- allows to organize queries hierarchically

Click on a query → display a list of matching publications in the Listing portlet.
Listing Portlet

**Cellulose percentage**

Release behaviour of preground HCl from hydrophilic matrix tablets containing pivaloyl powder in combination with hydrophobic polymers.

Mohammad R. Shariat-Hashemi, Hili A. Asadi, Fatemeh Arzhang, Ali Nejadsahleh, Ali Khoshsadr

The objective of this study was to investigate the release behaviour of preground hydrochloric acid from pivaloyl matrices in the presence hydrophilic polymers. The dissolution test was carried out at pH 1.2 and 6.8. The release rates of pivaloyl and hydrochloric acid (HPIC) were shown that increased in the percentage of HPIC in the binary mixture caused a significant increase in the release rate of pivaloyl. Pivaloyl-polymer matrices produced a lower drug release as compared to the release from the matrix alone. When sodium carboxymethyl cellulose (NaCMC) was incorporated into the pivaloyl, the results showed that matrices containing the ratio of pivaloyl:NaCMC of 1:1 to 1:2 were able to slow down the drug release significantly as compared to matrices made from only pivaloyl or NaCMC, an important agent suggesting that there could be a synergistic effect between pivaloyl and NaCMC. The release decreased in the pH 1.2 from 0.70 to 0.51 as the pH of the pivaloyl content was increased from 50 to 100 mg, which suggests that the presence of pivaloyl in HPIC matrices affected the release mechanism. Pivaloyl powder had the ability in the combination with other hydrophilic polymers to control the release profile. Care and consideration should be made when formulating hydrophilic matrices in different combinations.

**Expression of a library of fungal & glucosidases in Saccharomyces cerevisiae for the development of a biomass fermenting strain.**

Caroline Helle, Nicholas D. Goff, Nancy Bawa, Joes Hambardzum Tsigem, Lisa Mugharab, Reginald Storms, Vincent J. Martis

Converting cellulose biomass to ethanol involves the enzymatic hydrolysis of cellulose and the fermentation of the resulting glucose. The most successful enzymes are naturally thermostable, and lack the enzymes necessary to degrade cellulose to glucose. Towards the goal of engineering S. cerevisiae for hydrolysis of both cellulose and ethanol production from cellulose, a library of microbial glucosidases (MGs) from the SLG and BGL families were screened for their ability to be functionally expressed and displayed on the cell surface. Activity assays revealed that the BGL families had different substrate specificities, with only the BGLs displaying activity on their natural substrate, cellulose. However, growth on cellulose showed no correlation between the specific growth rates, the final cell yield, and the level of BGL activity that was expressed. One of the BGLs that expressed the highest levels of cellulase activity, sugarcane rBG7 (temperature), was then used to further studies directed at developing an efficient cellulase fermenting strain. Expressing BgBp11 from a plasmid yielded higher enzyme levels than those secreted into the medium when anchored to the cell surface; in contrast, enzyme levels were increased and a second BgBp11 were comparable when integrated into the chromosome. Flow cytometry analysis revealed that the two copies of integrated BgBp11 resulted in a higher percentage of the cell population that displayed the enzyme but with lower overall expression levels.

---

dos request various types of semantic assistance
view a list of named entities extracted from the publications
access Semantic Assistants menu
decide how to display results
access personalization options
Personalization

personalization menu

▶ switch between personalized and standard views
▶ display personalization options and users interest profile
▶ personalization options vary from portlet to portlet

listing portlet example: 3 personalization effects

▶ sort publications according to user’s interest profile
▶ highlight most interesting publications by a color marker
▶ highlight items from the user interest profile in the list
Visualisation

Introspective Views

Listing > Portlet: Personalization Options & Interest Profile

- sort-publications
- color-publications-by-interest
- highlight-interesting-items
Introspective Views

- each zone: items of certain interest degree
- each slice: items of a specific type
- hot zone: items that users are strongly interested in
- cold zone: items that users are not interested in
- getting overview, zooming, filtering, navigation, and search
- display relevant content upon request
- editing information in the model: adding/deleting items, changing interest degree, organizing items by type, defining user-specific types...
Conclusion

What is done
▷ web platform supporting curation of biochemical literature
▷ single-point of access to publications
▷ process and analyze publications using semantic assistants
▷ personalized view of publications and NLP results

What is next
▷ evaluation: user study
▷ more publication sources (API?)
▷ integration of other textual contents
▷ components collaboration
▷ interaction user/semantic services
Acknowledgment

- Justin Powlowski for his expert advice on the biology content and the researchers of the Centre for Structural and Functional Genomics at Concordia University for their patience and their feedback.

- Genome Canada and Génome Québec for funding part of this work.

- Part of this research was sponsored by the IBM Ph.D. Fellowship Awards Program and carried out in the framework of the Minerva Portals project in cooperation with IBM Deutschland Research & Development GmbH.