Maximizing the Value of STM Content through Semantic Enrichment

Frank Stumpf
December 1, 2009
What is Semantics and Semantic Processing?

<table>
<thead>
<tr>
<th>Content</th>
<th>Knowledge Framework</th>
<th>Technology Framework</th>
<th>Search</th>
</tr>
</thead>
<tbody>
<tr>
<td>Text</td>
<td>Human expertise</td>
<td>Machine readable format</td>
<td>Knowledge Discovery</td>
</tr>
<tr>
<td>Images</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Codes</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Convergence of Content, Cognition and Technology
The ultimate objective is to maximize the “findability” of knowledge hidden in a maze of content.

<table>
<thead>
<tr>
<th>Process</th>
<th>Methodology</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic Web</td>
<td>HTML, XML, RDF</td>
<td>Data Relationships</td>
</tr>
<tr>
<td>Semantic Search</td>
<td>Metadata, Indexing</td>
<td>Information Extraction</td>
</tr>
<tr>
<td>Semantic Indexing Classification</td>
<td>Taxonomy, Ontology</td>
<td>Knowledge Structuring</td>
</tr>
<tr>
<td>Semantic Enrichment</td>
<td>Annotations, Links, Citations</td>
<td>Knowledge Discovery</td>
</tr>
</tbody>
</table>
Challenges to “Findability”

- Convergence of disciplines - Bio engineering
- Multiple applications - Green technologies
- Ambiguities in concepts - “Alcohol”: song, journal title, organic chemical

Explosion in the volume of content

- 5.5 million researchers worldwide
- 1.4 million articles are written annually by these researchers
- 1.7 million patents
- Expanding grey literature

Complex relationships

Increasing complexity of content

- Chemical structures ✓
- Genetic sequences ✓
- Maps ✓

- Hyperlinks

Minds @ Work
How Semantic Enrichment Improves “Findability”

- **Precision Search**: Filter irrelevant data
- **Depth Search**: Extract hidden meaning
- **Beyond Search**: Explore further
Techniques and Tools to Improve “Findability”

Metadata & Keywords → SEARCH

Abstracts & Reviews → SELECT

Citations Analysis → SEEK

Text & Data Mining → DISCOVER

SCOPE Enknowledge Center
A Quatra Group Company
Solutions - Automated

- Examples
  - Open Calais
  - Newssift

- Type of Content
  - Newsletters
  - Websites

- Domains
  - News
  - General content

- Technology Used
  - RDF
  - NLP

- Semantic Enrichment Process
  - Extraction of named entities
Solutions - Semi Automated

**Type of Content**
- Journal articles
- Technical literature

**Examples**
- Nature Chemistry
- RSC
- Elsevier - Cell

**Domains**
- Medical
- Life Sciences
- Chemistry

**Technology Used**
- Ontologies
- Text mining

**Semantic Enrichment Process**
- Extraction of technical data
- Indexing of concepts
Manual

Type of Content
- Patents
- Genetics literature
- Biomedical images

Semantic Enrichment Process
- Conceptual summaries

Technology Used
- Engineering
- Medical
- Life Sciences

Example
- Dialog
Semantic Enrichment

Semantic Mark-up
- Inserting conceptual codes in structured documents

Semantic Coding
- Annotating concepts with medical codes

Semantic Indexing
- Extracting concepts to index documents
- Indexing medical images with diagnostic metadata

Image Indexing
- Search, retrieve, cluster, visualize knowledge generated from a simple user query

Semantic Search & Discovery

- Hypertext links
How STM Publishers Can Exploit the Power of Semantics

- Use semantic technologies in production & delivery platforms
- Provide technology-enabled services that help STM content users discover actionable Knowledge from data repositories
- Use machine learning, linguistics, and semantic technologies to power comprehensive search, navigation, and discovery over all forms of information
- Contextualize services and products
- Support task-specific knowledge work

Content is no longer a static asset for STM publishers but an evolving gene with more and more intelligence
Thank you!

Comments / Questions / Enquiries:

fstumpf@scopeknowledge.com

www.scopeknowledge.com
Explosion in the volume of content

- 5.5 million researchers worldwide
- 1.4 million articles are written annually by these researchers
- 1.7 million patents
- Expanding grey literature

Increasing complexity of content

- Chemical structures
- Genetic sequences
- Maps

- Hyperlinks

Challenges to “Findability”
Challenges to “Findability”

5.5 million researchers worldwide

1.4 million articles are written annually by these researchers

1.7 million patents

Explosion in the volume of content

Increasing complexity of content

- Chemical structures
- Genetic sequences
- Maps

- Hyperlinks
Challenges to “Findability”

- Convergence of disciplines, e.g., bioengineering
- Multiple applications, e.g., green technologies
- Ambiguities in concepts, e.g., alcohol as a song, journal title, organic chemical

Explosion in the volume of content
- 5.5 million researchers worldwide
- 1.4 million articles are written annually by these researchers
- 1.7 million patents
- Expanding grey literature

Increasing complexity of content
- Chemical structures
- Genetic sequences
- Maps

Maps
- Hyperlinks

CLOSE
EBSCO unveils mobile phones version of EBSCOhost platform 12 Nov 2009

Electronic research databases provider EBSCO Publishing (EBSCO), US, has released EBSCOhost Mobile, allowing researchers to access their EBSCOhost databases via smartphones and other handheld devices. The product seeks to provide a convenient way for customers using mobile devices to find quality content. The main EBSCOhost Mobile screen offers a number of options including choosing which EBSCOhost database(s) to search plus setting search options, accessing field codes and specifying preferences.

All databases and services currently available on the EBSCOhost platform will be available via EBSCOhost Mobile. The EBSCOhost interface claims to be the most-used research platform currently available in more than 100,000 libraries around the world. Since the interface is the basis for EBSCO Discovery Service, EBSCOhost Mobile is expected to dramatically impact mobile searching.

Many of the existing EBSCOhost features such as search modes, limiting to full text, date ranges, peer-reviewed content or by publication are available. Users will also be able to search images from their mobile devices. The result list is scrollable and available data includes abstract, citation, and full text. The user can download PDFs and save search results offline.
The Calais initiative is about enabling semantic applications by providing a metadata generation web service, sample applications using that service to jumpstart development efforts, and support for developers.

**The Calais Web Service**

*The Calais web service automatically attaches rich semantic metadata to the content you submit. Using natural language processing, machine learning and other methods, Calais categorizes and links your document with entities (people, places, organizations, etc.), facts (person "x" works for company "y"), and events (person "z" was appointed chairman of company "y" on date "x").*

* The Calais viewer works with Firefox and Internet Explorer - other browsers may yield unpredictable results.

**Enter text here:**

---

**Minds @ Work**
The Calais initiative is about enabling semantic applications by providing a metadata generation web service, sample applications using that service to jumpstart development efforts, and support for developers.

The Calais Web Service

The Calais web service automatically attaches rich semantic metadata to the content you submit. Using natural language processing, machine learning, and other methods, Calais categorizes and links your document with entities (people, places, organizations, etc.), facts (person \( x \) works for company \( y \)), and events (person \( z \) was appointed chairman of company \( y \) on date \( x \)).

* The Calais viewer works with Firefox and Internet Explorer; other browsers may yield unpredictable results.

Enter text here:

EBSCO unveils mobile phones version of EBSCOhost platform – 12 Nov 2009

Electronic research databases provider EBSCO Publishing (EBSCO), US, has released EBSCOhost Mobile, allowing researchers to access their EBSCOhost databases via smartphones and other handheld devices. The product seeks to provide a convenient way for customers using mobile devices to find quality content. The main EBSCOhost Mobile screen offers a number of options including choosing which EBSCOhost database(s) to search plus setting search options, accessing field codes and specifying preferences.

Submit

Unstructured content
Automated Semantic Tagging

The Calais initiative is about enabling semantic applications by providing a metadata generation web service, sample applications using that service to jumpstart development efforts, and support for developers.

The Calais Web Service
The Calais web service automatically attaches rich semantic metadata to the content you submit. Using natural language processing, machine learning and other methods, Calais categorizes and links your document with entities (people, places, organizations, etc.), facts (person “x” works for company “y”), and events (person “z” was appointed chairman of company “y” on date “x”).

* The Calais viewer works with Firefox and Internet Explorer; other browsers may yield unpredictable results.

Enter text here:

EBSCO unveil mobile phones version of EBSCOhost platform – 12 Nov 2009
Electronic research databases provider EBSCO Publishing (EBSCO), US, has released EBSCOhost Mobile, allowing researchers to access their EBSCOhost databases via smartphones and other handheld devices. The product seeks to provide a convenient way for customers using mobile devices to find quality content. The main EBSCOhost Mobile screen offers a number of options including choosing which EBSCOhost database(s) to search plus setting search options, accessing field codes and specifying preferences.
Automated Semantic Tagging

Extracted entities

RDF output

Tagged output
Semi-Automated Semantic Enrichment

Input

Text Mining

Manual QA

Database

< HTML>
< HTML>
< HTML>
Interference-free determination of abscisic acid and gibberellin in plant samples using excitation-emission matrix fluorescence based on oxidation derivatization coupled with second-order calibration methods

Yuan-Na Li, Hai-Long Wu *, Jin-Fang Nie, Shu-Fang Li, Yong-Jie Yu, Shu-Rong Zhang and Ru-Qiu Yu

State Key Laboratory of Chemo/Biosensing and Chemometrics, College of Chemistry and Chemical Engineering, Hunan University, Changsha, 410082, China. E-mail: hlwu@hnu.cn; Fax: (+86) 731 88221818

Received 26th May 2009, Accepted 17th September 2009

First published on the web 6th October 2009

A sensitive excitation-emission fluorescence method with a second-order calibration strategy is proposed to simultaneously determine abscisic acid (ABA) and gibberellin (GA) contents in extracts of leaves and buds of ginkgo. The methodology is based on the alternating normalization-weighted error (ANWE) and the parallel factor analysis (PARAFAC) algorithms, which make it possible that the ABA and GA concentration can be attained in extract of plants even in the presence of unknown interference from potential interfering matrix contaminants introduced during the sample pretreatment procedure. Satisfactory recoveries were obtained although the excitation and emission profiles of the analytes were heavily overlapped with each other and the background in the extracts. The limits of detection obtained for GA and ABA in leaf samples were 9.6 and 6.9 ng mL\(^{-1}\), respectively, which were in the concentration range from hundreds to several ng g\(^{-1}\) for GA and ABA in leaves in different periods. Furthermore, in order to investigate the performance of the developed method, some statistical parameters and figures of merit of ANWE and PARAFAC are evaluated. The method proposed lights a new avenue to determine quantitatively phytohormones in extracts of plants with a simple pretreatment procedure, and may hold potential to be extended as a promising alternative for more practical applications in plant growth processes.

Introduction

Plant growth and development is regulated by internal signals and by external environmental conditions. Phytohormones
Interference-free determination of abscisic acid and gibberellin in plant samples using excitation-emission matrix fluorescence spectroscopy and second-order calibration methods

Yuan-Na Li, Hai-Long Wu, Jian-Ping Nie, Shi-Pang Li, Yang-Jie Yu, Shu-It Li

State Key Laboratory of Chemical/Biochemical and Chemometrics, College of Engineering, Hunan University, Changsha, 410082, China. E-mail: libyuan2002@gmail.com

Tel: (438) 731 3812/518

Received 26th May 2009, Accepted 17th September 2009
First published on the web 6th October 2009

A sensitive excitation-emission fluorescence method with a second-order calibration strategy is proposed to simultaneously determine abscisic acid (ABA) and gibberellin (GA) contents in extracts of leaves and buds of sugarcane. The methodology is based on the alternating normalization-weighted error (ANWE) and the parallel factor analysis (PARAFAC) algorithms, which make it possible that the ABA and GA concentration can be attained in extract of plants even in the presence of unknown interferences from potential interfering matrix contaminants introduced during the sample pretreatment procedure. Satisfactory recoveries were obtained although the excitation and emission profiles of the analytes were heavily overlapped with each other and the background in the samples. The limits of detection obtained for GA and ABA in leaf samples were 9.6 and 6.9 ng mL⁻¹, respectively, which were in the concentration range (from hundreds to several μg g⁻¹) for GA and ABA in leaves in different periods. Furthermore, in order to investigate the performance of the developed method, some statistical parameters and figures of merit of ANWE and PARAFAC are evaluated. The method proposed offers a new avenue to determine quantitatively phytohormones in extracts of plants with a simple pretreatment procedure, and may hold potential to be extended as a promising alternative for more practical applications in plant growth processes.

Introduction

Plant growth and development is regulated by internal signals and by external environmental conditions. Phytohormones
SUMMARY

Signaling proteins driving the proliferation of stem and progenitor cells are often encoded by proto-oncogenes. EphB receptors represent a rare exception; they promote cell proliferation in the intestinal epithelium and function as tumor suppressors by controlling cell migration and inhibiting invasive growth. We show that cell migration and proliferation are controlled independently by the receptor EphB2. EphB2 regulated cell positioning is kinase-independent and mediated via phosphatidylinositol 3-kinase, whereas EphB2 fos most kinase activity regulates cell proliferation through an Abl-cyclin D1 pathway. Cyclin D1 regulation becomes uncoupled from EphB signaling during the progression from adenoma to colon carcinoma in humans, allowing continued proliferation with invasive growth. The dissociation of EphB2 signaling pathways enables the selective inhibition of the mitogenic effect without affecting the tumor suppressor function and identifies a pharmacological strategy to suppress adenoma growth.

INTRODUCTION

Essential pathways regulating cell proliferation are often shared between stem/progenitor cells and cancer cells. This poses a problem as these pathways cannot be targeted to specifically eliminate tumor cells without simultaneously risking the depletion of untransformed cells, which is often a limiting factor in chemotherapy when doses that may eradicate tumor cells give unacceptable side effects. EphB receptors represent a rare exception in that they promote proliferation in the normal intestinal epithelium but, paradoxically, act as tumor suppressors in colon cancer development (Battie et al., 2005; Holmberg et al., 2008). How can the same protein drive proliferation in the normal situation and function as a tumor suppressor in the same tissue?

Eph receptors constitute the largest subgroup of receptor kinase receptors. Their ephrin ligands, which are either transmembrane proteins or attached to the cell membrane with a GPI anchor, are also capable of signaling. Eph receptors and ephrins are best known for their roles in controlling cell migration and axon guidance (Frisch, 2000), but have more recently been identified as regulators of stem and progenitor cell proliferation.
Asymmetric total syntheses of (+)- and (-)-versicolamide B and biosynthetic implications

Kenneth A. Miller1,2, Sachiko Tsukamoto2 and Robert M. Williams2

The Diels–Alder reaction is one of the most well-studied, synthetically useful organic transformations. Although it has been postulated that a significant number of naturally occurring substances arise by biosynthetic Diels–Alder reactions, rigorous confirmation of a mechanistically distinct natural Diels–Alderase enzyme remains elusive. Within this context, several related fungi within the Aspergillus genus produce a number of metabolites of opposite absolute configuration, including (+)- or (-)-versicolamide B. These alkaloids are hypothesized to arise via biosynthetic Diels–Alder reactions, implying that each Aspergillus species possesses enantiotomerically distinct Diels–Alderases. In this paper, experimental validation of these biosynthetic proposals via deployment of the intramolecular hetero-Diels–Alder reaction as a key step in the asymmetric total synthesis of (+)- and (-)-versicolamide B is described. Laboratory validation of the proposed biosynthetic Diels–Alder construction coupled with the secondary metabolite profile of the producing fungi, reveals that each Aspergillus species has evolved enantiomerically distinct indole oxidases, as well as enantiomerically distinct Diels–Alderases.
Asymmetric total syntheses of (+)- and (-)-versicolamide B and biosynthetic implications

Korinna A. Miller, Sachiyo Tsukamoto and Robert M. Williams

Highlighting tool

The Diels-Alder reaction is one of the most well-studied, synthetically useful organic transformations. Although it has been posited that a significant number of naturally occurring substances arise by biosynthetic Diels-Alder reactions, rigorous confirmation of a mechanistically distinct natural Diels-Alderase enzyme remains elusive. Within this context, several related fungi within the Aspergillus genus produce a number of metabolites of opposite absolute configuration, including (+)- or (-)-versicolamide B. These alkaloids are hypothesized to arise via biosynthetic Diels-Alder reactions, implying that each Aspergillus species possesses enantiotomerically distinct Diels-Alderases. In this paper, experimental validation of these biosynthetic proposals via deployment of the intramolecular hetero-Diels-Alder reaction as a key step in the asymmetric total syntheses of (+)- and (-)-versicolamide B is described. Laboratory validation of the proposed biosynthetic Diels-Alder construction, coupled with the secondary metabolite profile of the producing fungi, reveals that each Aspergillus species has evolved enantiotomerically distinct fumoxide oxidases, as well as enantiotomerically distinct Diels-Alderases.

A seemingly limitless array of natural secondary metabolites are produced in nature from plants, microorganisms and animals in both terrestrial and marine environments. In the overwhelming majority of cases, these secondary metabolites, commonly referred to as "natural products", are produced as optically pure forms (when chiral). In the plant kingdom, examples of different species producing enantiotomerically opposite metabolites are well documented but rare. Some of the most well-known examples are terpenes, such as camphor, camphene and limonene. These substances arise from an initial enzyme-catalyzed oxidation of geranyl geranyl diphosphate, and the observation of antipodes in terpenes in nature has been attributed to enzymatic reactions specific to different species which often display enantioselectivity.
Asymmetric total syntheses of (+)- and (-)-versicolamide B and biosynthetic implications

Kenneth A. Miller, Satoshi Tsukamoto and Robert M. Williams

The Diels-Alder reaction is one of the most well-studied, synthetically useful organic transformations. Although it has been postulated that a significant number of naturally occurring substances arise by biosynthetic Diels-Alder reactions, rigorous confirmation of a mechanistically distinct natural Diels-Alderase enzyme remains elusive. Within this context, several related toxins within the *Aspergillus* genus produce a number of metabolites of opposite absolute configuration, including (+)- or (-)-versicolamide B. These alkaloids are synthesized to arise via biosynthetic Diels-Alder reactions and possess enantioselectivity in their intramolecular hetero-Diels-Alder reaction. Here, we report total syntheses of (+)- and (-)-versicolamide B and describe, experimentally and computationally, the biosynthetic basis of this natural stereochemical discrimination. Asymmetric total syntheses of (+)- and (-)-versicolamide B allows the development of new, more biologically active analogues, with the potential to become lead compounds in the treatment of important diseases.
Search our library

Advanced Search Options

BNF Category:

- None selected
- 01. Gastrointestinal system
- 02. Cardiovascular system
- 03. Respiratory system
- 04. Central nervous system
- 05. Infections
- 06. Endocrine system
- 07. Obstetrics, gynaecology and urinary tract disorders development
- 08. Malignant disease and immunosuppression
- 09. Nutrition and blood
- 10. Musculoskeletal and joint diseases
- 11. Eye
- 12. Ear, nose and oropharynx
- 13. Skin
- 14. Immunological products and vaccines
- 15. Anaesthesia
- Non-clinical

Semantic mark-up
WO 2000030778 A1

TITLE: Self-healing roll for surface conditioning of sheets, e.g. metal sheets, has non-woven web elements comprising entangled fibers held together by a bonding agent

NOVELTY - A self-healing article e.g. in the form of roll (21) comprises several compacted stacked web elements (22) having entangled fibers bonded together at points of mutual contact by a bonding agent. The article is resistant to an oxidizing agent and has a Shore A hardness of 70-93 and a void volume of 2-30%.

USE - For surface conditioning of sheets, e.g. metal sheets.

ADVANTAGE - The invention provides a self-healing article resistant to oxidizing agents having an increased life span. If used, results in fewer roll replacements and unscheduled production line downtimes. Chances of chemical contamination between treating solutions are also minimized.