Text & Data Mining (TDM)

Dr. Haralambos Marmanis
CTO / VP, Engineering
Copyright Clearance Center
What do we do?

Rightsholders

600+ million rights from
- Publishers
- Authors
- Creators

Content Users

- Licensing Solutions
- Rights Management
- Content Delivery
- Copyright Education

- 35,000 companies
- Employees worldwide
- Users in 180 countries
Overview of TDM

• What is it? How is it done?
• Why people care?
• An example: Drug discovery process
• Practical challenges
• CCC Solution Walkthrough
What is TDM?

The Discovery of (new) knowledge from a set of natural language resources (e.g. articles, books) and other structured or unstructured data
How is it done?

1. Identify a set of resources that are relevant to a particular research objective

2. Analyze and extract information specific to the research objective

3. Develop and explore the various relations between extracted objects of interest
Why people care?

Biomarker discovery
Drug repurposing
Drug safety
Competitive intelligence
Sentiment analysis

Current Document: pr003.txt  Document Sentiment Score: 0.3927253315949234
“Drug Discovery” Process

**Goal**: Develop new treatments for diseases through hypothesis formation.

**Methodology**:
- Keyword/Database Searching
- Review Literature
- Find relationships
- Generate hypotheses
- Test
- Product development
- Etc.
Literature-based Discovery

Raynaud's Phenomenon

Ariane Herrick, MD, FRCP
Corresponding author
Assistant Attending, Department of Rheumatology,
University of Washington, Seattle, WA, USA.
Email: ariane.herrick@FredHoffman.org
Current address: Department of Rheumatology,
University of Washington, Seattle, WA, USA

Opinion statement

Raynaud's phenomenon (RP) is a clinical syndrome characterized by digital ischemia, often precipitated by cold exposure, stress, or emotional upset. It is characterized by a triphasic response involving digital vasospasm, pallor, cyanosis, and hyperemia. The syndrome is diagnosed based on clinical features, without the need for specific laboratory tests. The clinical diagnosis is made when digital ischemia occurs in response to cold exposure or stress.

Introduction

Mild Raynaud described the phenomenon of local anesthetic of the periphery in 1862 [1]. The clinical triphasic color change in white (palor), blue (cyanosis), and red (flushing) was first described by Lhermitte and Joubert in 1889 [2]. Raynaud's phenomenon can occur as a part of a number of different conditions, including connective tissue disorders such as systemic sclerosis (SSc), systemic lupus erythematosus (SLE), dermatomyositis, scleroderma, and other connective tissue diseases. It can also occur as a complication of conditions such as hypothyroidism, hypervitaminosis A, and connective tissue disorders such as systemic sclerosis (SSc).

Raynaud's phenomenon is considered a chronic vasospastic disorder. The primary target of the disease is the digital vessels, affecting primarily the small and medium-sized arteries. The disease is often characterized by intermittent digital ischemia, which is often triggered by cold exposure, stress, or emotional upset.

Raynaud's phenomenon can be classified into three types: primary, secondary, and mixed. Primary Raynaud's phenomenon occurs in the absence of other underlying conditions, whereas secondary Raynaud's phenomenon is associated with underlying systemic diseases such as SSc, systemic lupus erythematosus (SLE), and dermatomyositis.

The mechanism of Raynaud's phenomenon is not fully understood. It is thought to involve a complex interplay between the autonomic nervous system, the immune system, and the vasculature. The exact cause of the vasospasm is not known, but it is thought to involve the activation of sympathetic nervous system, which leads to the release of norepinephrine and serotonin, resulting in vasoconstriction.

The treatment of Raynaud's phenomenon is based on individual patient needs and may include lifestyle modifications, medications, and surgical interventions. Lifestyle modifications such as avoiding cold exposure and stress management can be effective in mild cases. Medications such as calcium channel blockers, alpha-blockers, and antiplatelet agents can be used to treat more severe cases. Surgical interventions such as sympathectomy may be considered in refractory cases.

References


Acknowledgments

BK: Raynaud's phenomenon.

Funding

This research was supported by the National Institutes of Health (NIH) grant number K07 AR072958.

Conflict of interest

The authors declare that they have no conflict of interest.

Ethical approval

This study was approved by the institutional review board of the University of Washington.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Appendix A

Supplementary material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.1016/j.molphen.2021.05.001.

Figure 1: The mechanism of Raynaud's phenomenon

Figure 2: Raynaud's phenomenon in a patient with systemic sclerosis

Figure 3: The effect of calcium channel blockers on digital blood flow
Too Much Research

53M Records in Scopus
800,000 Journal Articles published per year
Even within a single research area

Lots of disorders …
- Angina
- Acute coronary syndrome
- Alexia
- Anomic aphasia
- Aortic dissection
- Aortic regurgitation
- Aortic stenosis
- Apoplexy
- Apraxia
- Arrhythmias
- Asymmetric septal hypertrophy (ASH)

Lots of documents…
- 35,000+ on Improve Circulation
- 7,000+ per disease area

Atherosclerosis
- Atrial flutter
- Atrial septal defect
- Atrioventricular canal defect
- Atrioventricular septal defect
- Avascular necrosis
More practical challenges

- Many sources of content
- Many formats
- Difficult to obtain **full-text** in XML
- Difficult to **integrate** content into TDM software.
- Hard to negotiate and manage **licenses** and feeds from all publishers.
Data Processing Workflow

- Individual documents or corpora
- Information retrieval
- Collection of papers to be mined
- Normalised documents
- Copied & annotated
- Derived dataset capable of semantic interrogation
- Textual analysis: lexical, entity identification
- Semantic query
- Discovered Knowledge
- Extracted Information

Information Retrieval

Knowledge Discovery

Palantir

IBM. WATSON

Linguamatics
Solving the retrieval problem

A single retrieval platform:

- **Speed up** time to obtain properly licensed content for text mining
- **Discover** and **download full-text in XML**, not just abstracts
- **Main corpus** includes Subscribed and Not-Subscribed content
- **Normalize** XML format across many publishers
- Provide a **Web UI** and **RESTful API** services
1. Publishers provide content and rights

2. Researchers create content sets by using search or other discovery criteria

3. Researchers slice and dice results and identify an appropriate corpus for their project

4. XML corpus can be imported into various TDM tools
## ADMIN PORTAL

<table>
<thead>
<tr>
<th>Journal</th>
<th>ISSN</th>
<th>Platform</th>
<th>Start Date</th>
<th>End Date</th>
<th>Months Back</th>
<th>Embargo (months)</th>
<th>Mechanism</th>
<th>Date Added</th>
<th>Actions</th>
</tr>
</thead>
</table>
### Your Projects

<table>
<thead>
<tr>
<th>Creation Date</th>
<th>Project</th>
<th>Description</th>
<th>Actions</th>
<th>Status</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>19 Nov 2014</td>
<td>Her2 Breast Cancer</td>
<td></td>
<td>View options</td>
<td>Completed</td>
<td>View</td>
</tr>
<tr>
<td>19 Nov 2014</td>
<td>NPC1</td>
<td></td>
<td>View options</td>
<td>Completed</td>
<td>View</td>
</tr>
<tr>
<td>18 Nov 2014</td>
<td>testing build on 11/16 - cancer</td>
<td></td>
<td>View options</td>
<td>Completed</td>
<td>View</td>
</tr>
<tr>
<td>13 Nov 2014</td>
<td>Breast cancer demo</td>
<td></td>
<td>View options</td>
<td>Completed</td>
<td>View</td>
</tr>
<tr>
<td>10 Nov 2014</td>
<td>demo test cancer</td>
<td></td>
<td>View options</td>
<td>Open</td>
<td></td>
</tr>
<tr>
<td>07 Nov 2014</td>
<td>NPC1 - second try</td>
<td></td>
<td>View options</td>
<td>Preparing</td>
<td>download</td>
</tr>
<tr>
<td>03 Nov 2014</td>
<td>Hemagglutinin</td>
<td>Project preview says 140 results</td>
<td>View options</td>
<td>Ready for download</td>
<td>View</td>
</tr>
<tr>
<td>03 Nov 2014</td>
<td>Cancer w date range</td>
<td>added date range</td>
<td>View options</td>
<td>Completed</td>
<td>View</td>
</tr>
<tr>
<td>23 Oct 2014</td>
<td>Alzheimers</td>
<td>A project to create a description that describes how this is a project</td>
<td>View options</td>
<td>Preparing</td>
<td>download</td>
</tr>
<tr>
<td>17 Oct 2014</td>
<td>Diabetes</td>
<td></td>
<td>View options</td>
<td>Completed</td>
<td>View</td>
</tr>
</tbody>
</table>

1 - 10 of 10
CREATE PROJECT

Project name:
Polio

Project description: (Optional)
Find all documents with mention of Polio (Polio myelitis)

SEARCH RETRIEVAL METHOD

Search query analysis Nearest neighbor analysis Article ID list

Enter your search query:

- Full Text
  - AND
    - Polio
  - Full Text

Search citations

Show limiters

Do you want to view a preview of the article before requesting the full text?

View Preview

OR, do you want to go ahead and get the full text associated with the matches to your query?

Get Full Text

stm
PROJECT PREVIEW

Number of Results by Publisher

- Publisher A
- Publisher B

Number of Results by Year

- 2009
- 2010
- 2011
- 2012

Number of Results by Journal

<table>
<thead>
<tr>
<th>Journal</th>
<th>ISSN</th>
<th>Rightsholder</th>
<th>Number of results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Future Cardiology</td>
<td>1479-6678</td>
<td>Rightsholder 1</td>
<td>364</td>
</tr>
<tr>
<td>Pharmacogenomics</td>
<td>1452-2416</td>
<td>Rightsholder 2</td>
<td>352</td>
</tr>
<tr>
<td>Aging Health</td>
<td>1745-509X</td>
<td>Rightsholder 3</td>
<td>332</td>
</tr>
<tr>
<td>Clinical Lipidology</td>
<td>1758-4299</td>
<td>Rightsholder 1</td>
<td>275</td>
</tr>
<tr>
<td>Diabetes Management</td>
<td>1758-1907</td>
<td>Rightsholder 2</td>
<td>268</td>
</tr>
<tr>
<td>Regenerative Medicine</td>
<td>1746-0731</td>
<td>Rightsholder 3</td>
<td>265</td>
</tr>
<tr>
<td>Clinical Lipidology</td>
<td>1758-4302</td>
<td>Rightsholder 3</td>
<td>254</td>
</tr>
</tbody>
</table>
Age and psychological influences on immune responses to trivalent inactivated influenza vaccine in the meditation or exercise for preventing acute respiratory infection (MEPARI) trial

Hayney, Mary S; Coe, Christopher L; Muller, Daniel; Obasi, Chidi N; Backo, Uba; Ewers, Tola; Barrett, Bruce, Human Vaccines & Immunotherapeutics, 2014 Jan 01, Vol. 10 Issue 1, pages 83-81

ISSN: 2164-9515
DOI: n/a
Publisher: DEFAULT_TDM
Language: (English)

the groups is noteworthy. The parent study, from which the current data were derived, showed that both exercise and MBSR reduced the number of days sick from all-cause acute respiratory infection (ARI) illness.22 The incidence, duration and global severity...

Direct inhibition of hexokinase activity by metformin at least partially impairs glucose metabolism and tumor growth in experimental breast cancer

Marini, Cecilia; Salanti, Barbara; Massolo, Michela; Amaro, Adriana; Esposito, Alessia Isabella; Orego, Anna Maria; Capitanio, Selene; Emonti,
Filter your results:

Project: Diabetes test 2

QUERY: Full Text: Diabetes

Add Additional criteria to refine these results further

Results: 1 - 100 of 399

Age and psychological influences on immune responses to trivalent inactivated influenza vacc the meditation or exercise for preventing acute respiratory infection (MEPARI) trial
Time’s Up!

About your speaker:
Name: Dr. Haralambos Marmanis
Company: Copyright Clearance Center
Tel:  +1 978.646.2723
Email: hmarmanis@copyright.com
Social Media:
https://www.linkedin.com/in/marmanis
https://twitter.com/marmanis
http://www.amazon.com/H.-Marmanis/e/B002BOA8O6