Ontology Supported Knowledge Discovery in the Field of Human Performance and Cognition

Christopher Thomas  
*Wright State University - Main Campus*

Pablo N. Mendes  
*Wright State University - Main Campus*

Delroy H. Cameron  
*Wright State University - Main Campus, cameron.20@wright.edu*

Amit P. Sheth  
*Wright State University - Main Campus, amit.sheth@wright.edu*

Krishnaprasad Thirunarayan  
*Wright State University - Main Campus, t.k.prasad@wright.edu*

*See next page for additional authors*

Follow this and additional works at: [https://corescholar.libraries.wright.edu/knoesis](https://corescholar.libraries.wright.edu/knoesis)  
Part of the Bioinformatics Commons, Communication Technology and New Media Commons, Databases and Information Systems Commons, OS and Networks Commons, and the Science and Technology Studies Commons

Repository Citation  
https://corescholar.libraries.wright.edu/knoesis/974

This Presentation is brought to you for free and open access by the The Ohio Center of Excellence in Knowledge-Enabled Computing (Kno.e.sis) at CORE Scholar. It has been accepted for inclusion in Kno.e.sis Publications by an authorized administrator of CORE Scholar. For more information, please contact corescholar@www.libraries.wright.edu, library-corescholar@wright.edu.
Authors
Christopher Thomas, Pablo N. Mendes, Delroy H. Cameron, Amit P. Sheth, Krishnaprasad Thirunarayan, and Cartic Ramakrishnan

This presentation is available at CORE Scholar: https://corescholar.libraries.wright.edu/knoesis/974
Ontology supported Knowledge Discovery in the field of Human Performance and Cognition

Kno.e.sis Center  http://knoesis.org
Wright State University

C. Thomas, P. Mendes, D. Cameron, A. Sheth, TK Prasad
Thanks, C. Ramakrishnan
Motivation

• Speed up the “search” part of research
• Develop a system that helps in focused exploration of a domain
Procedural memory in recent-onset Parkinson’s disease.


Experimental Psychology Laboratory, Université de Haute-Bretagne, Rennes, France.

Parkinson’s disease is accompanied by cognitive disorders which may affect procedural memory. Procedural memory uses a specific knowledge resource that expresses itself through pre-established acting procedures. The aim of this study was to better define the characteristics of procedural memory, first of all, by trying to determine the level of involvement of that memory in the acquisition process (during learning and/or during procedure maintenance), then by specifying the effect of the type of resource involved (verbal or motor). To achieve this, we compared the mnemonic performances of 20 recent-onset parkinsonian patients with those of 20 healthy controls, using two memory tasks with a fixed rule (poetry, visuomotor tracking). Result analysis revealed that parkinsonian patients had more difficulty than controls in learning the two rules, regardless of the material involved. Their difficulties were often associated with an impairment of executive functions, and the procedural memory problems described in parkinsonian patients are linked to the involvement of these resources in the various tasks.

PMID: 10036333 [PubMed - indexed for MEDLINE]
INTRODUCTION: Patients submitted to surgery to treat a brain aneurysm, who have suffered a subarachnoid haemorrhage, sometimes present cognitive disorders that can affect their social, familial, academic or occupational relationships. Memory disorders are frequent, although other cognitive functions may also be affected. AIMS: The purpose of this research work was to study performance in logical verbal memory and visual-constructual memory in subjects following a surgical intervention (at least a year ago) to treat an aneurysm in the territory of the anterior circulation of the brain. We also wanted to analyse whether the location of the aneurysm in the brain had any effect on memory performance.

PATIENTS AND METHODS: We examined a sample of 24 adult subjects of both sexes, with no previous history of cognitive or psychiatric disorders, who had undergone surgical treatment of brain aneurysms in the middle cerebral, anterior communicating and posterior communicating arteries. Neuropsychological tests were performed to assess the general cognitive status, as well as logical verbal and visual-constructual memory. RESULTS: 79% of the patients present a general cognitive status within the range of what could be considered to be normal. In logical verbal memory, 92% present performances within the limits of the expected range of values and 83% did the same in visual-constructual memory. Depending on the location of the aneurysm, significant differences were only found in the delayed evocation of logical verbal material. CONCLUSIONS: A year after the intervention, most of the patients present a pattern of normality in the general cognitive status, and in logical verbal and visual-constructual memory. Yet, in spite of the good neurological resolution, alterations to memory are still to be found, although less frequently. The anatomical location of the aneurysm in the brain affects performance in tasks involving delayed logical verbal memory.
Keyword Search

| magnesium | Search | Enter a query (e.g., magnesium) then click Search |

**WORKBENCH**

Oral **magnesium** treatment inhibits PDT in patients with stable CAD.

The highest potassium concentration was 8.0 mmol/l, **magnesium** 2.15 mmol/l.

Since **magnesium deficiency** has a pro-inflammatory effect, the expected consequence would be an increased risk of developing insulin resistance when **magnesium** deficiency is associated with a high-sucrose diet.

This results in continuous cardiac **stress** during the whole pregnancy.

This phosphorylation was markedly inhibited by **magnesium** sulfate (3.0 mM).

**magnesium** sulfate tocolysis remains a North American anomaly.

**magnesium** sulfate prophylaxis with **magnesium** sulfate, with expeditious delivery.

In addition, calcium, creatinine, and **magnesium** were measured in urine.

<table>
<thead>
<tr>
<th>Relation</th>
<th>List of Entities</th>
<th>File name</th>
</tr>
</thead>
<tbody>
<tr>
<td>entry combination of</td>
<td>Stress</td>
<td>17402562-6</td>
</tr>
<tr>
<td>leads to</td>
<td>6674050-S</td>
<td>1997-235-1</td>
</tr>
<tr>
<td>1332048-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1544946-11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16555586-7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>58230791-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19682121-4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18489501-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19146798-5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19173040-9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19290477-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19244720-1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19244720-6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19244725-9</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Traditionally the creation of domain models is a long and expensive process
Motivation

• For this reason we propose a largely automatic domain model generation on the basis of existing knowledge repositories and text corpora, namely MeSH, MedLine and Wikipedia.
Results
Top Levels by importance (#Concepts)
Selective Hierarchy – CogSci/NeuroSci
Zooming in on Cognition-related concepts
Biomolecules-related concepts

- j.2:_sb_224289_in_controlling_extracellular_5_ht
- j.2:_excessive_5_ht_activity_in_brain_areas
- j.2:_increased_neuronal_density_and_thalamic_volume_reduced_thalamic_activation_filtering_sensitization_of_dopaminergic_modulation
- j.2:_drn_5_ht_neurons
- j.2:_the_roles_of_5_ht_1a_and_3_ht_1b_autoreceptors
- j.2:_hallucinations_reduced_cortical_and_thalamic_volume_reduced_thalamic_activation_filtering_sensitization_of_dopaminergic_modulation
- Ht_0712
- j.2:_due_to_a_reduced_corn_degradation_of_norepinephrine_and_dopamine
- j.2:_increased_excitatory_drive_on_dopamine_neurons
- j.2:_the_behaviourally_elicted_terminal_release_of_dopamine
- j.2:_dopamine_depletion
- j.2:_important_for_depression_pathophysiology_eg_dopamine_da
- j.2:_abnormal_rhythms_of_the_central_neurotransmitters
- category_phosphodiesterase_inhibitors
- category_proteins
- category_biomolecules
- category_neurotransmitters
- category_neurotransmitters
- Gamma_aminobutyric_Acid
- j.2:_gamma_aminobutyric_acid_receptor_gababr
- j.2:_purkinje_cells_gi_o_protein_coupled_b_type

Knowledge Enabled Information and Services Science
Technologies
Steps

1. Carve a focused domain model out of Wikipedia
2. Identify mentions of entities and relationships in the relevant scientific literature (Pubmed abstracts) to support non-hierarchical guidance.
3. Map extracted entity mentions to concepts and extracted predicates to relationships
4. Applications to access research literature guided by the domain model.
1. Carve a focused domain model out of Wikipedia
2. Identify mentions of entities and relationships in the relevant scientific literature (Pubmed abstracts) to support non-hierarchical guidance.
3. Map extracted entity mentions to concepts and extracted predicates to relationships
4. Applications to access research literature guided by the domain model.
Step 1: Doozer

An Expand and Reduce approach to Automatic domain model creation

Christopher Thomas, Pankaj Mehra, Roger Brooks, Amit Sheth

Knowledge Enabled Information and Services Science
Wisdom of the Crowds

• The most comprehensive and up to date account of the present state of knowledge is given by

   **Everybody**

= The Web in general
= Blogs
= Wikipedia
Goal:
Harness the Wisdom of the Crowds to Automatically define a domain with up-to-date concepts
• Expand and Reduce approach
  – Start with ‘high recall’ methods
    • Exploration - Full text search
    • Exploitation – Node Similarity Method
    • Category growth
  – End with “high precision” methods
    • Apply restrictions on the concepts found
    • Remove unwanted terms and categories
Expand - conceptually

Graph-based expansion

Delete results with low Lucene score.
Node Similarity computation

\[ s(a, b) = \frac{\sum_{\{i, j \mid \langle N_i(a), N_j(b) \rangle \in M \}} \text{avg}(w(N_i(a)), w(N_j(b)))}{\text{avg} \left( \sum_{i \in |N(a)|} w(N_i(a)), \sum_{j \in |N(b)|} w(N_j(b)) \right)} \]
Collecting Instances

Full Text Search

Semantic Similarity Method

Query: "server* compute* informati* internet www ftp"

FTS

root nameserver

DNS root zone

Server (computing)

Wikipedia

NSD

AlterNIC

EDNS

DNS zone

ICANN

Internet
Creating a Hierarchy

Knowledge Enabled Information and Services Science
Creating a Hierarchy

Seed Query

Fulltext Concept Search (Somnath)

Wikipedia

Graph Search (Denis)

Graph Search (Denis)

Graph Search (Denis)

Knowledge Enabled Information and Services Science
Hierarchy Creation - summary
Reduce steps

• Remove all terms that have low pertinence to the domain
• Intersect hierarchy with broader focus domain
• Reduce hierarchy depth
Remove unwanted individuals
Remove unwanted categories
Flatten categories
Evaluation wrt. expert model

Evaluation wrt. MeSH versions of 2004 (04) and 2008 (08) for both the restricted and the full set of MeSH term
1. Carve a focused domain model out of Wikipedia

2. Identify mentions of entities and relationships in the relevant scientific literature (Pubmed abstracts) to support non-hierarchical guidance.

3. Map extracted entity mentions to concepts and extracted predicates to relationships

4. Applications to access research literature guided by the domain model.
• Vocabularies, Thesauri, Ontologies, exist in several fields
  – How can we use them?
    • lexicon: Fish Oils, Raynaud’s Disease, etc.
    • types/labels: **Fish Oils instance of Lipid**
    • relationships between types: **Lipid affects Disease**

• Identification of simple ontology terms in text is not enough
  – Compound Entities
  – Complex Relationships

But
• Entities (MeSH terms) in sentences occur in modified forms
  - “adenomatous” modifies “hyperplasia”
  - “An excessive endogenous or exogenous stimulation” modifies “estrogen”
• Entities can also occur as composites of 2 or more other entities
  - “adenomatous hyperplasia” and “endometrium” occur as “adenomatous hyperplasia of the endometrium”
Extraction Algorithm

Anti-Ro(SSA) autoantibodies are associated with T cell receptor beta genes in systemic lupus erythematosus patients.

(A)

[Anti-Ro(SSA) autoantibodies UMLS:Immunologic_Factor] are associated with [T cell receptor beta genes UMLS:Gene_or_Genome] in [systemic lupus erythematosus patients UMLS:Patient_or_Disabled_Group].
Extraction Approach

- Parse sentences with a dependency parser
- Use a few domain-independent rules to segment sentences into Subj → Pred → Object
- Subjects and Objects represent compound entities
Preliminary results

Subject types, inferred from the HEAD of the compound

- [Genetic_Function] The cardiac myosin heavy chain Arg-403→Gln mutation
- [Disease_or_Syndrome] hypertrophic cardiomyopathy
- [Amino_Acid_Peptide_or_Protein] The cardiac ventricular myosin
- [Organism_Attribute] heavy chain phenotype
- [Amino_Acid_Peptide_or_Protein] the fibronectin receptor, talin, vinculin and actin
- [Gene_or_Genome] the RAD51 and RAD52 genes
- [Tissue] the plasma membrane
- [Body_Part_Organ_or_Organ_Component] the ligated kidneys
- [Body_Part_Organ_or_Organ_Component] the developing kidney
- [Amino_Acid_Peptide_or_Protein] other cytoskeletal proteins such as tau or actin
- [Amino_Acid_Peptide_or_Protein] a ubiquitous actin-binding protein
- [Amino_Acid_Peptide_or_Protein] the ubiquitous actin
- [assay] The centrifugation assay and the DNase I inhibition assay
- [Social_Behavior] a role for synaptojanin family members in actin function
<table>
<thead>
<tr>
<th>Relationship</th>
<th>Sentence Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>increased</td>
<td>A pre-treatment of cells with SGE from partially fed ticks in amounts salivary glands → increased → the level of both viral nucleocapsid N protein phosphoprotein P in a dose-dependent manner</td>
</tr>
<tr>
<td>inhibits</td>
<td>alpha-catenin → inhibits → beta-catenin signaling</td>
</tr>
<tr>
<td>inhibits</td>
<td>MgCl2 → inhibits → these effects of profilin, most likely</td>
</tr>
<tr>
<td>causes</td>
<td>The cardiac myosin heavy chain Arg-403 Gln mutation → causes → hypertrophic cardiomyopathy</td>
</tr>
<tr>
<td>causes</td>
<td>Moreover, addition of profilin to steady-state actin filaments → causes → slow depolymerization</td>
</tr>
<tr>
<td>causes</td>
<td>(11-22 microM) into infected PtK2 cells → causes → a marked slowing of actin tail elongation and bacterial migration</td>
</tr>
<tr>
<td>binds</td>
<td>the cytoplasmic domain of E-cadherin → binds → either beta-catenin or plakoglobin</td>
</tr>
<tr>
<td>binds</td>
<td>a constituent → binds → RBC alpha-spectrin antibody plus the presence of significant quantities of actin</td>
</tr>
</tbody>
</table>
An excessive endogenous or exogenous stimulation modifies estrogen, inducing adenomatous hyperplasia of the endometrium.

- **Modifiers**
  - hasModifier

- **Modified entities**
  - hasPart

- **Composite Entities**
  - induces
  - hasPart
Using the generated RDF

- **migraine** (D008881)
  - caused by
  - hasPart
  - me_3142
    - by_a_primary_abnormality_of_platelet_behavior

- **platelet** (D001792)
  - hasPart
  - me_2286
    - _13%_and_17%_adp_and_collagen_induced_platelet_aggregation

- **collagen** (D003094)
  - hasPart
  - stimulated

- **magnesium** (D008274)
  - hasPart
  - stimulated
Steps

1. Carve a focused domain model out of Wikipedia
2. Identify mentions of entities and relationships in the research literature.
3. Map extracted entity mentions to concepts and extracted predicates to relationships
4. Applications to access research literature guided by the domain model.
Step 3: Identifying Synonymy and Antonymy between Verbs for Relationship Matching

Christopher Thomas, Wenbo Wang
Motivation

• An ontology has clearly demarcated, formal relationships whereas natural language uses multiple ways of representing these relationships

• Focus on verbs expressing relationships

• NLP-extracted triples use verbs to indicate relationships, but are not aligned with the relationships yet.

• Align similar verbs → Align extracted predicates to formal relationships
Examples of relationships to merge

- Dilation of the fourth ventricle **indicated** atrophy of the middle cerebellar peduncle.
  - E.g. <Dilation of the fourth ventricle> → **GOOD**
    - indicate
      <atrophy of the middle cerebellar peduncle>

- The lesions in the rolipram-treated group also **showed** increased astrogliosis and increased CREB phosphorylation in the activated microglia and astrocytes.
  - E.g. <lesions in the rolipram-treated group> → **GOOD**
    - indicate
      <increased astrogliosis and increased CREB phosphorylation in the activated microglia and astrocytes>
Short Review – training set creation

- Normalization: Remove phrases containing less than 2 verbs
- POS Tagging: Eating -> <VBG>, eats -> <VBZ>
- Synonyms & antonyms extraction: ANT: Like <-> Hate, SYN: calculate <-> compute
- Pattern extraction: He learns and memorizes materials
- Pattern2Relationship matrix
- VerbPair2Pattern matrix: What is the correlation of this verb pair and that pattern?
• Synonyms & antonyms extraction from Wordnet
  – A single verb has multiple meanings
    • contract, take, get (be stricken by an illness, fall victim to an illness) "He got AIDS"; "She came down with pneumonia"; "She took a chill"
    • film#1, shoot#4, take#16 (make a film or photograph of something) "take a scene"; "shoot a movie"
    • Take has 42 meanings in Wordnet
  – Some meanings of a verb are less frequently used
  – If the number of meanings of a verb exceeds threshold, we will abandon this verb
  – If the frequency of a specific meaning of a verb is low, we will abandon this meaning
Solution

- Do pattern-based text analysis to find predicates that are used in similar contexts and map them to upper-level relationships, e.g. those found in UMLS
Steps

1. Carve a focused domain model out of Wikipedia
2. Identify mentions of entities and relationships in the research literature.
3. Map extracted entity mentions to concepts and extracted predicates to relationships
4. Applications to access research literature guided by the domain model.
Trellis: Knowledge-Driven Web Browsing

Step 4

Pablo N. Mendes, Cartic Ramakrishnan, Delroy Cameron and Amit P. Sheth
Knoesis Center, May 27 2009.
Introduction

- Background
- System Overview
- Demo
- Future Work
Background

- **Web Browsing**
  - **Search**
    - Keyword search
      - Point of access to related content
  - **Sift**
    - Retrieve documents
      - Hyperlinks
  - **Query Refinement**
Background

Observations

Human Cognition

Model

Information Need (Documents)

Keyword

Entity

Relationship

Pathway

Graduate Student
Research Assistant
Turkey
San Diego
System Overview

- **New Browsing Paradigm**
  - Search
  - Navigate Context
  - Organize results

- **Trellis**
  - Intricate support structure for vines
  - Vines as information need/context through navigation
System Overview

Fig. 1 Trellis Architecture
Actions discussed include hemodynamic monitoring, the administration of oxygen, and the use of morphine, diuretics, nitroglycerin, angiotensin-converting enzyme inhibitors, angiotensin II receptor blockers, spironolactone, beta-blockers, calcium channel blockers, magnesium, digoxin, and positive inotropic drugs.

Moreover, potentiation of the K⁺-effect by magnesium ions has been observed.

The effect of potassium ions on the formation of adenosine 3',5'-cyclic monophosphate (cAMP) in the rat cerebral cortex was studied under conditions where development of spreading depression had been blocked by pretreatment of the cerebral cortex with magnesium ions.

In this review, we describe and discuss the myriad of integrated physiological mechanisms through which magnesium may beneficially alter outcomes.

These findings suggest a potential mechanism whereby magnesium may beneficially alter outcomes through decreased effects on spreading depression.

Indeed, serum ferritin, magnesium, and ascorbate concentrations were higher in the ADHD group, but iron, zinc, and vitamin B6 were not different.

Recently, there has been burgeoning experimental, clinical, and epidemiological data that provides strong evidence that dietary magnesium intake and supplementation are inversely associated with the risk for MetS and its components.
Calcium Channel Blockers -> *prevent* -> Migraine
Magnesium -> isa -> Calcium Channel Blocker

Migraine -> prevent -> Migraine

Magnesium -> prevent -> Migraine
Future Work

- **Ranking - Boosting Precision/Recall**
  - User feedback
  - Corpus Statistics

- **Scalability**
  - Information Extraction from natural language

- **Named Entity Recognition**
  - User driven

- Knowledge Discovery
Thank you