Finding Seeds of Future Discoveries in Current Literature

Tanja Urbančič
University of Nova Gorica, Nova Gorica, Slovenia and
Jozef Stefan Institute, Ljubljana, Slovenia

Ingrid Petrič
University of Nova Gorica, Nova Gorica, Slovenia

Nada Lavrač, Bojan Cestnik, Matjaž Juršič, Borut Sluban
Jozef Stefan Institute, Ljubljana, Slovenia

Marta Macedoni-Lukšič
University Children's Hospital, Ljubljana, Slovenia (previous),
Institute for Autism (now)

Outline

- **Part 1: Introduction**
  - Creativity and knowledge discovery
  - Literature as a source of cross-context knowledge discoveries
  - Open vs. closed discovery
- **Part 2: Open discovery framework**
  - Exploring rarity (RaJoLink)
  - Case 1: RaJoLink in the Migraine domain
  - Case 2: Towards better understanding of Autism
- **Part 3: Closed discovery framework**
  - Speeding-up the process:
    exploring outliers, ranking candidate bridging terms, workflows, HCI (CrossBee)
- **Part 4: What next?**

“What could be learned from looking at nothing?”

(F. Slakey)

- **Hubble Space Telescope**
  - Carried into orbit by a space shuttle in 1990
  - Positioned above the Earth’s atmosphere
  - Extremely high-resolution images
  - Many Hubble observations led to breakthroughs in astrophysics

- **Hubble Deep Field**
  - Image assembled from 342 separate exposures taken over 10 consecutive days (18-28 Dec 1995)
  - Focused on a very small region in the constalation of Ursa Major (“an empty black void”)
  - The picture revealed over 1.500 galaxies and became a landmark image in the study of the early universe

Creativity and Discoveries

- **Boden (The Creative Mind – Myths and Mechanisms, 2004):**

  Creativity as "the ability to come up with ideas or artifacts that are new, surprising and valuable"

  Three types of creativity: combinatorial, exploratory, transformational

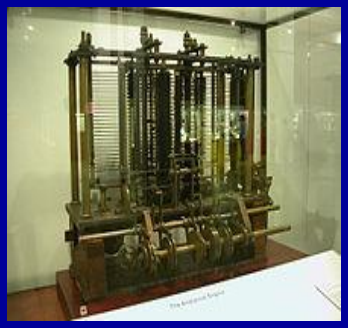
- **In many discoveries, existing knowledge is creatively combined with new ideas.**

- **Koestler (The Act od Creation, 1964):**

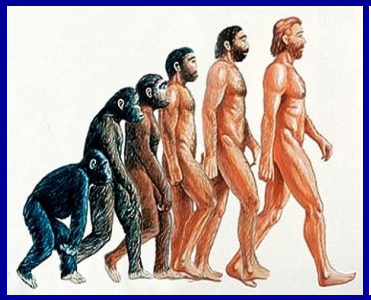
  "Creative act uncovers, selects, re-shuffles, combines, synthesizes already existing facts, ideas, faculties, skills. The more familiar the parts, the more striking the new whole."

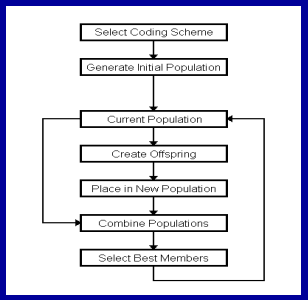
- **Ideas often come from other contexts**

Examples from the history of computer science



- From Jacquard's loom (1801) to Babbage's Analitic Engine (1837)
  - Idea of punched cards representing data, problems etc. (instead of colorful textil patterns)



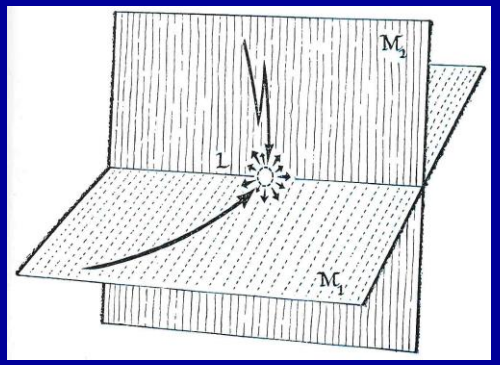


- From evolution in nature to evolutionary computing (Lawrence J. Fogel, 1964)
  - Idea of populations of candidate solutions developing through simulated evolution (instead of "survival of the fittest" in nature)

Cross-context link discovery

- **Bisociations as a basis of human creativity**

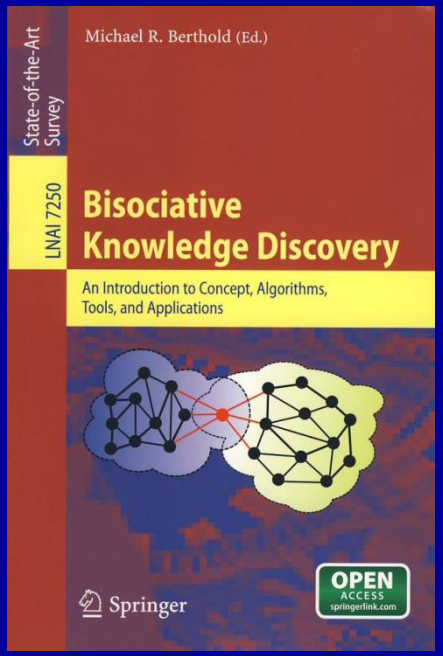
"… the perceiving of a situation or idea, L, in two self-consistent but habitually incompatible frames of reference, M1 and M2. The event L ... is not merely linked to one associative context but bisociated with two." (Koestler, 1964)



- **Finding links which lead "out-of-the-plane"**

- BISON: Bisociation Networks for Creative Information Discovery, 7FP project (2008-2010)

- **Two concepts are bisociated if and only if:**
  - There is no direct, obvious evidence linking them
  - One has to cross contexts to find the link
  - This new link provides some novel insight



Cross-context link discovery from literature

Swanson, Medical literature as a potential source of new knowledge, 1990

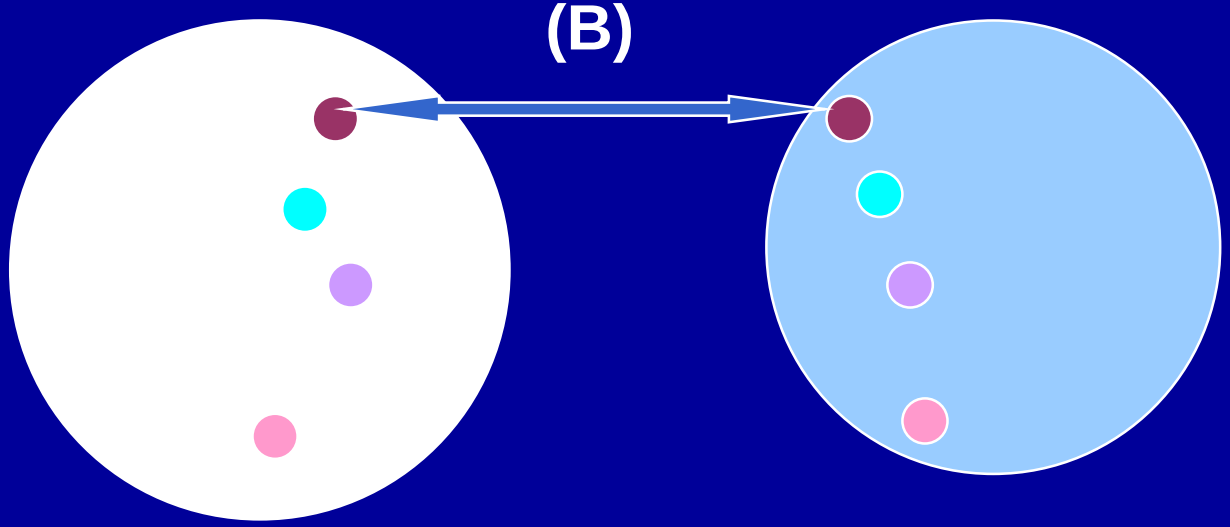
Literature about magnesium (A)

(38.000 articles)

(B)

Literature about migraine (C)

(4.600 articles)

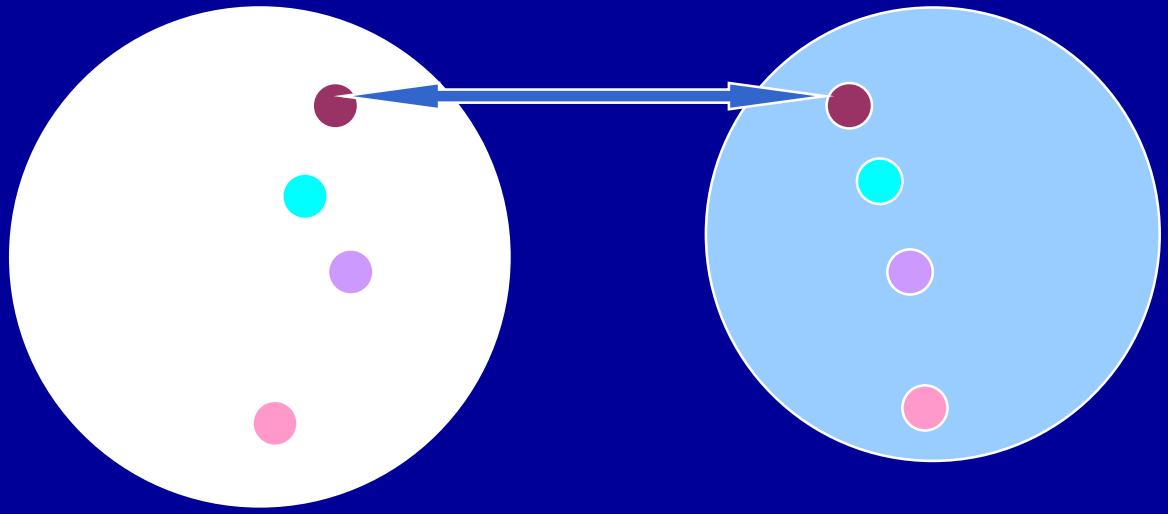


Swanson's ABC model
B: Linking (bridging) terms

Literature about magnesium (A)

Literature about migraine (C)

B-terms: calcium channel blockers, stress and A type behavior, ...



Argument 1
(magnesium literature)

- Mg is a natural calcium channel blocker.
- Stress and Type A behavior can lead to body loss of Mg.
- Magnesium has anti-inflammatory properties.
- . . .

Argument 2
(migraine literature)

- Calcium channel blockers can prevent migraine attacks.
- Stress and Type A behavior are associated with migraine.
- Migraine may involve sterile inflammation of the cerebral blood vessels.
- . . .

Generation of a hypothesis
A may influence C

For a given C, how do we find A?

Swanson:

"Search proceeds via some intermediate literature (B) toward an unknown destination A. … Success depends entirely on the knowledge and ingenuity of the searcher."

Can we provide a more systematic support?

Scientific literature as a potential source of new knowledge: What do we have at disposal?

Example:

- **Biomedical bibliographical database PubMed**
- **US National Library of Medicine**
- **More than 21M citations**
- **More than 5.600 journals**
- **2.000 – 4.000 references added each working day!**

Do we use such sources of knowledge effectively?

- **Situation:**
  - Speed of knowledge growth, huge ammounts of literature available on-line
  - High specialization of researchers
  - Potentially useful connections between “islands” of knowledge may remain hidden

- **Our objective:**
  - To develop methods and SW tools to support researchers in their knowledge discovery

Closed vs. open discovery

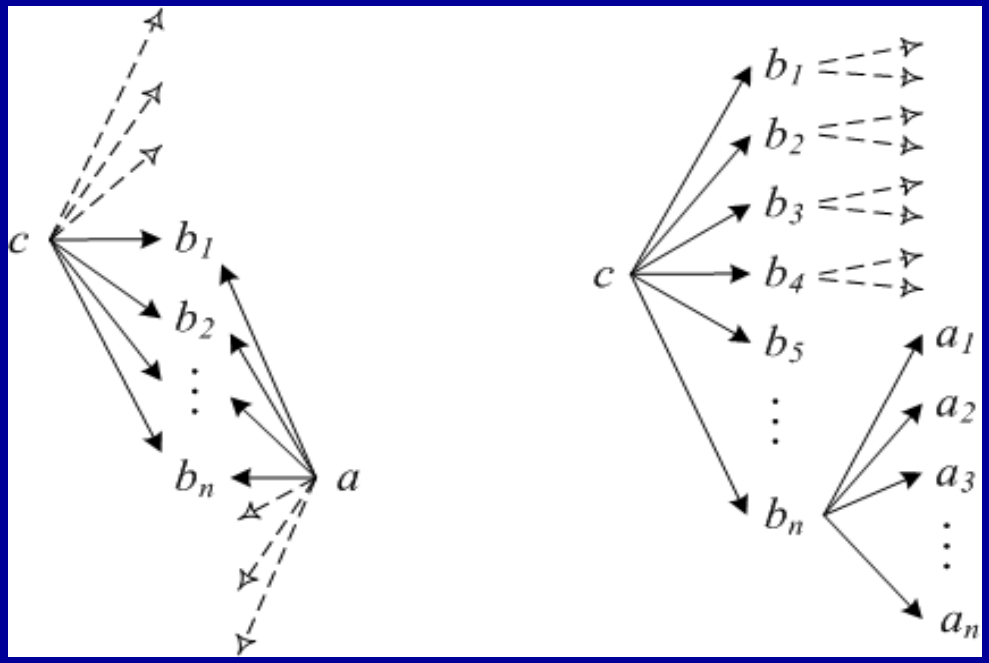
(as defined by Weeber et al., 2001)

Closed discovery

C and A known in advance

Open discovery

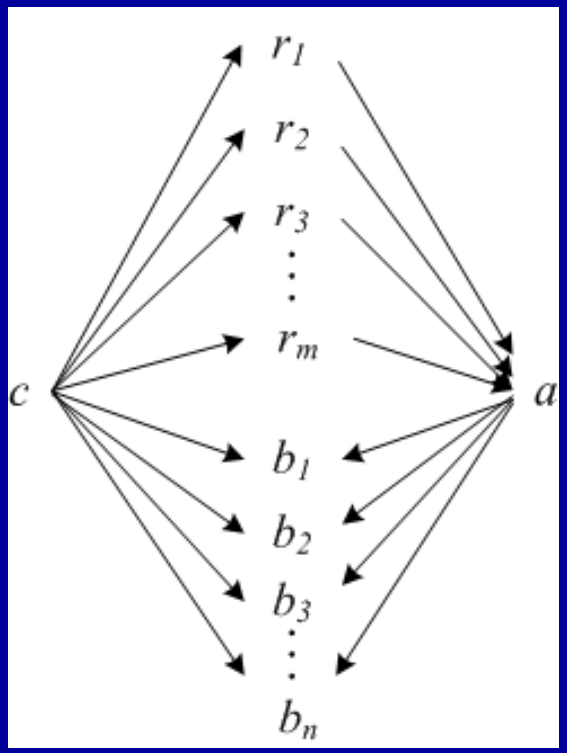
Only C known in advance



Outline

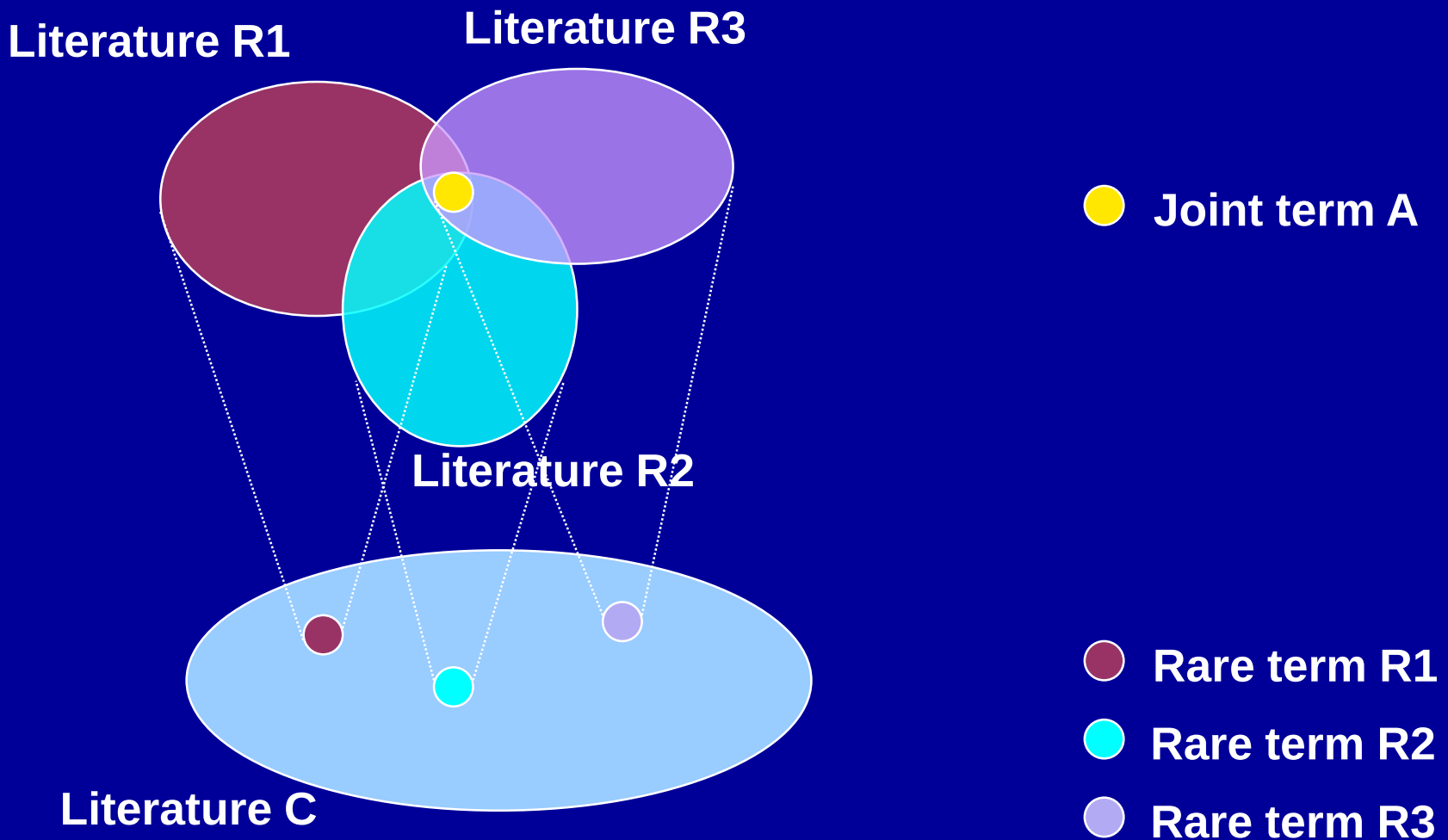
- **Part 1: Introduction**
  - Creativity and knowledge discovery
  - Literature as a source of cross-context knowledge discoveries
  - Open vs. closed discovery
- **Part 2: Open discovery framework**
  - Exploring rarity (RaJoLink)
  - Case 1: RaJoLink in the Migraine domain
  - Story 2: Towards better understanding of Autism
- **Part 3: Closed discovery framework**
  - Speeding-up the process:
    exploring outliers, ranking candidate bridging terms, workflows, HCI (CrossBee)
- **Part 4: What next?**

Combined open and closed discovery in RaJoLink (Petrič, Urbančič, Cestnik, Macedoni-Lukšič, 2009)

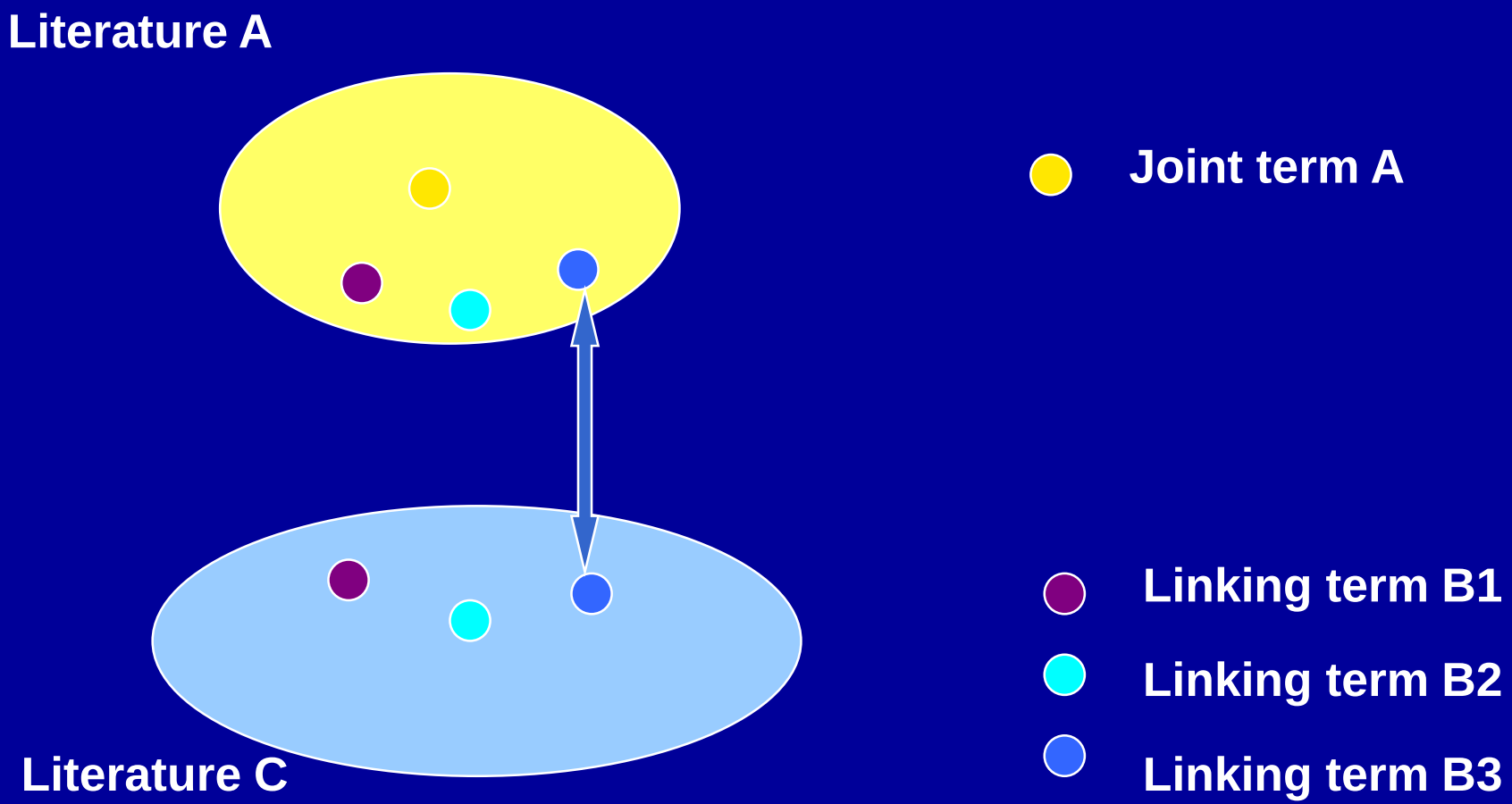


- **Open discovery**
  (upper part of the figure):
  - **Identifying Rare terms *r***
  - **Finding a Joint term *a***
- **Closed discovery**
  (lower part of the figure):
  - **Searching for Linking terms *b***

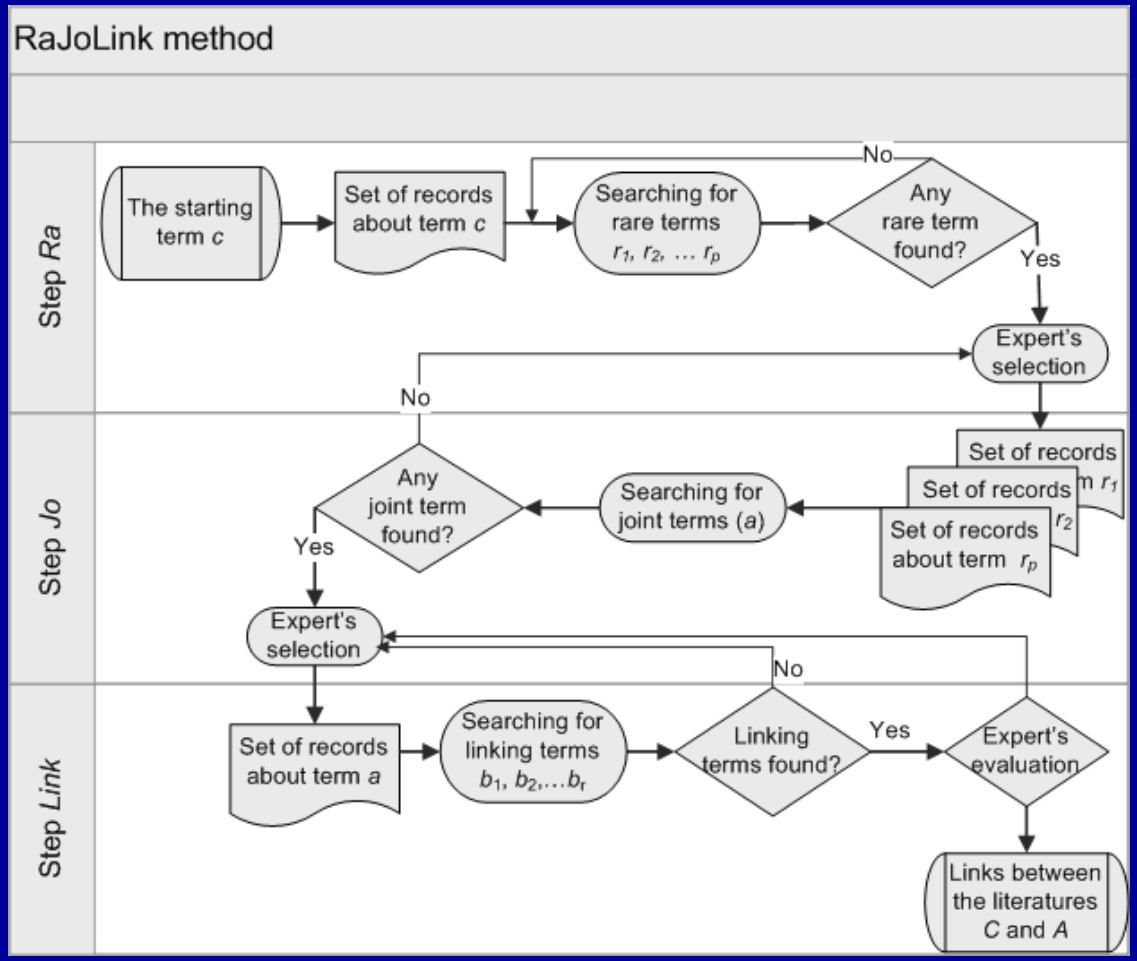
RaJoLink method: Idea
(open discovery part)



RaJoLink method: Idea
(closed discovery part)



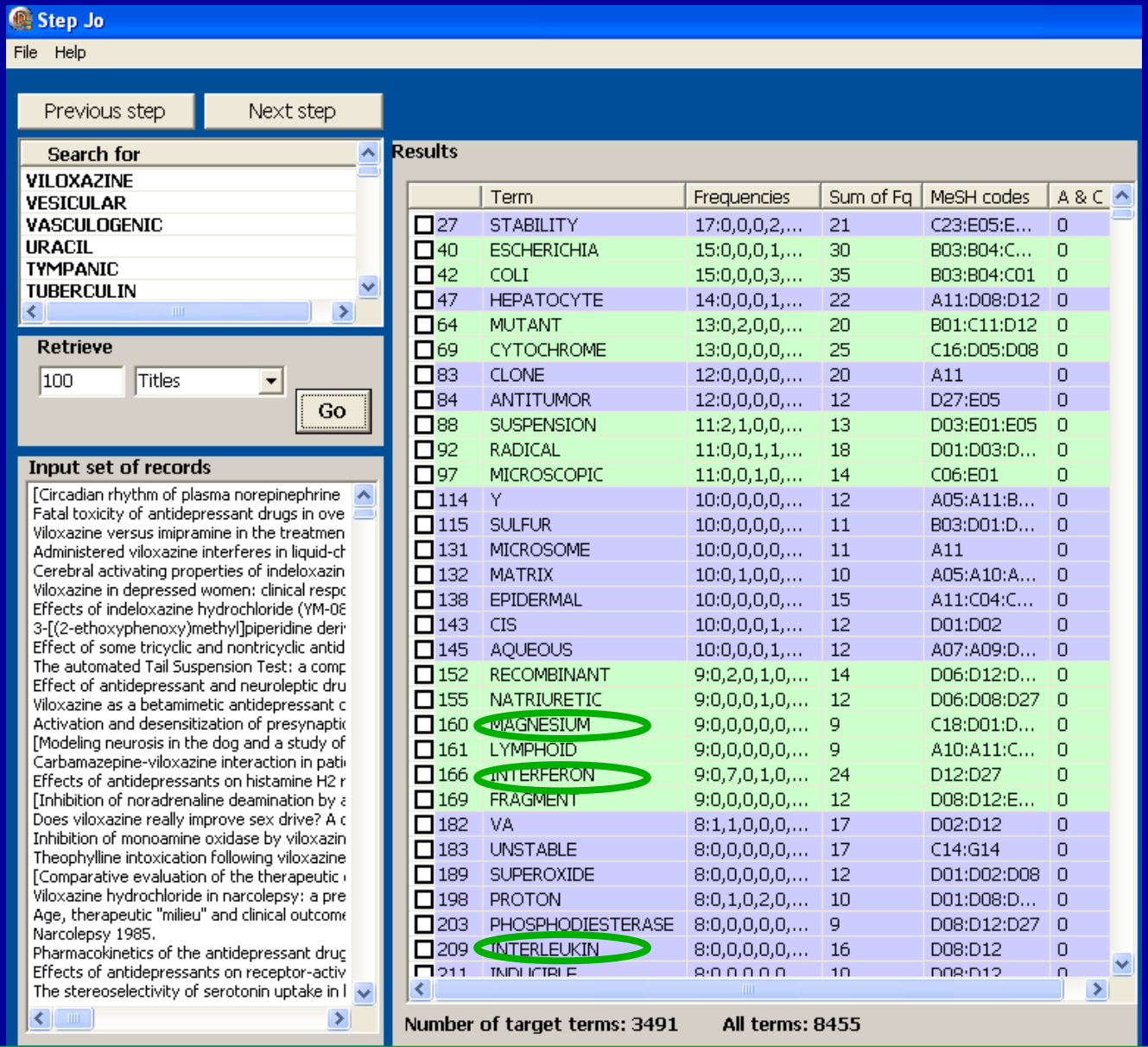
Procedures of the RaJoLink



Case 1: “Re-investigating” migraine

- Problem: migraine headaches
- The golden standard evaluation
- Terms extracted from medical articles published before 1988
- Generation of hypotheses related to migraine domain with the support of RaJoLink tool

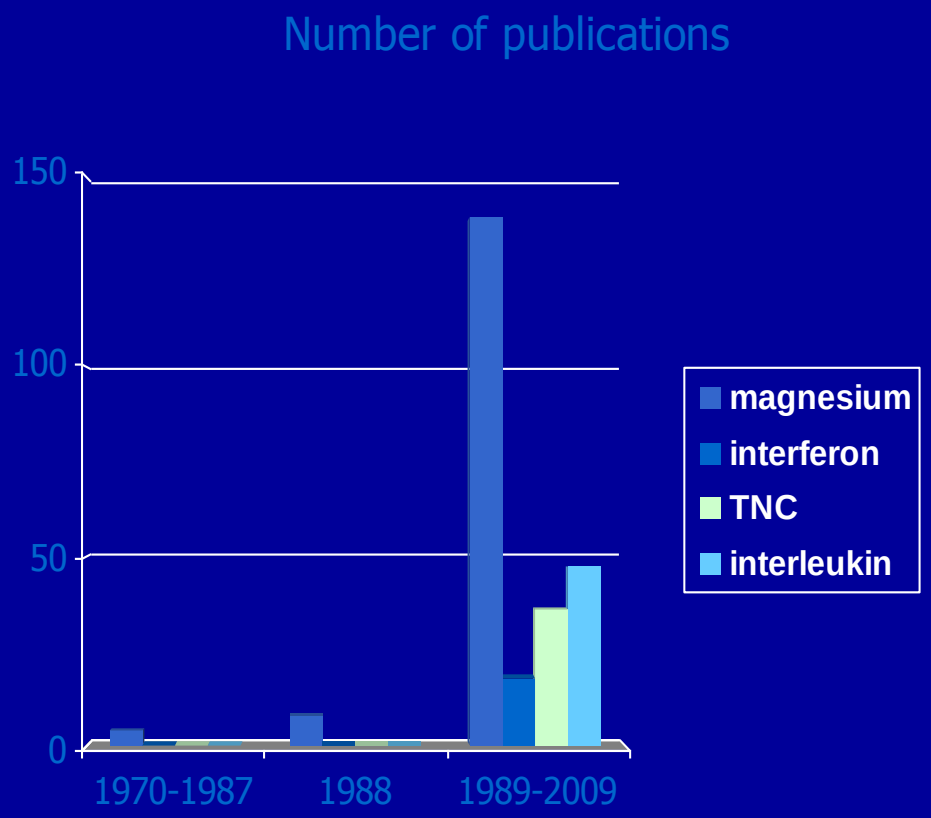
Results in migraine experiment



Results in migraine experiment

- Magnesium
  "re-discovered"
- Other 3 important connections identified with RaJoLink:
  - interferon
  - interleukin
  - tumor necrosis factor

  Their relation with migraine appeared in PubMed after 1988!

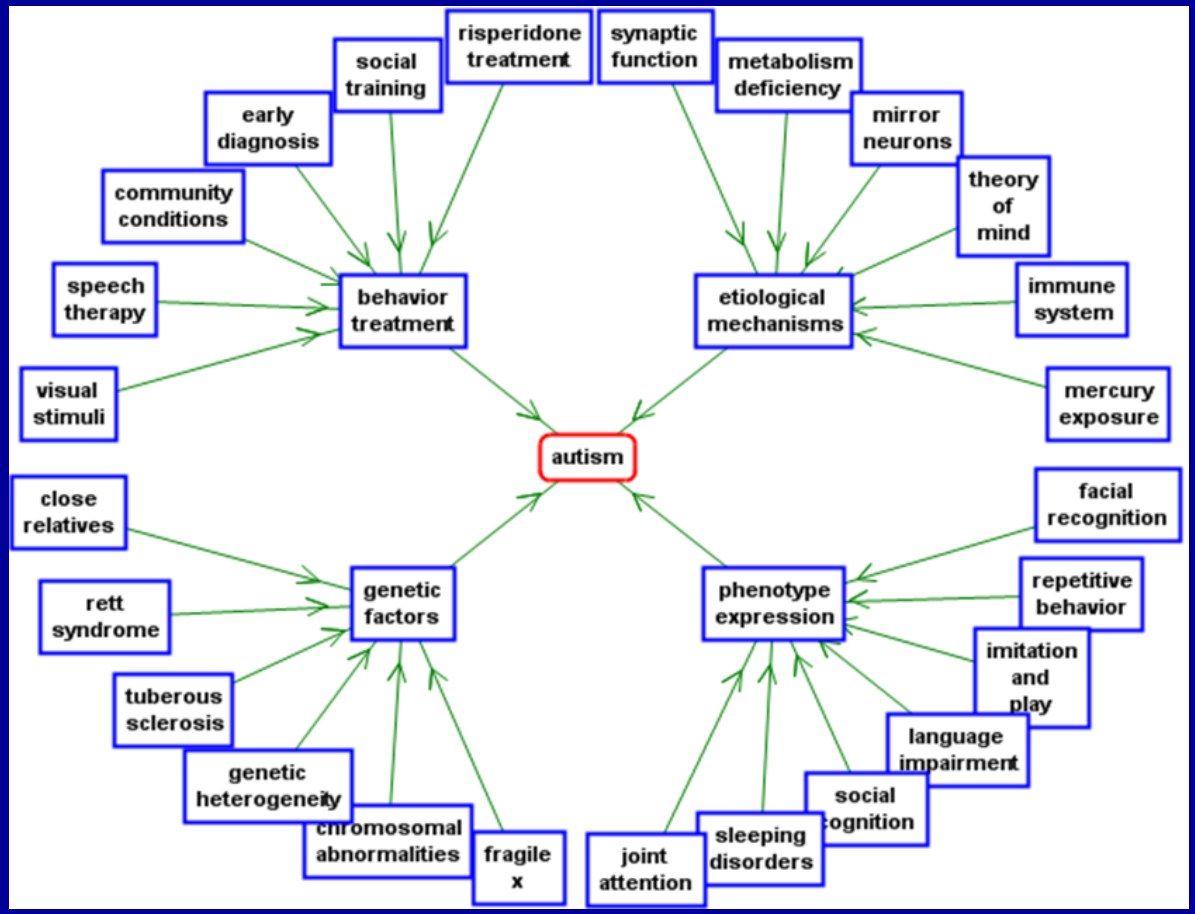


Case 2: Autistic spectrum disorders

- Abnormal development of cognitive, communication and social interaction skills
- Estimated prevalence 5.8 / 1000, increasing
- Causal explanation in 10-15% of cases
- Complex connections of genetic and environmental factors
- Research fragmented (behavioral psychology, genetics, biochemistry, brain anatomy and physiology, …)
- Early diagnosis and treatment very important

Ontology of autism domain

- Tool: OntoGen (Fortuna, Grobelnik, Mladenić, 2006)
- Designed for construction of topic ontologies
- Clustering algorithms used for topic suggestion

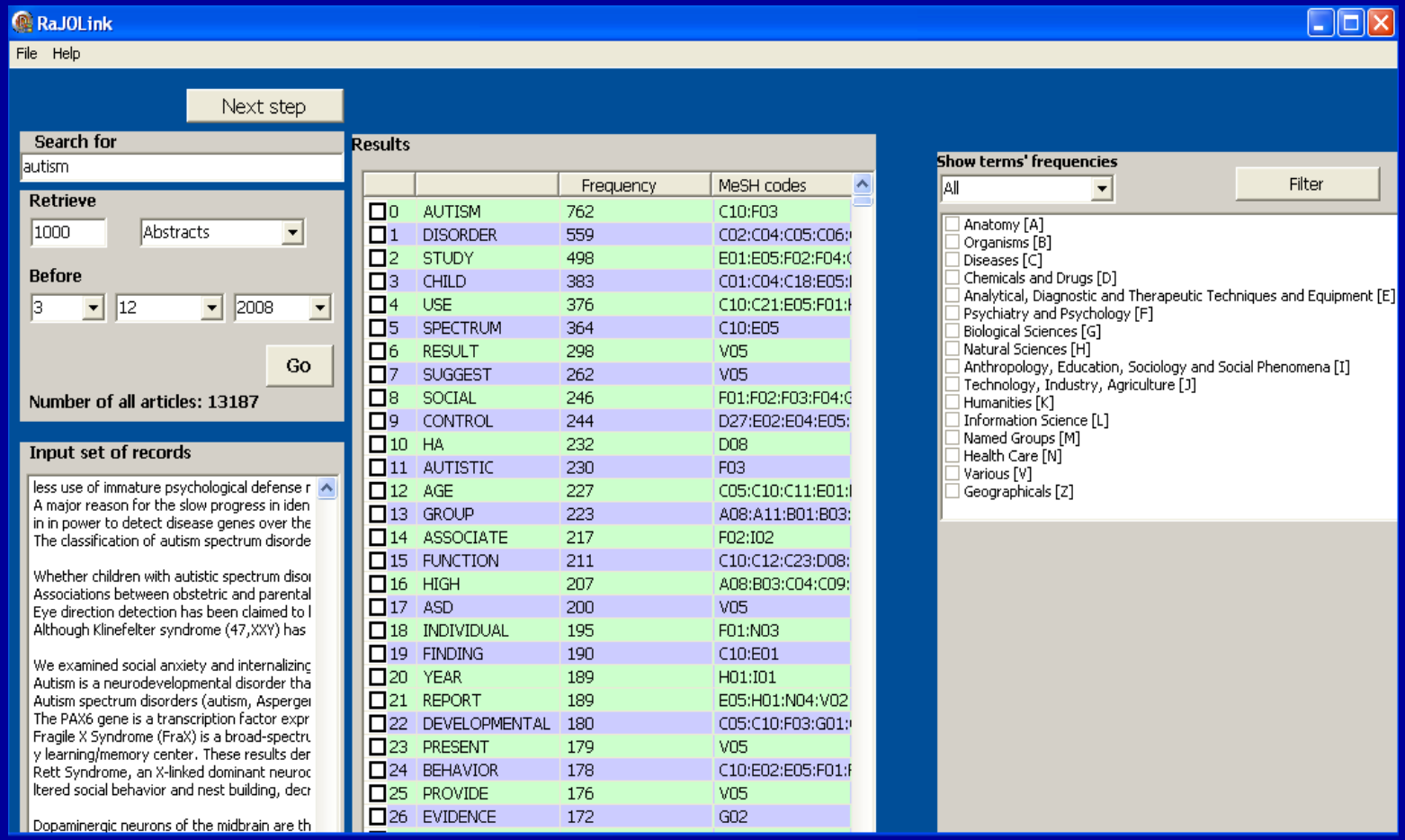


RaJoLink in autism domain

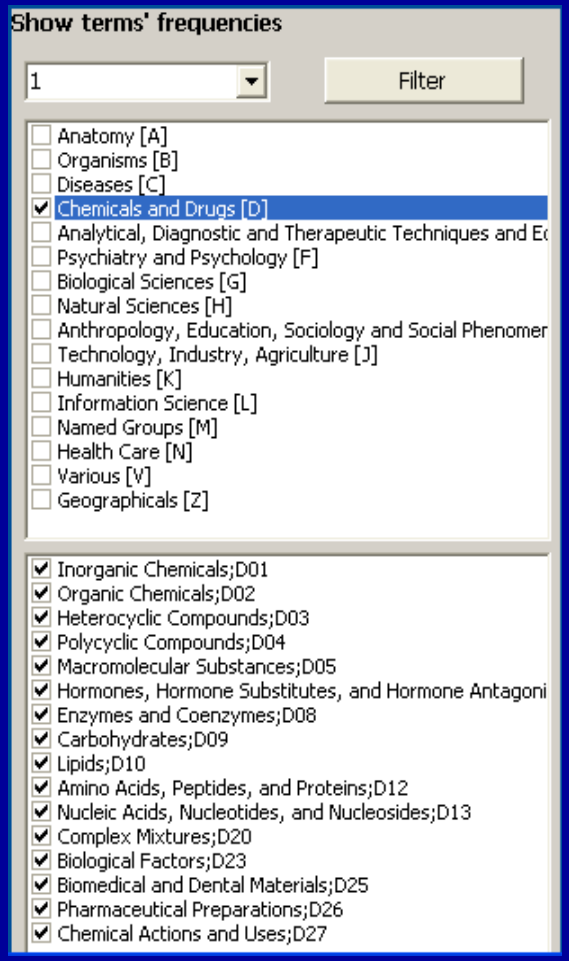
- **Data source: PubMed** ***http://www.ncbi.nlm.nih.gov/sites/entrez***
  - **More than 10.000 documents with autis***
  - **354 articles with entire text in PubMed**
  - **214 of them published in the last 10 years**

- **RaJoLink experiments**
  - **Tool: RaJoLink implements RaJoLink method (Petrič et al. 2007, Urbančič et al. 2007, Petrič et al. 2009)**
  - **Cooperation with medical expert**

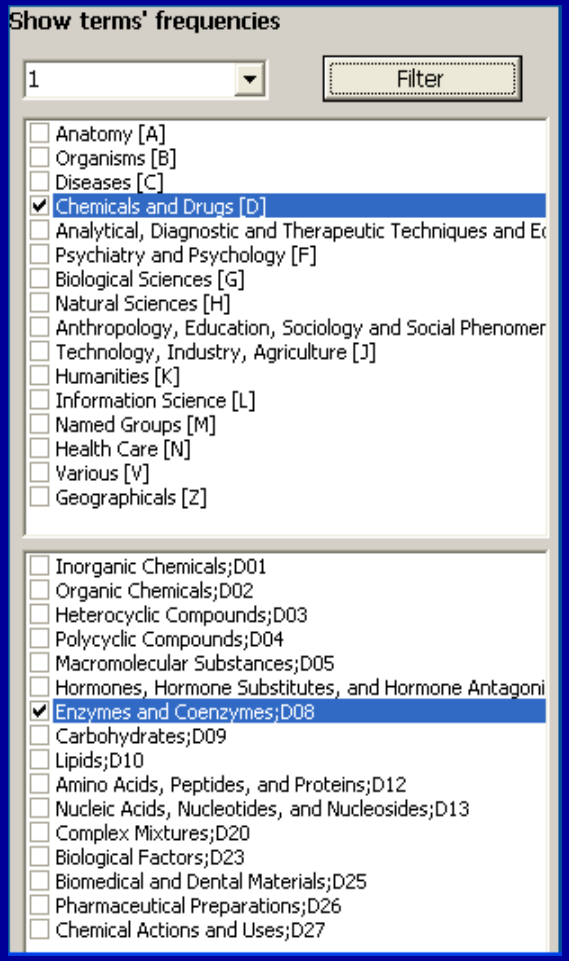
RaJoLink: Step Ra



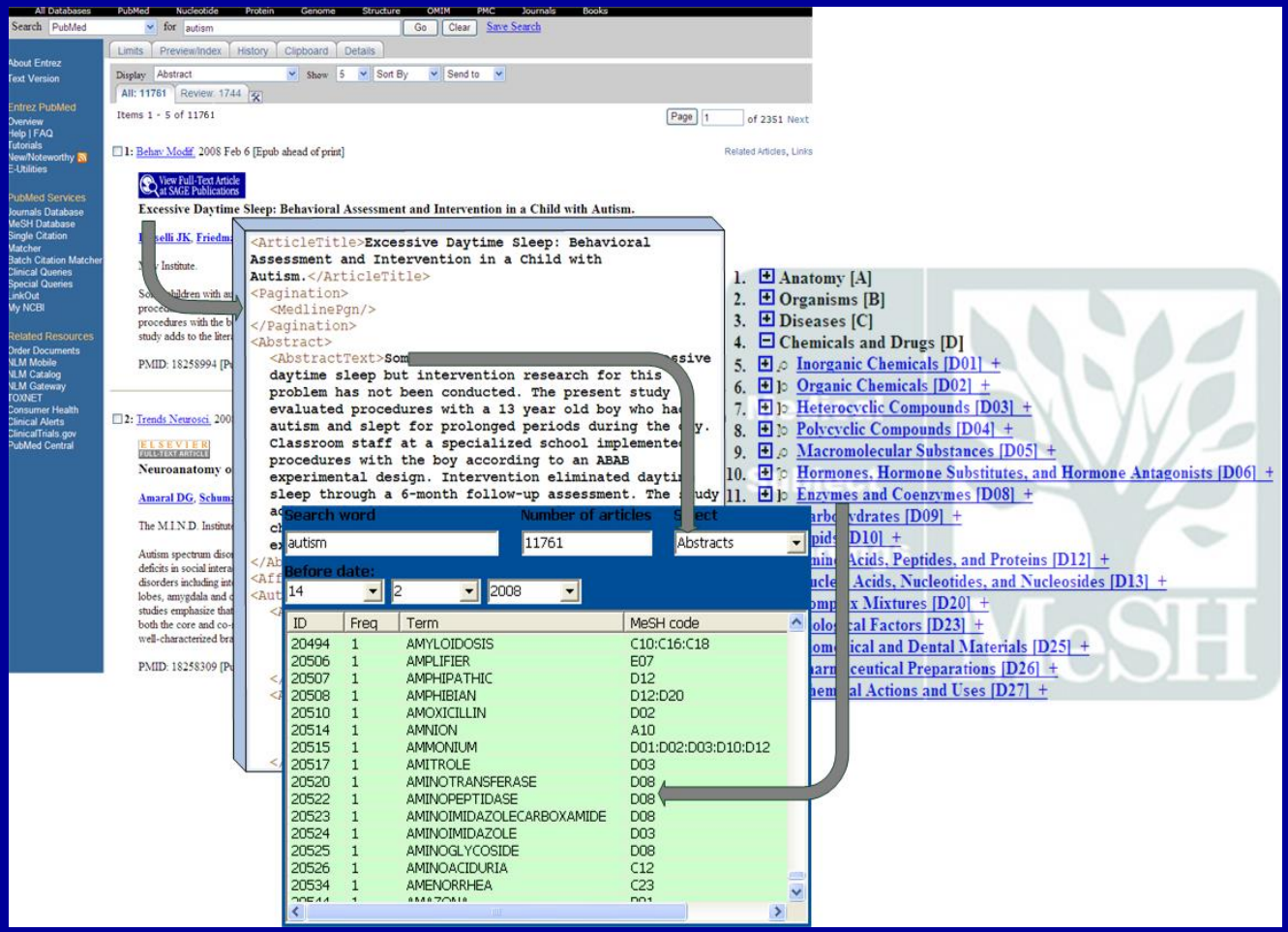
MeSH: Medical Subject Headings



Filtering according to MeSH classification



Step Ra - results



From domain-specific rare terms the following were selected:

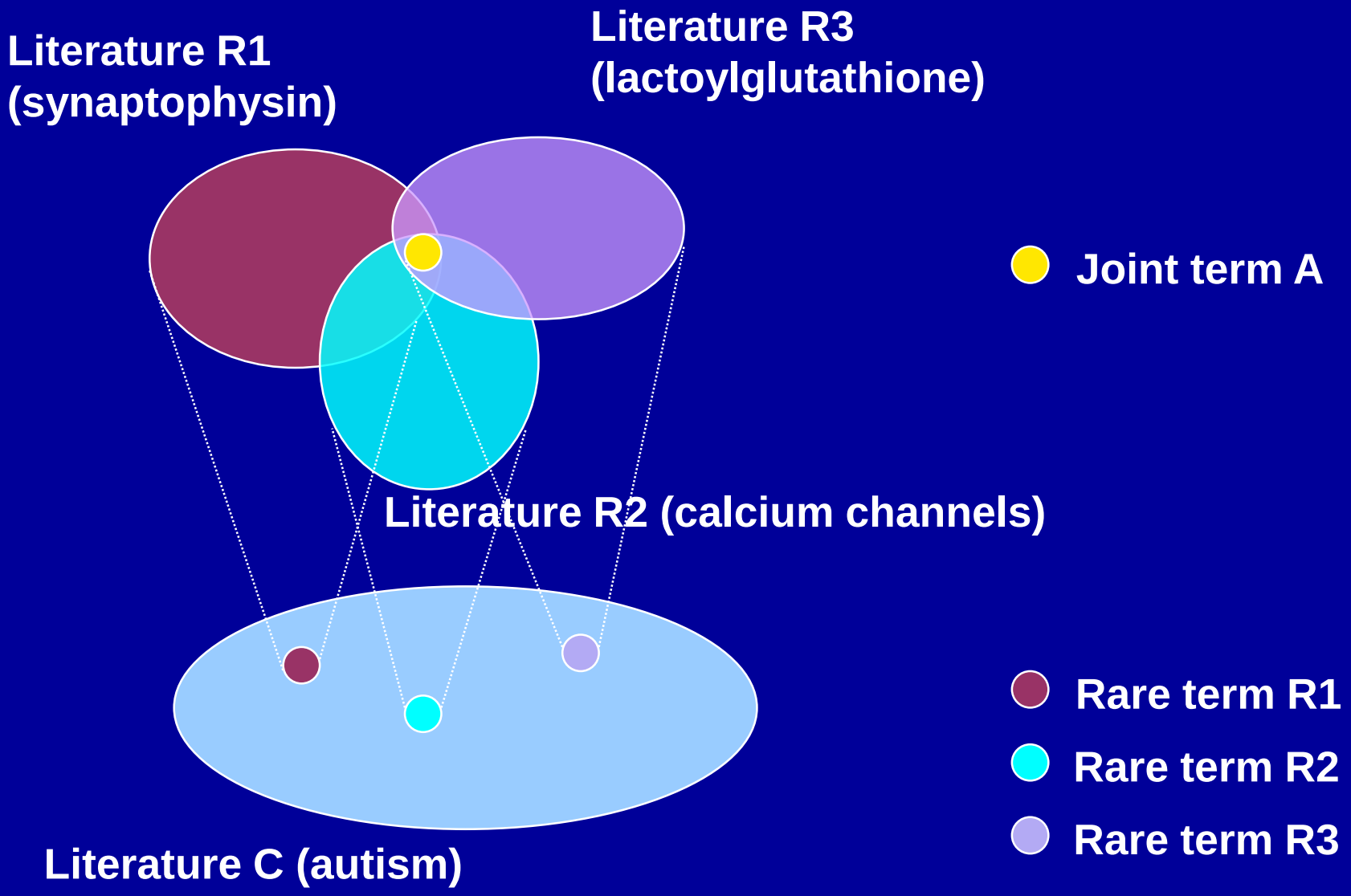
- lactoylglutathione (768 articles in PubMed, 1 of them mentioning autism)
- synaptophysin (3742 articles in PubMed, 2 of them mentioning autism)
- calcium channels (55284 articles in PubMed, 7 of them mentioning autism)

Background knowledge used in selection:

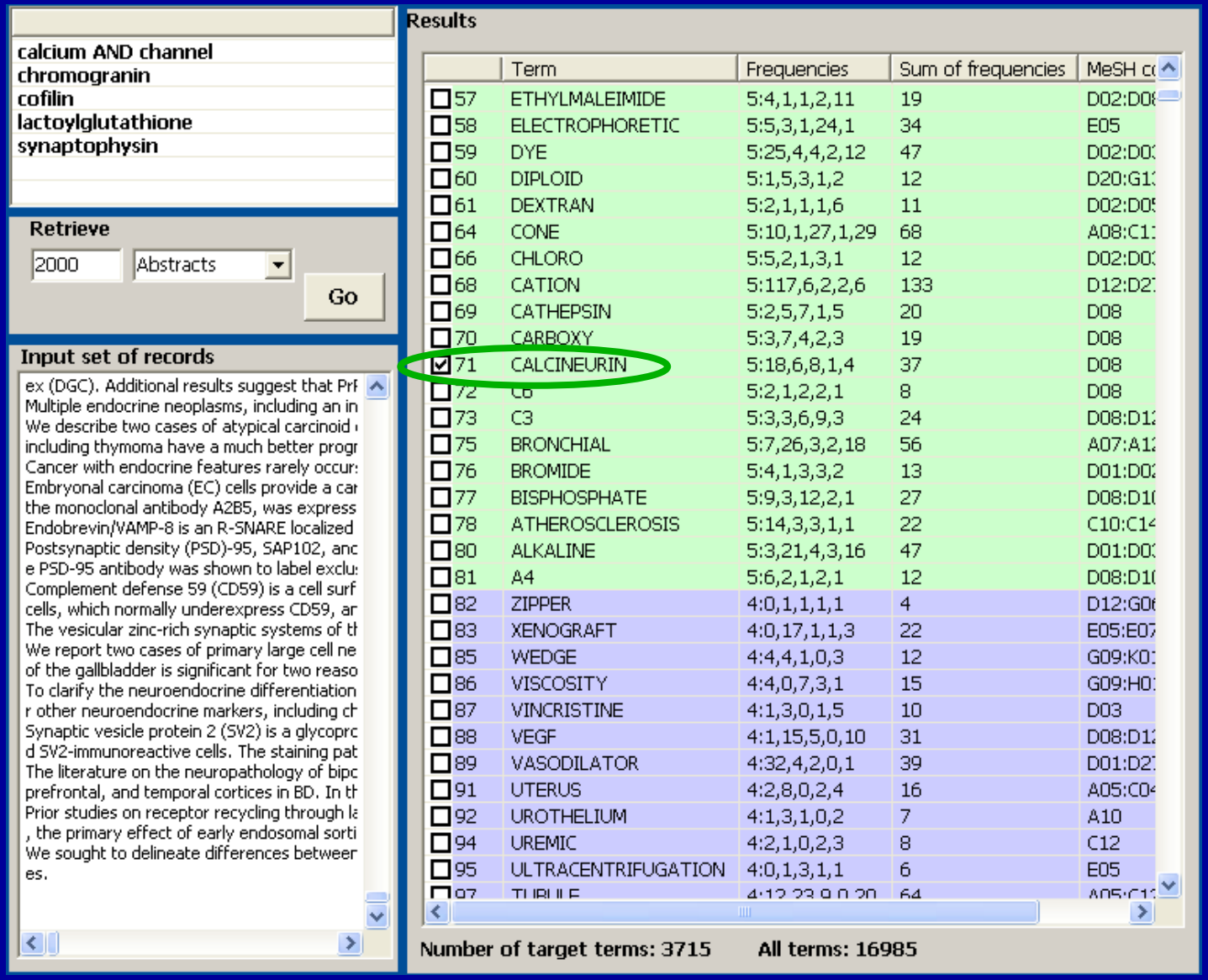
- increase of polarity of glyoxalase I in autistic brain, glyoxalase system involves lactoylglutathione
- altered synaptic function in autism, synaptophysin is a protein localized to synaptic vesicles
- abnormal calcium signalling in some autistic children

Do selected rare terms have something in common?

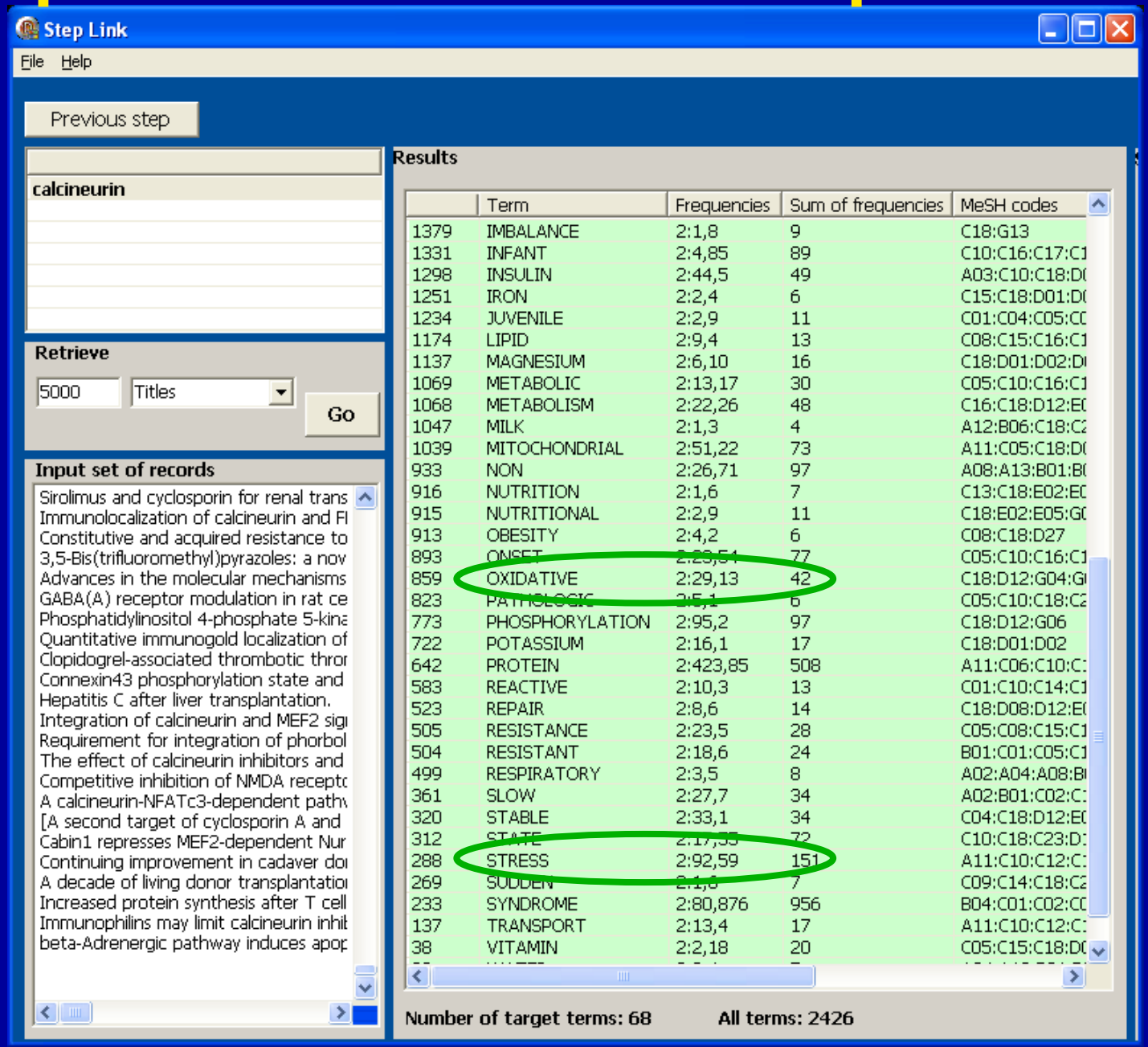
RaJoLink method: Idea



Step Jo: results in autism experiment



Step Link: results in autism experiment



Step Link – hypothesis testing

Expert Opin Pharmacother. 2009 Sep;10(13):2127-43.

Autism: an emerging 'neuroimmune disorder' in search of therapy.

Theoharides TC, Kempuraj D, Redwood L.

RESULTS/CONCLUSION: Increased oxidative stress and immune dysregulation are present in ASDs. Mast-cell activation may contribute to gut-blood-brain barrier disruption and brain inflammation. No effective treatments have emerged. Well-designed clinical trials with nonpsychotropic drugs were few and ASD characteristics varied considerably, making conclusions difficult. Psychotropic drugs are often used for stereotypic and aggressive behaviors. Unique combinations with antioxidant and anti-inflammatory flavonoids hold promise. New potential translational research areas and possible treatments are suggested.

Neurobiol Dis. 2007 May;26(2):342-52. Epub 2007 Jan 25.

Beta-amyloid causes downregulation of calcineurin in neurons through induction of oxidative stress.

Celsi F, Svedberg M, Unger C, Cotman CW, Carrì MT, Ottersen OP, Nordberg A, Torp R.

Calcineurin is an abundant cytosolic protein that is implicated in the modulation of glutamate release. Here we show that the expression level of this enzyme is reduced in primary neuronal cultures treated with beta-amyloid. Parallel experiments in ETNA cell lines expressing SOD1 suggested that the effect of beta-amyloid on calcineurin expression is mediated by oxidative stress. The relevance of the in vitro experiments was

Argument 1 (calcineurin literature)

- Erin et al. (2003) observed that calcineurin occured as a complex with Bcl-2 in various regions of rat and mouse brain.
- Chen et al. (2003) reported about the decrease in protein ubiquitination in synaptosomes and in nonneural cells that may play role in the regulation of synaptic function by a calcineurin antagonist FK506.
- Winder and Sweatt (2001) described the critical role of protein phosphatase 1, …. and calcineurin in the activity-dependent alterations of synaptic plasticity that depends on calcineurin.

Argument 2 (autism literature)

- Fatemi et al. (2001) reported a reduction of Bcl-2 (a regulatory protein for control of programmed brain cell death) levels in autistic cerebellum.
- Belmonte et al. (2004) reviewed neuropathological studies of cerebral cortex in autism indicating abnormal synaptic and columnar structure and neuronal migration defects.
- Huber et al. (2002) showed evidences about an important role of fragile X protein, an identified cause of autism, in regulating activity-dependent synaptic plasticity in the brain.

Do identified pairs of documents point towards useful hypotheses?

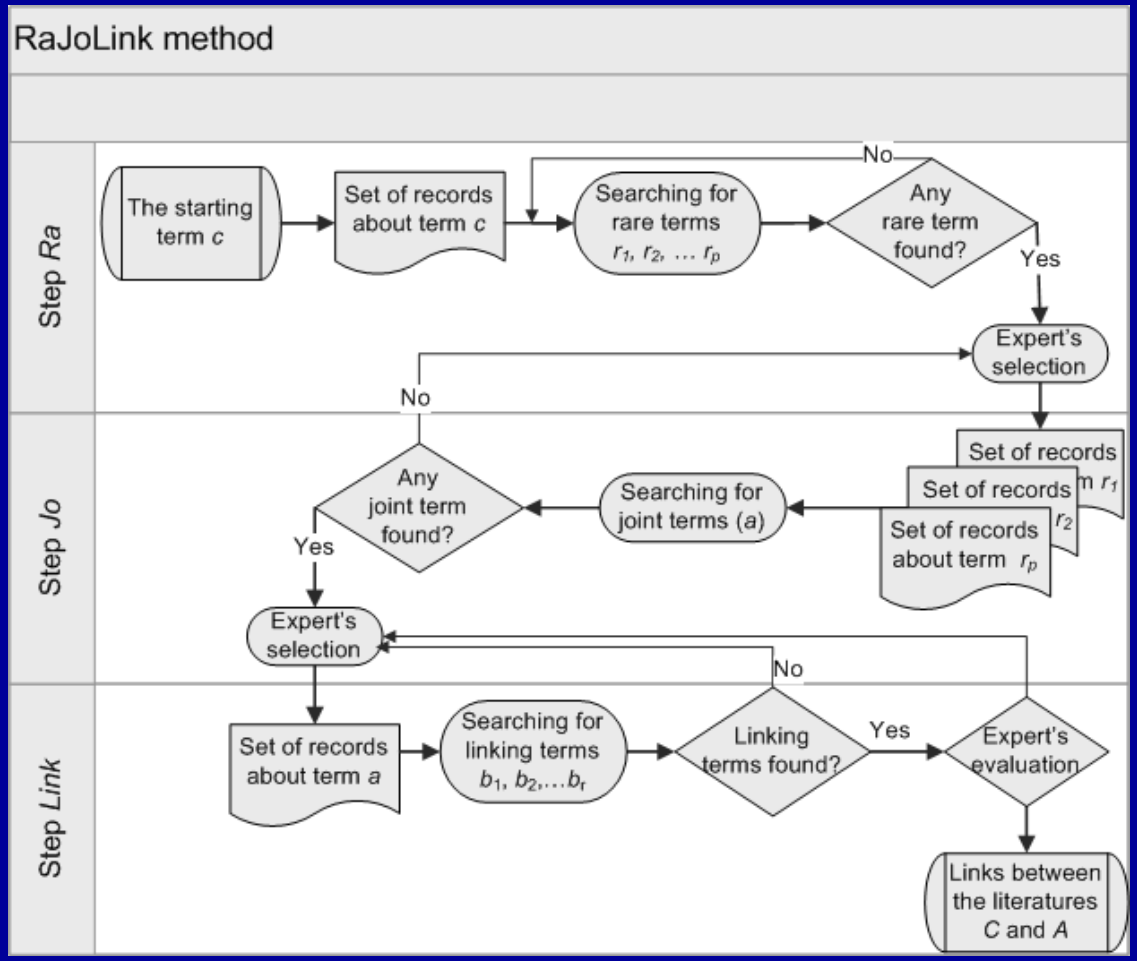
Expert's evaluation

Recent studies indicate that calcineurin participates in intracellular signaling pathways, which regulate synaptic plasticity and neuronal activities. An impaired synaptic plasticity is thought to be also a consequence of the lack of FMR1 protein in fragile X syndrome which is one of the identified cases of autism.

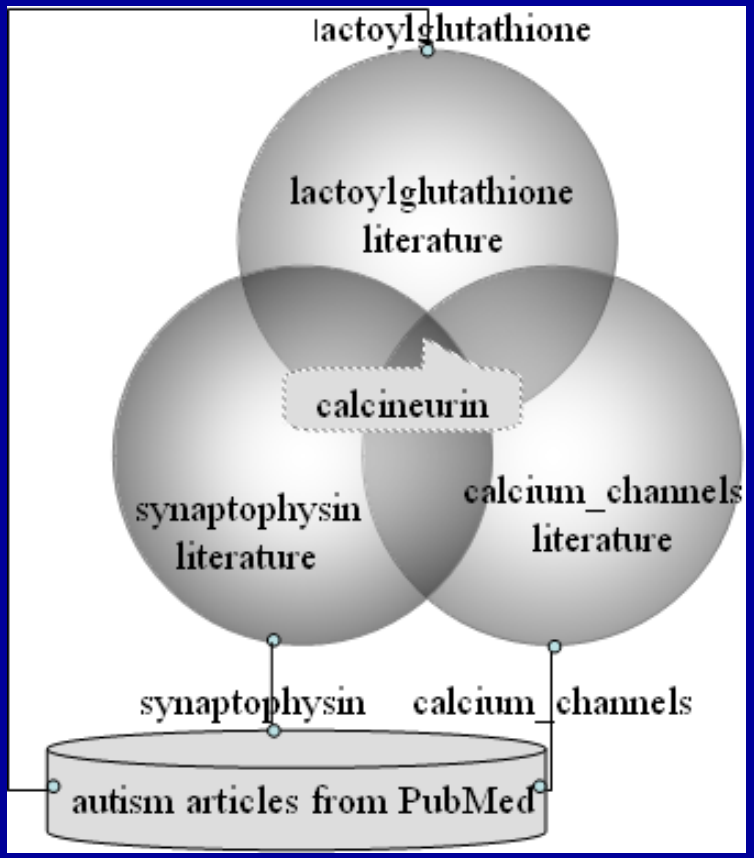
Expert's guidance

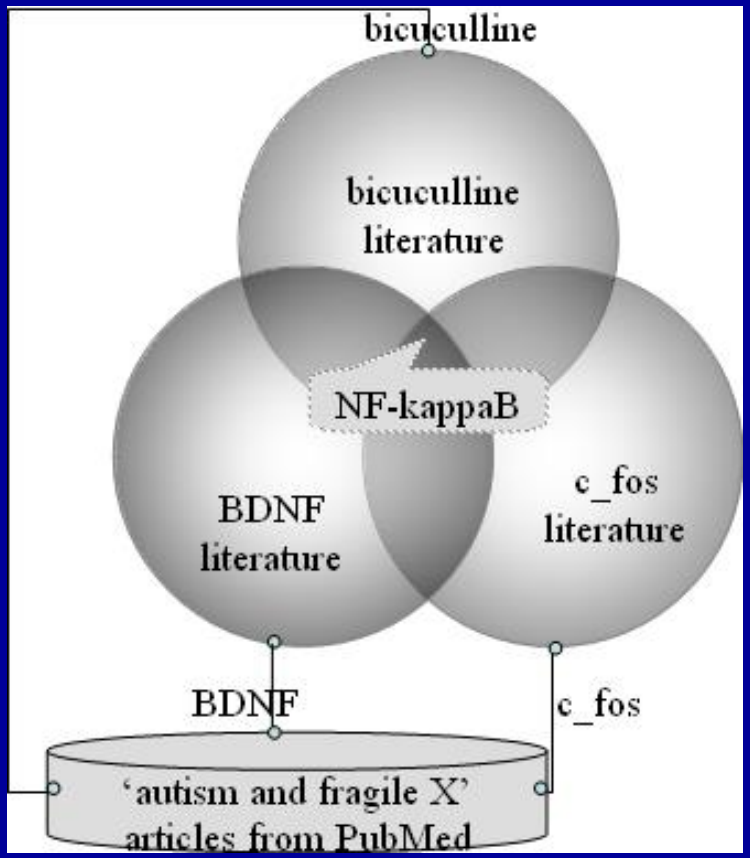
"Continue with fragil X, this would really be interesting…"

Procedures of the RaJoLink



Results obtained on autism domain, and on autism+fragile_X domain





Autism literature	NF-kappaB literature
Araghi-Niknam and Fatemi [2] showed reduction of ***Bcl-2***, an important marker of apoptosis, in frontal, parietal and cerebellar cortices of autistic individuals.	Mattson [12] reported in his review that activation of NF-kappaB in neurons can promote their survival by inducing the expression of genes encoding antiapoptotic proteins such as ***Bcl-2*** and the antioxidant enzyme Mn-superoxide dismutase.
Vargas et al. [21] reported altered ***cytokine*** expression profiles in brain tissues and cerebrospinal fluid of patients with autism.	Ahn and Aggarwal [1] reported that on activation NF-kappaB regulates the expression of almost 400 different genes, which include enzymes, ***cytokines*** (such as TNF, IL-1, IL-6, IL-8, and chemokines), adhesion molecules, cell cycle regulatory molecules, viral proteins, and angiogenic factors.
Ming et al. [13] reported about the increased urinary excretion of an ***oxidative stress*** biomarker - 8-iso-PGF2alpha in autism.	Zou and Crews [24] reported about increase in NF-kappaB DNA binding following ***oxidative stress*** neurotoxicity.

Expert's evaluation

It is thought that autism could result from an interaction between genetic and environmental factors with an oxidative stress and immunological disorders as potential mechanisms linking the two [3], [13]. Both of the mechanisms are related to NF-kappaB as the result of our analysis. The activation of the transcriptional factor NF-kappaB was shown to prevent neuronal apoptosis in various cell cultures and in vivo models [12]. Oxidative stress and elevation of intracellular calcium levels are particularly important inducers of NF-kappaB activation. In addition, various other genes are responsive to the activation of the NF-kappaB, including those for cytokines. In this way the NF-kappaB can be involved in the complex linkage between the immune system and autism [3], [21]. So, according to our analysis one possible point of convergence between "oxidative stress" and "immunological disorder" paradigm in autism is NF-kappaB.

Results - Towards better understanding of autism

- **Urbančič T., Petrič I., Cestnik B., Macedoni-Lukšič M. Literature mining: towards better understanding of autism. In: Bellazzi, R. et al., editors. AIME 2007. Proceedings of the 11th Conference on Artificial Intelligence in Medicine in Europe; Amsterdam, The Netherlands. 217-226 (2007).**

  "… according to our analysis one possible point of convergence between oxidative stress and immunological disorder paradigm in autism is NF-kappaB."

- **Naik U.S., Gangadharan, Abbagani K., Nagalla B., Dasari N., Manna S. K. A Study of Nuclear Transcription Factor –Kappa B in Childhood Autism.** *PLoS ONE* **6(5): e19488 (2011).**

  "We have noticed significant increase in NF-kB DNA binding activity in peripheral blood samples of children with autism."

  "This finding has immense value in understanding many of the known biochemical changes reported in autism."

OPEN ACCESS Freely available online

PLoS one

A Study of Nuclear Transcription Factor-Kappa B in Childhood Autism

Usha S. Naik[1], Charitha Gangadharan[2], Kanakalatha Abbagani[1], Balakrishna Nagalla[3], Niranjan Dasari[1], Sunil K. Manna[2]*

1 Department of Psychiatry, Osmania Medical College, Hyderabad, India, **2** Laboratory of Immunology, Centre for DNA Fingerprinting and Diagnostics, Nampally, Hyderabad, India, **3** National Institute of Nutrition, Hyderabad, India

Abstract

Background: Several children with autism show regression in language and social development while maintaining normal motor milestones. A clear period of normal development followed by regression and subsequent improvement with treatment, suggests a multifactorial etiology. The role of inflammation in autism is now a major area of study. Viral and bacterial infections, hypoxia, or medication could affect both foetus and infant. These stressors could upregulate transcription factors like nuclear factor kappa B (NF-κB), a master switch for many genes including some implicated in autism like tumor necrosis factor (TNF). On this *hypothesis*, it was proposed to determine NF-κB in children with autism.

Methods: Peripheral blood samples of 67 children with autism and 29 control children were evaluated for NF-κB using electrophoretic mobility shift assay (EMSA). A phosphor imaging technique was used to quantify values. The fold increase over the control sample was calculated and statistical analysis was carried out using SPSS 15.

Results: We have noted significant increase in NF-κB DNA binding activity in peripheral blood samples of children with autism. When the fold increase of NF-κB in cases (n = 67) was compared with that of controls (n = 29), there was a significant difference (3.14 vs. 1.40, respectively; $p < 0.02$).

Conclusion: This finding has immense value in understanding many of the known biochemical changes reported in autism. As NF-κB is a response to stressors of several kinds and a master switch for many genes, autism may then arise at least in part from an NF-κB pathway gone awry.

Citation: Naik US, Gangadharan C, Abbagani K, Nagalla B, Dasari N, et al. (2011) A Study of Nuclear Transcription Factor-Kappa B in Childhood Autism. PLoS ONE 6(5): e19488. doi:10.1371/journal.pone.0019488

Editor: Monica Uddin, University of Michigan, United States of America

Received September 29, 2010; **Accepted** April 8, 2011; **Published** May 9, 2011

Funding: The study was funded by a Central Government Grant from the Department of Biotechnology, Ministry of Science and Technology, Government of India. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: manna@cdfd.org.in

Open discovery - Conclusions

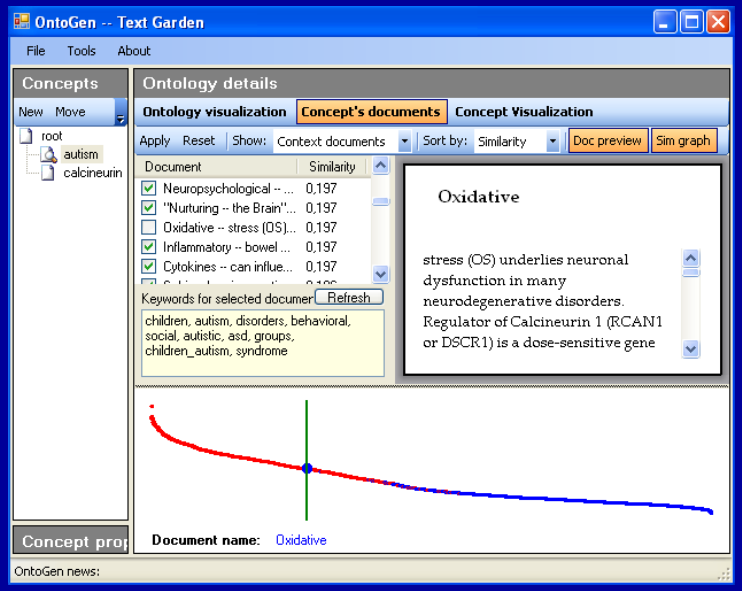
- It is worthwhile to explore rare terms for generation of hypotheses (RaJoLink method).
- Expert's involvement is crucial for speeding up the process (selections) and for evaluations of candidate hypotheses, but it's well supported by the method.
- Expert evaluation confirmed the relevance of discovered relations in the autism domain (published in medical journals).
- RaJoLink: Method for finding seeds of future discoveries in nowadays literature

Outline

- **Part 1: Introduction**
  - Creativity and knowledge discovery
  - Literature as a source of cross-context knowledge discoveries
  - Open vs. closed discovery
- **Part 2: Open discovery framework**
  - Exploring rarity (RaJoLink)
  - Case 1: RaJoLink in the Migraine
  - Case 2: With RaJoLink Towards better understanding of Autism
- **Part 3: Closed discovery framework**
  - Speeding-up the process:
    exploring outliers, ranking candidate bridging terms, workflows, HCI (CrossBee)
- **Part 4: What next?**

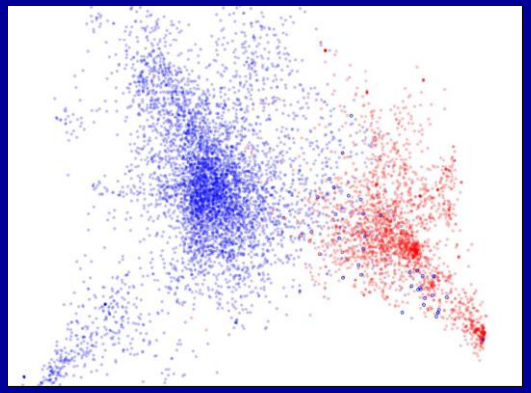
Closed discovery: improving search for bridging terms

- **Exploring outliers** (Petrič, Cestnik, Lavrač, Urbančič, 2012)
  - **Similarity graph of all documents from domains A and C, obtained with Ontogen** (Fortuna et al., 2006)
  - **Outliers from literature C are positioned among the documents from literature A.**
  - **Outliers can be successfully used for detecting bridging terms.**

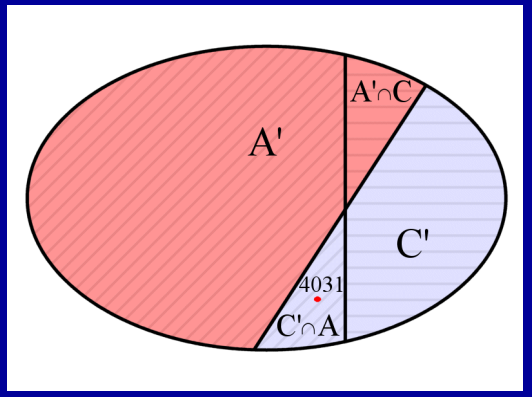


Closed Discovery – Role of Outlier documents

- The majority of bridging concepts can be found in outlier documents.
  (Sluban, Juršič, Cestnik, Lavrač, 2012)



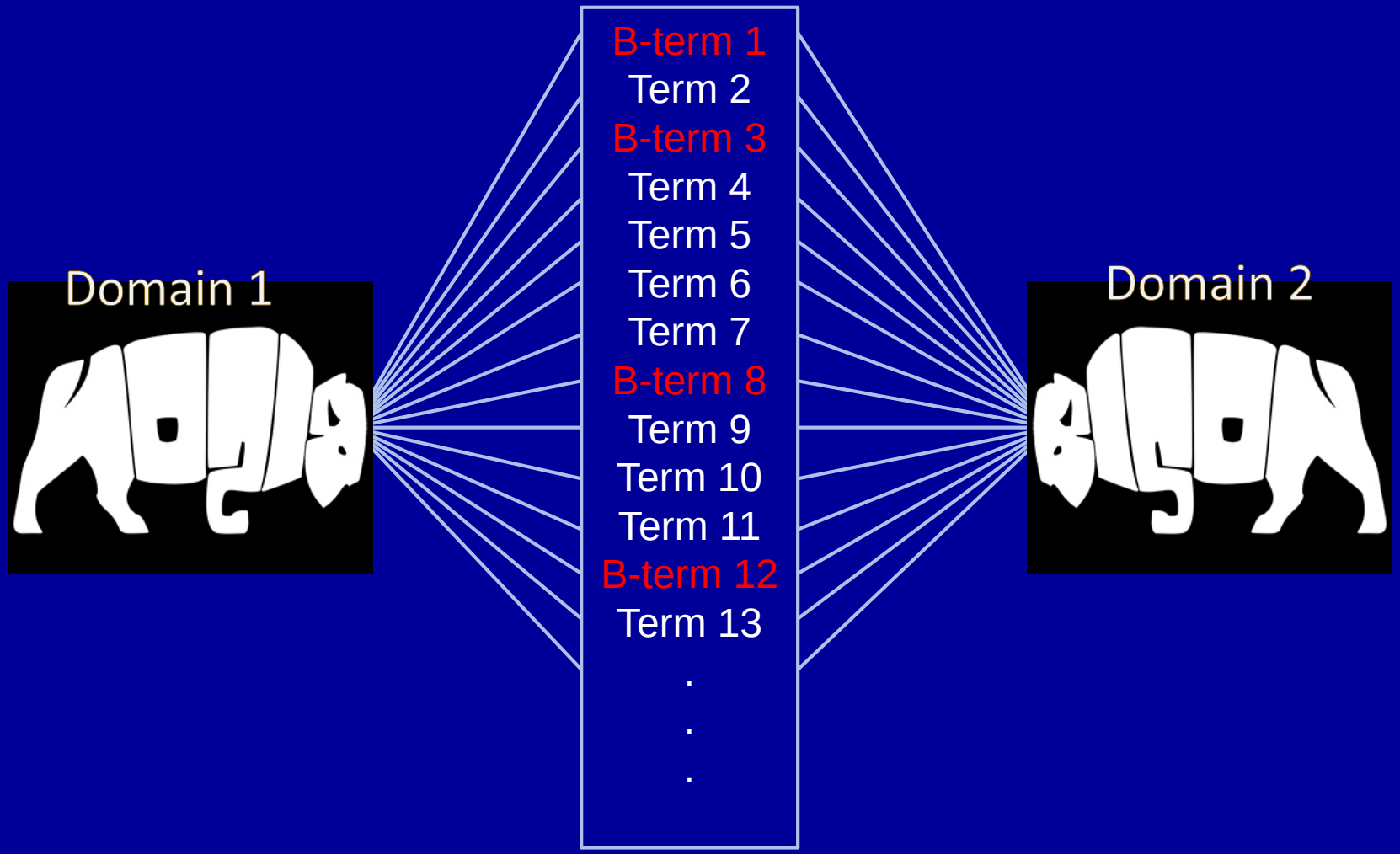
- Focusing on outliers from literatures *A* and *C* resulted in substantial search reduction.
  (Petrič, Cestnik, Lavrač, Urbančič, 2012)



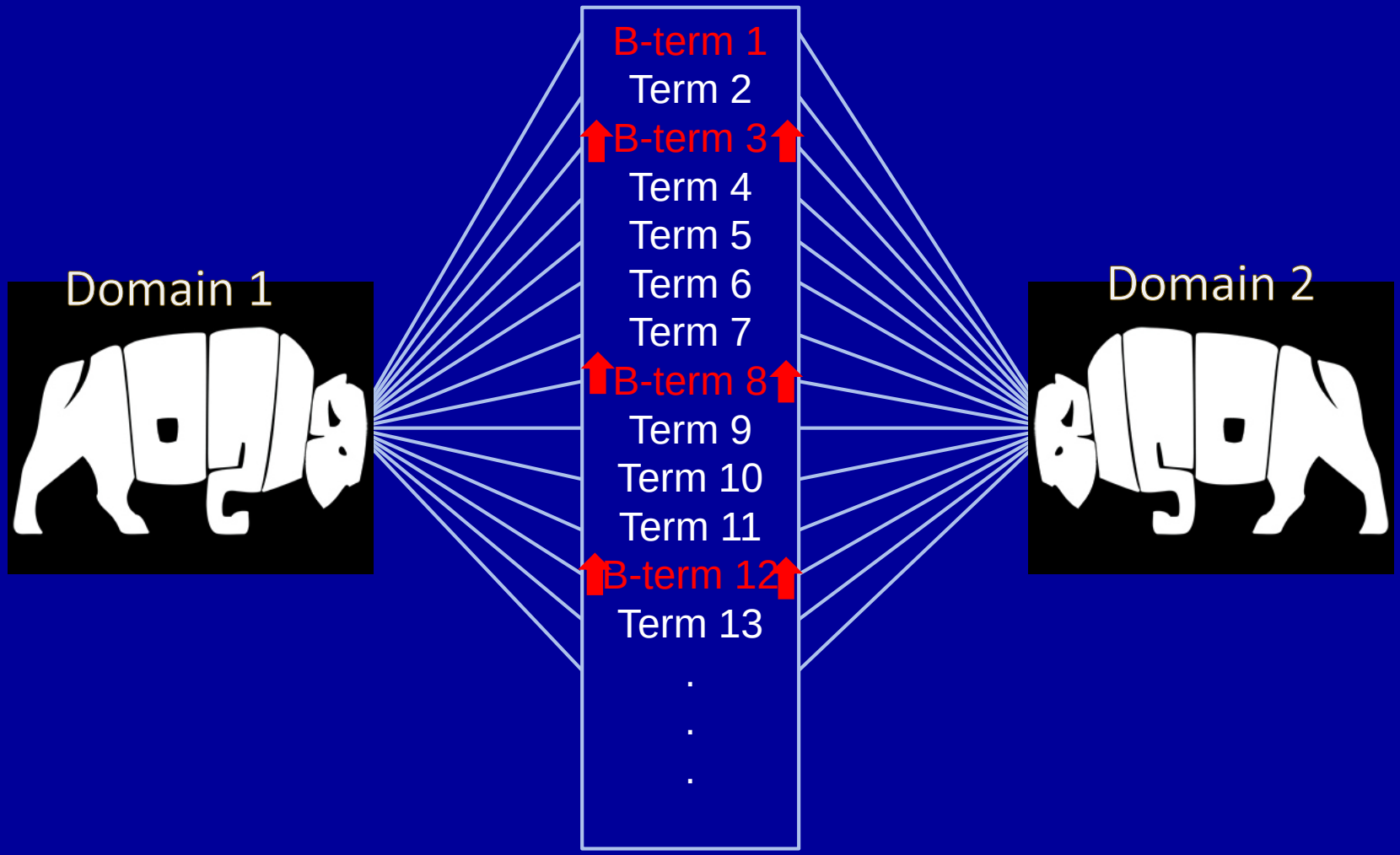
Closed discovery – improving search for bridging terms

- **Web application CrossBee (Juršič 2013):**
  - **Ranking of candidate b-terms according to their bisociation potential**
  - **Side-by-side document view, customized for expert inspection of potential cross-domain links**
  - **Standard document navigation – filtering, keyword and similarity search**
  - **Color-based domain separation scheme**
  - **User interface customization (emphasizing features on/off, …)**
  - **Hierarchical semi-automated document clustering** *(Top Circle)*

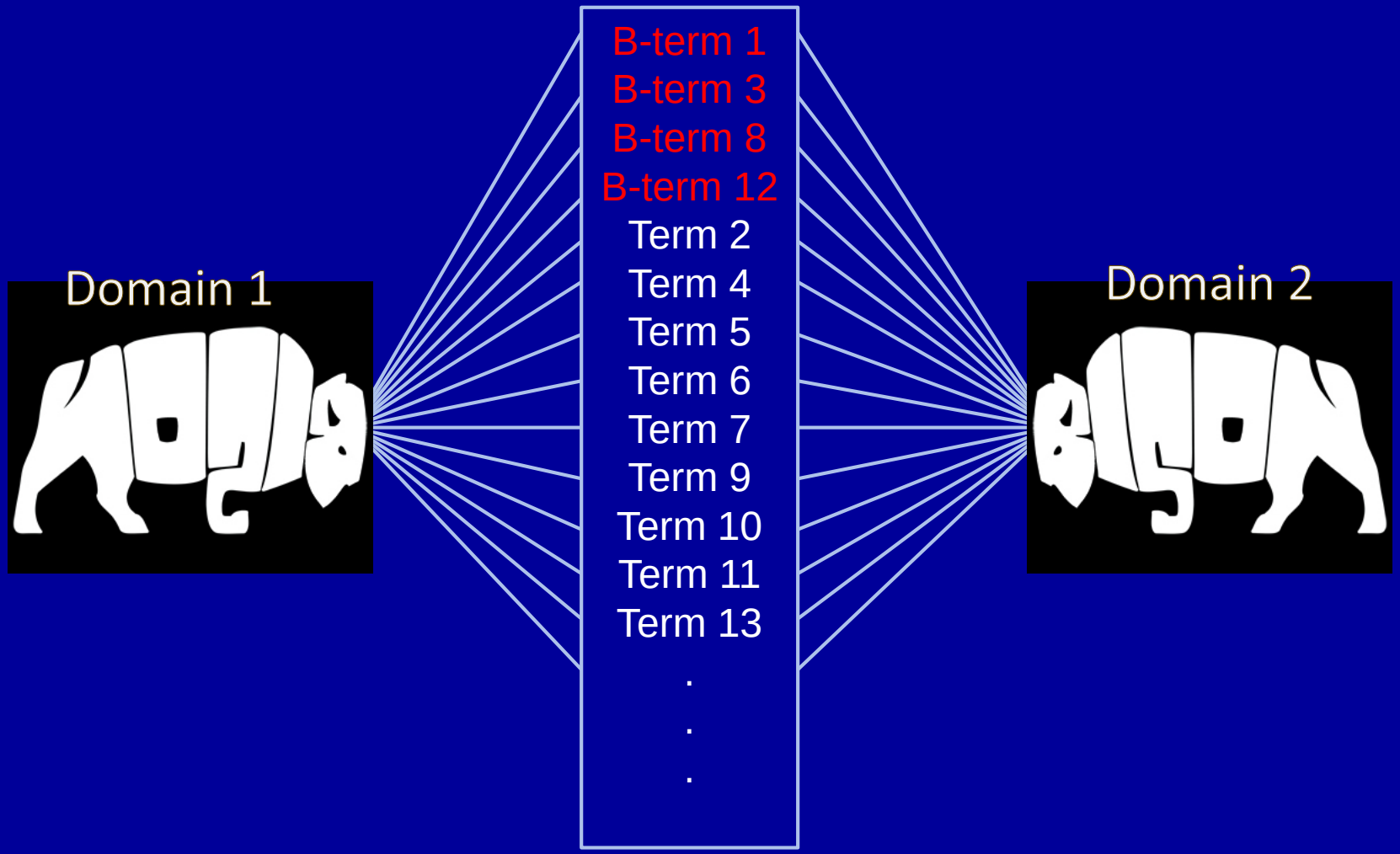
B-TERM DISCOVERY



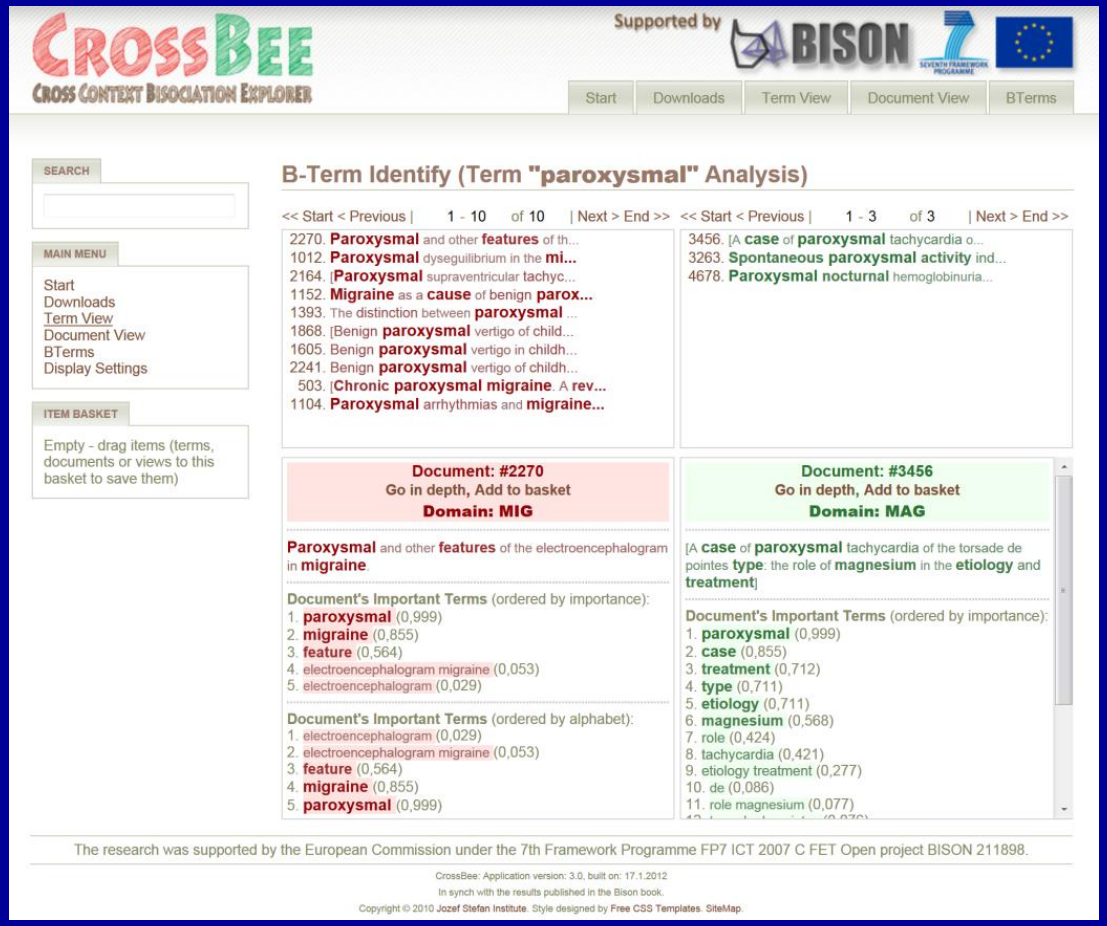
B-TERM DISCOVERY



B-TERM DISCOVERY

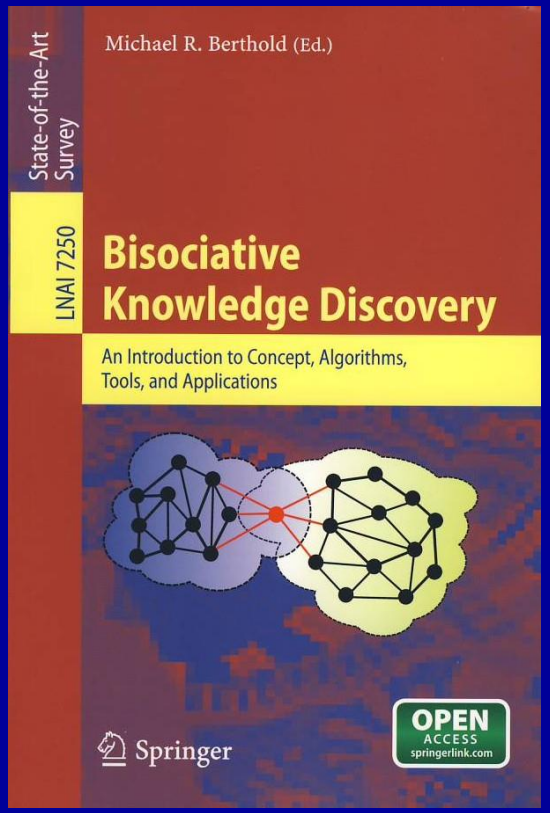


Syde-by-side document inspection in CrossBee



Conclusions

- New discoveries based on cross-domain connections uncovered from literature
- Literature mining tools
- Open discovery
  (suggesting which domains to connect)
- Closed discovery
  (supportive evidence through bridging terms)
- Applications of literature mining so far mostly in biomedical domains
  (e.g. genes and diseases, mechanisms underlying different disorders, etc.)
- Potential for other domains



What next?

- Open discovery and CrossBee – upgrades and improvements
- Continued research in computational creativity for knowledge discovery
- ConCreTe: Concept Creation Technology
- Case 3: Call for ideas