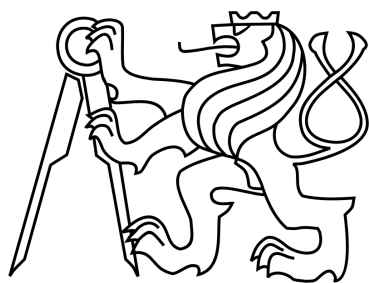




Hledání nápadů v textových zdrojích





- **Motivace:** hledání postupů, které podpoří vznik **bisociací** uvnitř bohatých (medicínských) informačních zdrojů pro hledání nových hypotéz, které vysvětlují zkoumaný fenomén, prostřednictvím bisociačních vazeb mezi různými

- Nástroje podporující objevování souvislostí:
 - **RaJoLink** (Petrič and Cestnik, 2007)
 - **OntoGen** (Fortuna et al., 2006)
 - **CrossBee** (Juršič et al., 2012)

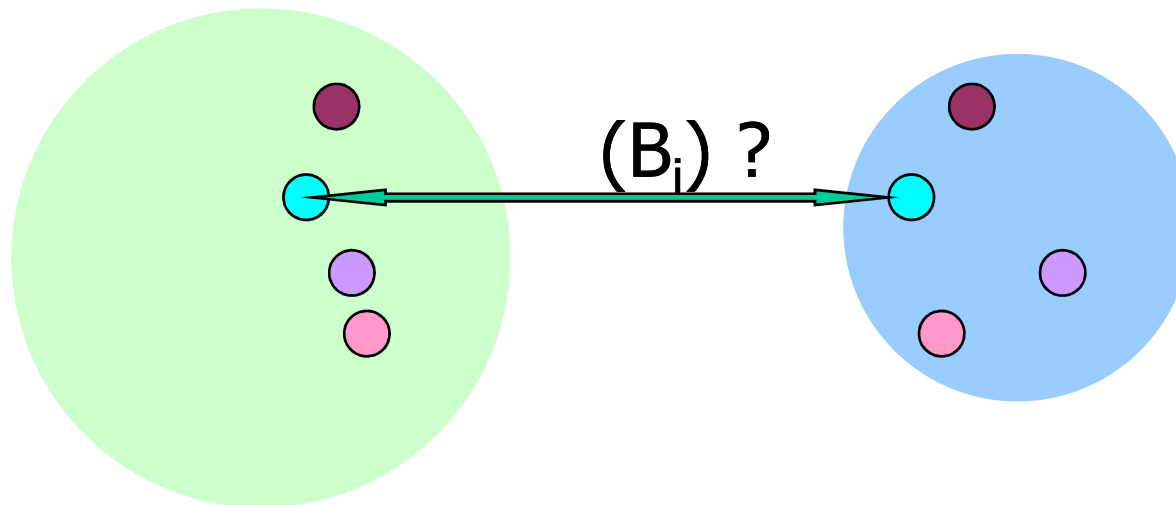
- Informační zdroje:
 - MEDLINE (<http://www.ncbi.nlm.nih.gov/sites/entrez>)
 - MeSH (<http://www.nlm.nih.gov/mesh/>)

Swansonův model hledání souvislostí v literálních zdrojích (Swanson, 1986)



Literatura o
magnesiumu (A)

Literatura o
migréně (C)



38 tis. článků

4,6 tis. článků



Argumenty objevující se v dobře vybraných článcích z různých literatur

Lit. o magnesiu

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

Lit. o migréně

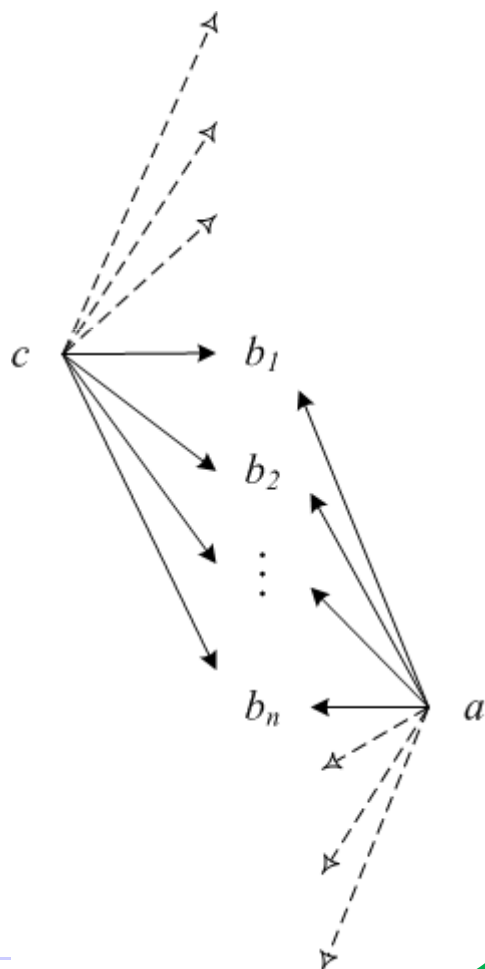
- ❖ Calcium channel blockers can prevent migraine attacks.
- ❖ Stress and Type A behaviour are associated with migraine.
- ❖ Migraine may involve sterile inflammation of the cerebral blood vessels.

Uzavřený vs. otevřený (*closed vs. open*) proces objevování vazeb (Weeber et al., 2001)



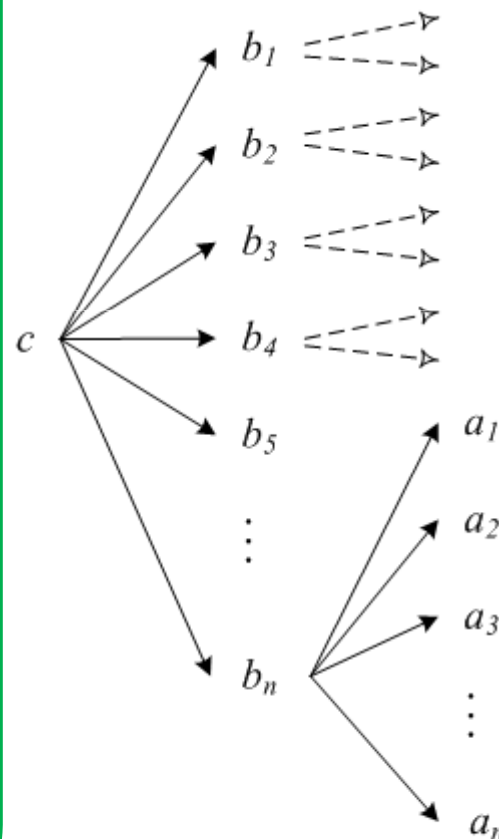
Příklad „Migréna – magnézium“

uzavřený
proces



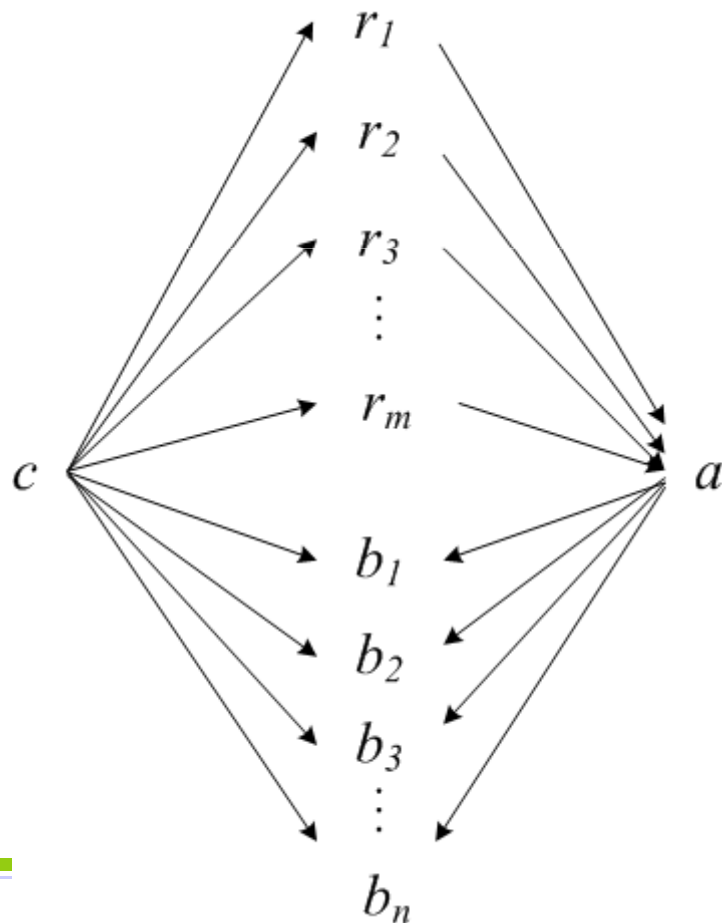
Příklad „Migréna - ?“

otevřený
proces





Combined open and closed discovery in RaJoLink (Petrič et al., 2009)



Otevřený proces pro generování hypotéz

❖ Identifikace **řidkých** (*rare*) termů pro koncept **c**

❖ Hledání **společných** (*joint*) termů pro řídké → vytvoření návrhu na koncept **a**

Uzavřený proces hledání vazeb mezi **c** a **a** – testování hypotéz

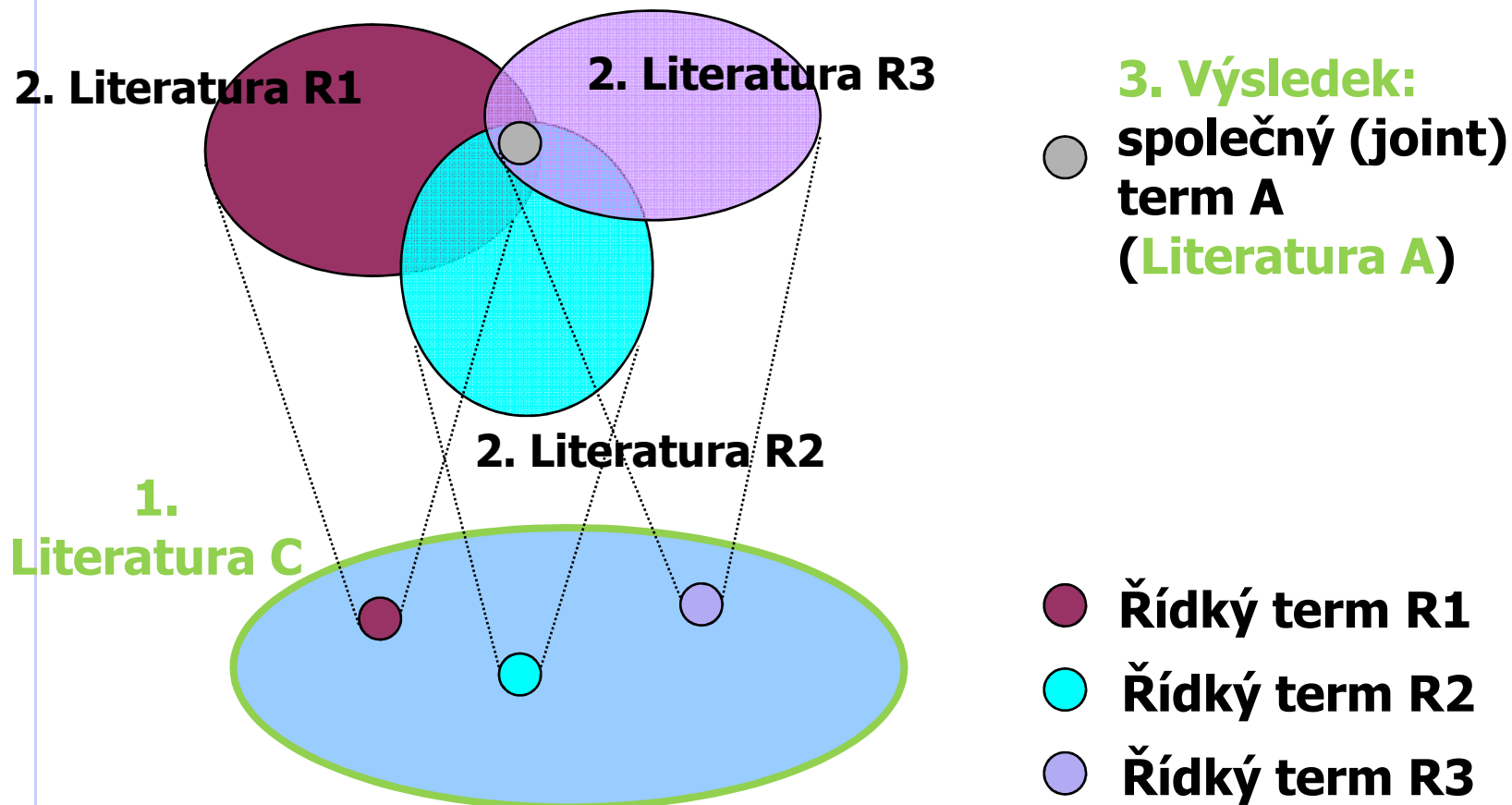
❖ Hledání **vazebních** (*linking*) termů **b**, které tvoří můstek mezi uvažovanými koncepty

Metoda RaJoLink: část "open discovery"



Hledání kandidáta na hypotézu pomocí řídkých termů.

Term je ***n*-řídký** (*rare*) pro studovanou literaturu **C**, pokud se vyskytuje v méně než ***n*** člancích v **C** (***n*** je parametr, který se často volí 1).

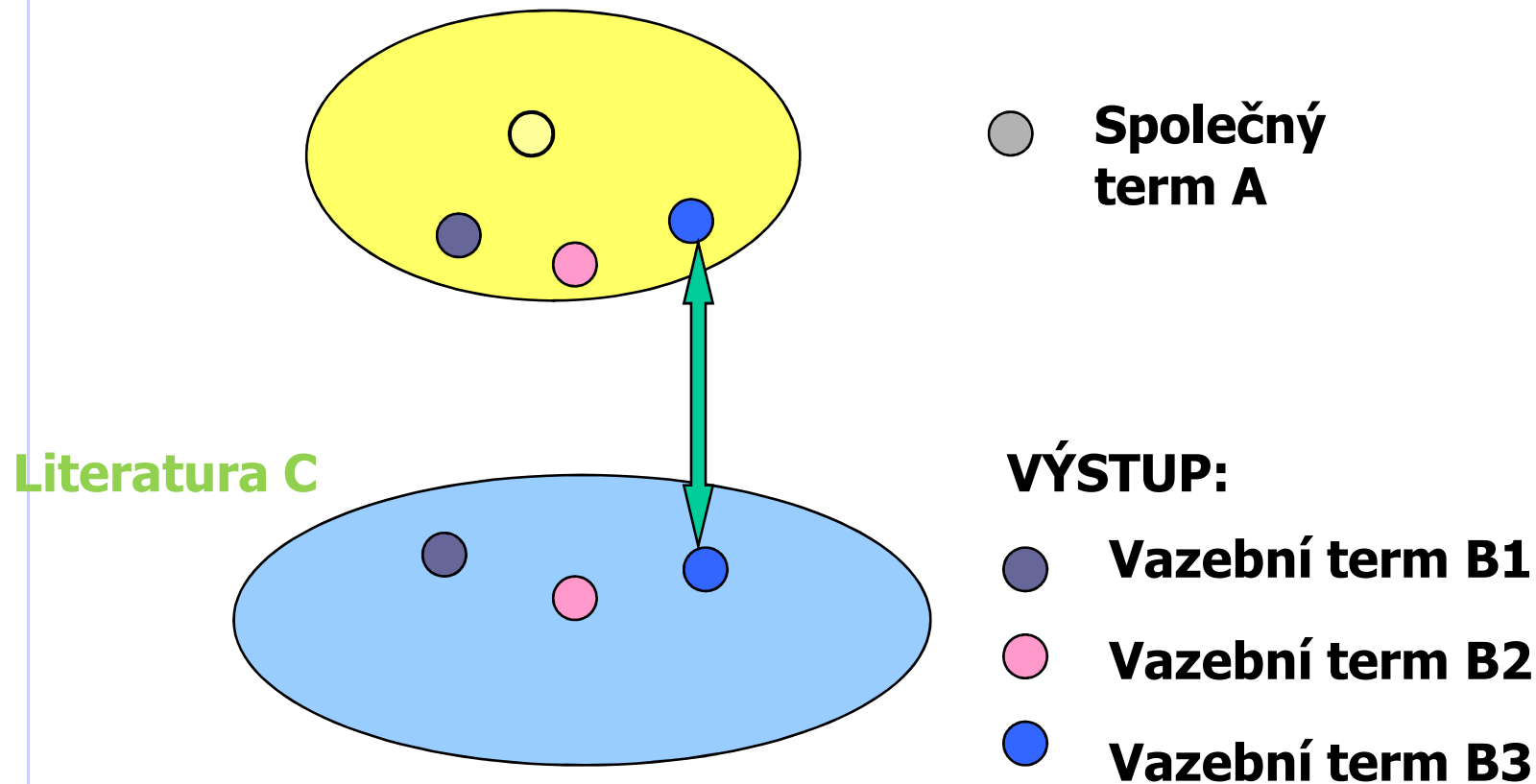


Metoda RaJoLink: uzavřené hledání



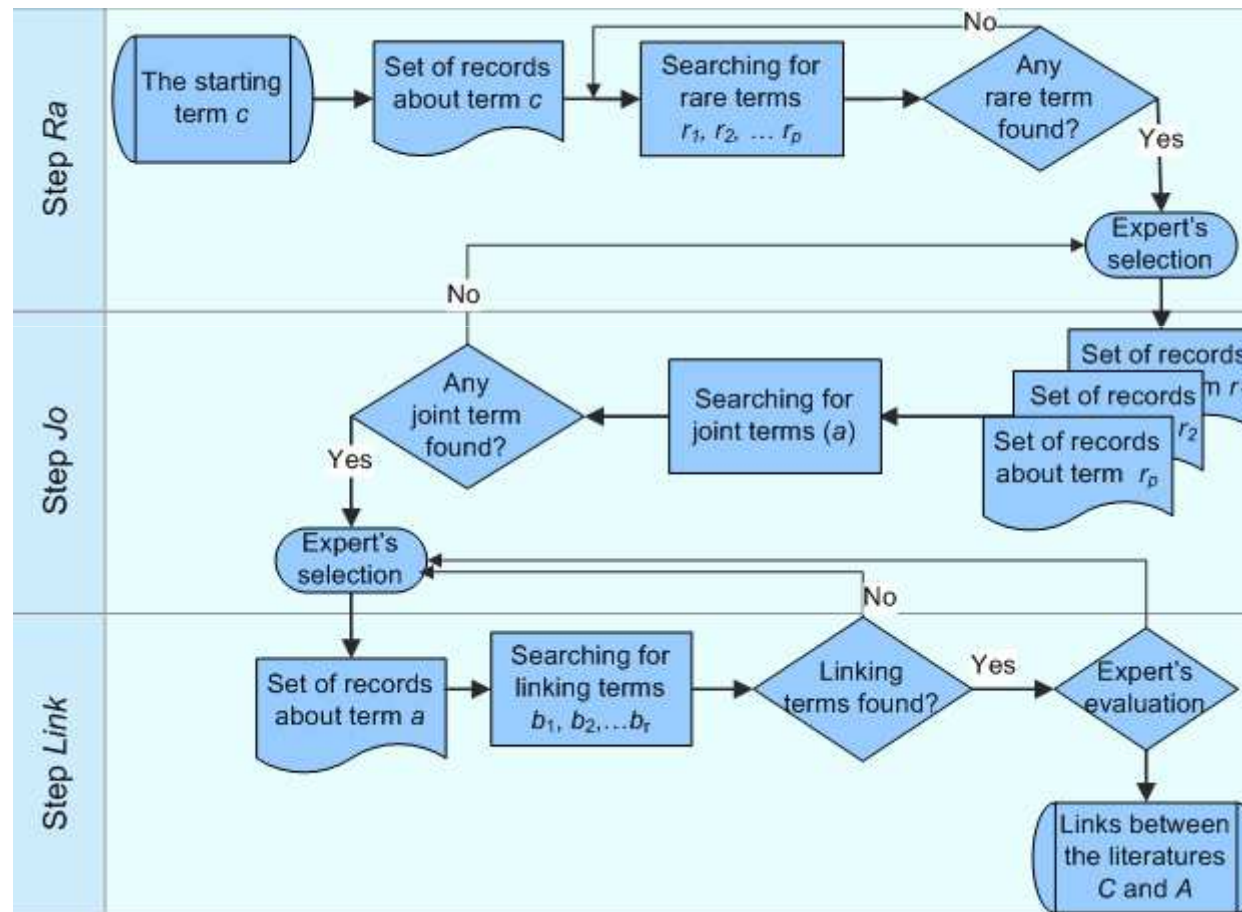
VSTUP: Literatura **A**, Literatura **B**

Literatura A vznikla jako literatura zahtnující průnik literatur pro vybrané řídké termy





The RaJoLink method's procedures (Petrič et al., 2009)





Steps of the RaJoLink method – step Ra

Step	Input	Action	Human involvement	Output
<i>Ra</i>	Set of records about domain of interest (about phenomenon C)	1.1 Extraction of texts		
		1.2 Data preprocessing		
		1.3 Identification of rare terms		
		1.4 Terms filtering	Indication of interesting rare terms	Rare terms C_{r_1} $C_{r_2} \dots C_{r_p}$

Step Ra



All Databases PubMed Nucleotide Protein Genome Structure OMIM PMC Journals Books

Search PubMed for autism Go Clear Save Search

Limits Preview/Index History Clipboard Details

Display: Abstract Show 5 Sort By Send to

All: 11761 Review: 1744

Items 1 - 5 of 11761 Page 1 of 2351 Next

- About Entrez
- Text Version
- Entrez PubMed
- Overview
- Help / FAQ
- Tutorials
- New/Noteworthy
- E-Utilities
- PubMed Services
- Journals Database
- MeSH Database
- Single Citation Matcher
- Batch Citation Matcher
- Clinical Queries
- Special Queries
- LinkOut
- My NCBI
- Related Resources
- Order Documents
- NLM Mobile
- NLM Catalog
- NLM Gateway
- TOXNET
- Consumer Health
- Clinical Alerts
- Clinicaltrials.gov
- PubMed Central

1: Behav. Modif. 2008 Feb 6 [Epub ahead of print]

View Full-Text Article at SAGE Publications

Excessive Daytime Sleep: Behavioral Assessment and Intervention in a Child with Autism.

Iselli JK, Friedm...
Institute.
School children with au...
procedu...
procedures with the b...
study adds to the liter...

PMID: 18258994 [P...

2: Trends Neurosci. 200...

REVIEWER
NEUROANATOMY o...

Amaral DG, Schum...

The M.I.N.D. Institut...

Autism spectrum disor...
deficits in social intera...
disorders including int...
lobes, amygdala and c...
studies emphasize that...
both the core and co-...
well-characterized bra...

PMID: 18258309 [P...

```
<ArticleTitle>Excessive Daytime Sleep: Behavioral
Assessment and Intervention in a Child with
Autism.</ArticleTitle>
<Pagination>
<MedlinePgn/>
</Pagination>
<Abstract>
<AbstractText>Som... excessive
daytime sleep but intervention research for this
problem has not been conducted. The present study
evaluated procedures with a 13 year old boy who has
autism and slept for prolonged periods during the day.
Classroom staff at a specialized school implemente
procedures with the boy according to an ABAB
experimental design. Intervention eliminated daytime
sleep through a 6-month follow-up assessment. The study
```

Search word: autism Number of articles: 11761 Select: Abstracts

Before date: 14 2 2008

ID	Freq	Term	MeSH code
20494	1	AMYLOIDOSIS	C10:C16:C18
20506	1	AMPLIFIER	E07
20507	1	AMPHIPHATIC	D12
20508	1	AMPHIBIAN	D12:D20
20510	1	AMOXICILLIN	D02
20514	1	AMNION	A10
20515	1	AMMONIUM	D01:D02:D03:D10:D12
20517	1	AMITROLE	D03
20520	1	AMINOTRANSFERASE	D08
20522	1	AMINOPEPTIDASE	D08
20523	1	AMINOIMIDAZOLECARBOXAMIDE	D08
20524	1	AMINOIMIDAZOLE	D03
20525	1	AMINOGLYCOSIDE	D08
20526	1	AMINOACIDURIA	C12
20534	1	AMENORRHEA	C23

1. Anatomy [A]
2. Organisms [B]
3. Diseases [C]
4. Chemicals and Drugs [D]
5. Inorganic Chemicals [D01] +
6. Organic Chemicals [D02] +
7. Heterocyclic Compounds [D03] +
8. Polymeric Compounds [D04] +
9. Macromolecular Substances [D05] +
10. Hormones, Hormone Substitutes, and Hormone Antagonists [D06] +
11. Enzymes and Coenzymes [D08] +
- Carbohydrates [D09] +
- Lipids [D10] +
- Amino Acids, Peptides, and Proteins [D12] +
- Nucleic Acids, Nucleotides, and Nucleosides [D13] +
- Complex Mixtures [D20] +
- Biological Factors [D23] +
- Chemical and Dental Materials [D25] +
- Pharmaceutical Preparations [D26] +
- Chemical Actions and Uses [D27] +



Steps of the RaJoLink method – step Jo



Step	Input	Action	Human involvement	Output
<i>Jo</i>	Sets of records about C_{r_1} $C_{r_2, \dots}$ C_{r_p}	2.1 Extraction of texts		
		2.2 Data preprocessing		
		2.3 Search for joint terms	Selection of a significant joint term	Joint term <i>a</i>

Step Jo



calcium AND channel
chromogranin
cofilin
lactoylglutathione
synaptophysin

Retrieve

2000 Abstracts

Go

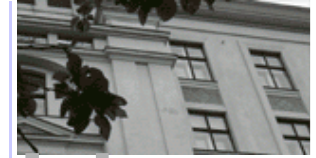
Input set of records

ex (DGC). Additional results suggest that Prf
Multiple endocrine neoplasms, including an in
We describe two cases of atypical carcinoid,
including thymoma have a much better progr
Cancer with endocrine features rarely occur:
Embryonal carcinoma (EC) cells provide a car
the monoclonal antibody A2B5, was express
Endobrevin/VAMP-8 is an R-SNARE localized
Postsynaptic density (PSD)-95, SAP102, anc
e PSD-95 antibody was shown to label exclu
Complement defense 59 (CD59) is a cell surf
cells, which normally underexpress CD59, ar
The vesicular zinc-rich synaptic systems of th
We report two cases of primary large cell ne
of the gallbladder is significant for two reaso
To clarify the neuroendocrine differentiation
r other neuroendocrine markers, including ch
Synaptic vesicle protein 2 (SV2) is a glycoprc
d SV2-immunoreactive cells. The staining pat
The literature on the neuropathology of bipc
prefrontal, and temporal cortices in BD. In th
Prior studies on receptor recycling through th
, the primary effect of early endosomal sorti
We sought to delineate differences between
es.

Results

	Term	Frequencies	Sum of frequencies	MeSH co
<input type="checkbox"/>	57 ETHYLMALDEIMIDE	5:4,1,1,2,11	19	D02:DO0
<input type="checkbox"/>	58 ELECTROPHORETIC	5:5,3,1,24,1	34	E05
<input type="checkbox"/>	59 DYE	5:25,4,4,2,12	47	D02:DO0
<input type="checkbox"/>	60 DIPLOID	5:1,5,3,1,2	12	D20:G10
<input type="checkbox"/>	61 DEXTRAN	5:2,1,1,1,6	11	D02:DO0
<input type="checkbox"/>	64 CONE	5:10,1,27,1,29	68	A08:C10
<input type="checkbox"/>	66 CHLORO	5:5,2,1,3,1	12	D02:DO0
<input type="checkbox"/>	68 CATION	5:117,6,2,2,6	133	D12:D20
<input type="checkbox"/>	69 CATHEPSIN	5:2,5,7,1,5	20	D08
<input type="checkbox"/>	70 CARBOXY	5:3,7,4,2,3	19	D08
<input checked="" type="checkbox"/>	71 CALCINEURIN	5:18,6,8,1,4	37	D08
<input type="checkbox"/>	72 C6	5:2,1,2,2,1	8	D08
<input type="checkbox"/>	73 C3	5:3,3,6,9,3	24	D08:D10
<input type="checkbox"/>	75 BRONCHIAL	5:7,26,3,2,18	56	A07:A10
<input type="checkbox"/>	76 BROMIDE	5:4,1,3,3,2	13	D01:DO0
<input type="checkbox"/>	77 BISPHOSPHATE	5:9,3,12,2,1	27	D08:D10
<input type="checkbox"/>	78 ATHEROSCLEROSIS	5:14,3,3,1,1	22	C10:C10
<input type="checkbox"/>	80 ALKALINE	5:3,21,4,3,16	47	D01:DO0
<input type="checkbox"/>	81 A4	5:6,2,1,2,1	12	D08:D10
<input type="checkbox"/>	82 ZIPPER	4:0,1,1,1,1	4	D12:G00
<input type="checkbox"/>	83 XENOGRAFT	4:0,17,1,1,3	22	E05:E07
<input type="checkbox"/>	85 WEDGE	4:4,4,1,0,3	12	G09:K00
<input type="checkbox"/>	86 VISCOSITY	4:4,0,7,3,1	15	G09:H00
<input type="checkbox"/>	87 VINCRISTINE	4:1,3,0,1,5	10	D03
<input type="checkbox"/>	88 VEGF	4:1,15,5,0,10	31	D08:D10
<input type="checkbox"/>	89 VASODILATOR	4:32,4,2,0,1	39	D01:D20
<input type="checkbox"/>	91 UTERUS	4:2,8,0,2,4	16	A05:CO0
<input type="checkbox"/>	92 UROTHELIUM	4:1,3,1,0,2	7	A10
<input type="checkbox"/>	94 UREMIC	4:2,1,0,2,3	8	C12
<input type="checkbox"/>	95 ULTRACENTRIFUGATION	4:0,1,3,1,1	6	E05
<input type="checkbox"/>	97 TURBULE	4:12,23,0,0,20	64	A05:C10

Number of target terms: 3715 All terms: 16985



Steps of the RaJoLink method – step Link

Step	Input	Action	Human involvement	Output
<i>Link</i>	Joint set of records about <i>a</i> and articles about <i>c</i>	3.1 Extraction of texts		
		3.2 Data preprocessing		
		3.3 Identification of content related <i>A</i> and <i>C</i> records		
		3.4 Search for linking terms <i>b</i>	Selection of meaningful linking terms	Linking terms $B_1,$ B_2, \dots, B_r

Step Link



Step Link

File Help

Previous step

calcineurin

Retrieve

5000 Titles Go

Input set of records

Sirolimus and cyclosporin for renal trans
 Immunolocalization of calcineurin and FI
 Constitutive and acquired resistance to
 3,5-Bis(trifluoromethyl)pyrazoles: a nov
 Advances in the molecular mechanisms
 GABA(A) receptor modulation in rat ce
 Phosphatidylinositol 4-phosphate 5-kinz
 Quantitative immunogold localization of
 Clopidogrel-associated thrombotic thror
 Connexin43 phosphorylation state and
 Hepatitis C after liver transplantation.
 Integration of calcineurin and MEF2 sigi
 Requirement for integration of phorbol
 The effect of calcineurin inhibitors and
 Competitive inhibition of NMDA receptc
 A calcineurin-NFATc3-dependent path
 [A second target of cyclosporin A and
 Cabin1 represses MEF2-dependent Nur
 Continuing improvement in cadaver doi
 A decade of living donor transplantati
 Increased protein synthesis after T cell
 Immunophilins may limit calcineurin inh
 beta-Adrenergic pathway induces apoc

Results

	Term	Frequencies	Sum of frequencies	MeSH codes
1379	IMBALANCE	2:1,8	9	C18:G13
1331	INFANT	2:4,85	89	C10:C16:C17:C1
1298	INSULIN	2:44,5	49	A03:C10:C18:D0
1251	IRON	2:2,4	6	C15:C18:D01:D0
1234	JUVENILE	2:2,9	11	C01:C04:C05:CC
1174	LIPID	2:9,4	13	C08:C15:C16:C1
1137	MAGNESIUM	2:6,10	16	C18:D01:D02:D0
1069	METABOLIC	2:13,17	30	C05:C10:C16:C1
1068	METABOLISM	2:22,26	48	C16:C18:D12:EC
1047	MILK	2:1,3	4	A12:B06:C18:C2
1039	MITOCHONDRIAL	2:51,22	73	A11:C05:C18:D0
933	NON	2:26,71	97	A08:A13:B01:BC
916	NUTRITION	2:1,6	7	C13:C18:E02:EC
915	NUTRITIONAL	2:2,9	11	C18:E02:E05:GC
913	OBESITY	2:4,2	6	C08:C18:D27
893	ONSET	2:23,54	77	C05:C10:C16:C1
859	OXIDATIVE	2:29,13	42	C18:D12:G04:GI
823	PATHOLOGIC	2:5,1	6	C05:C10:C18:C2
773	PHOSPHORYLATION	2:95,2	97	C18:D12:G06
722	POTASSIUM	2:16,1	17	C18:D01:D02
642	PROTEIN	2:423,85	508	A11:C06:C10:C:
583	REACTIVE	2:10,3	13	C01:C10:C14:C1
523	REPAIR	2:8,6	14	C18:D08:D12:EC
505	RESISTANCE	2:23,5	28	C05:C08:C15:C1
504	RESISTANT	2:18,6	24	B01:C01:C05:C1
499	RESPIRATORY	2:3,5	8	A02:A04:A08:BI
361	SLOW	2:27,7	34	A02:B01:C02:C:
320	STABLE	2:33,1	34	C04:C18:D12:EC
312	STATE	2:17,33	72	C10:C18:C23:D:
288	STRESS	2:92,59	151	A11:C10:C12:C:
269	SUDDEN	2:4,3	7	C09:C14:C18:C2
233	SYNDROME	2:80,876	956	B04:C01:C02:CC
137	TRANSPORT	2:13,4	17	A11:C10:C12:C:
38	VITAMIN	2:2,18	20	C05:C15:C18:D0

Number of target terms: 68 All terms: 2426

Step Link - alternative



OntoGen -- Text Garden

File Tools About

Concepts

New Move

- root
 - autism
 - calcineurin

Ontology details

Ontology visualization **Concept's documents** **Concept Visualization**

Apply Reset Show: Context documents Sort by: Similarity Doc preview Sim graph

Document	Similarity
<input checked="" type="checkbox"/> Neuropsychological -- ...	0,197
<input checked="" type="checkbox"/> "Nurturing -- the Brain"...	0,197
<input type="checkbox"/> Oxidative -- stress (OS)...	0,197
<input checked="" type="checkbox"/> Inflammatory -- bowel ...	0,197
<input checked="" type="checkbox"/> Cytokines -- can influe...	0,197
<input type="checkbox"/> ...	0,100

Keywords for selected document Refresh

children, autism, disorders, behavioral, social, autistic, asd, groups, children_autism, syndrome

Oxidative

stress (OS) underlies neuronal dysfunction in many neurodegenerative disorders. Regulator of Calcineurin 1 (RCAN1 or DSCR1) is a dose-sensitive gene

Document name: Oxidative

OntoGen news:



❖ Open discovery:

- ◆ RaJoLink represents a more interdisciplinary approach to hypotheses generation that bridges the overspecialization in the sciences. We provide connections between biomedical literature by analysis and explanation of rare terms.

❖ Closed discovery:

- ◆ With the combination of outlier detection and high frequency analysis approach we demonstrated that outlying documents could be used as a heuristic guidance to speed-up the search for the linking terms and alleviate the burden on the expert when hypotheses have to be tested.

❖ Recent experiments:

- ◆ Detection of published evidence of autism findings that coincide with specific calcineurin and NF-kappaB observations (Petrič et al., 2007, Urbančič et al., 2007).
- ◆ The gold standard evaluation: RaJoLink led to the Swanson's relation of magnesium with migraine and to other three discoveries important for migraine.



- ❖ Petrič, I.; Urbančič, T.; Cestnik, B. Discovering hidden knowledge from biomedical literature. *Informatica* 31(1):15-20 (2007).
- ❖ Petrič, I.; Urbančič, T.; Cestnik, B. Literature mining: potential for gaining hidden knowledge from biomedical articles, In: Bohanec, M.; Gams, M.; Rajkovič, V.; Urbančič, T.; Bernik, M.; Mladenič, D. et al., editors. IS-2006. Proceedings of the 9th International multi-conference Information Society; Ljubljana, Slovenia. 52-55 (2006).
- ❖ Petrič, I.; Urbančič, T.; Cestnik, B.; Macedoni-Lukšič, M. Literature mining method RaJoLink for uncovering relations between biomedical concepts. *Journal of Biomedical Informatics* 42(2): 219-227 (2009).
- ❖ Urbančič, T.; Petrič, I.; Cestnik, B.; Macedoni-Lukšič, M. Literature mining: towards better understanding of autism. In: Bellazzi R; Abu-Hanna A; Hunter J, editors. AIME 2007. Proceedings of the 11th Conference on Artificial Intelligence in Medicine in Europe; Amsterdam, The Netherlands. 217-226 (2007).