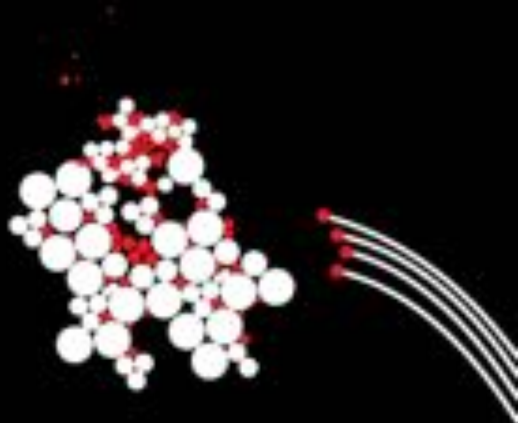
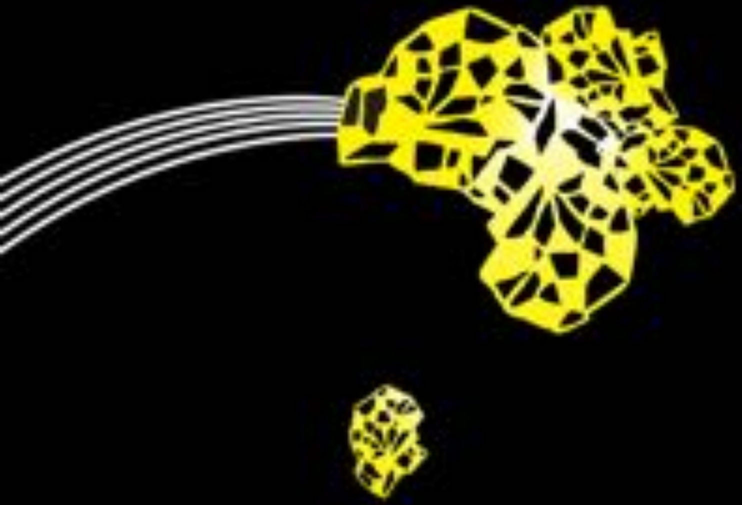
UNIVERSITY OF TWENTE.



Conceptual Indexing

Dolf Trieschnigg, Djoerd Hiemstra & Theo Huibers

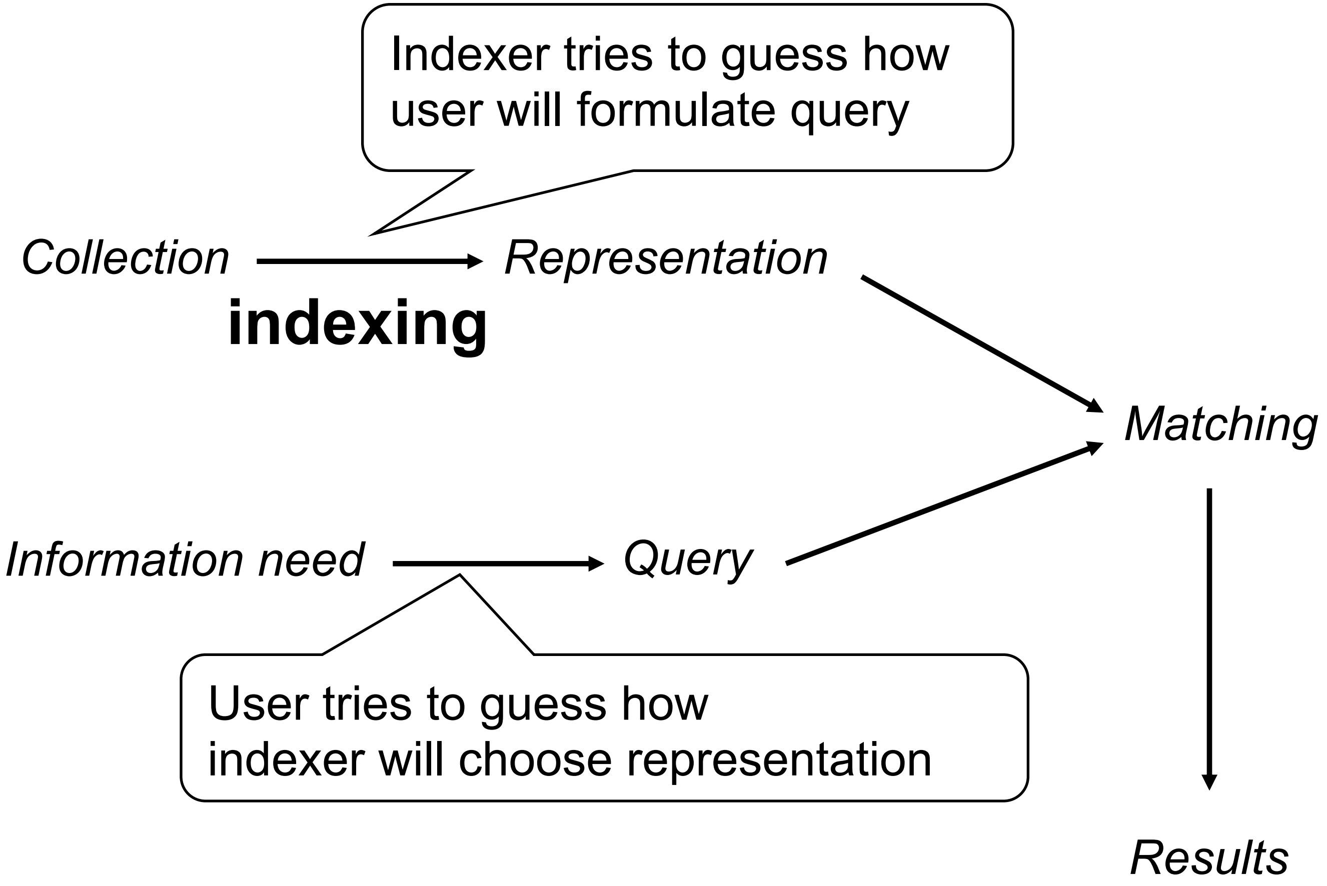


Overview

- Characteristics of indexing & indexing languages
- Basic measures of performance
- Some *crowdexing*
- Indexing in two domains
  - Biomedicine
    - What is biomedicine
    - Biomedical terminology and IR
    - PubMed & MeSH
  - Folktales

With pop quizes!

Characteristics of indexing & indexing languages



Indexer tries to guess how user will formulate query

Collection → Representation

indexing

Matching

Information need → Query

User tries to guess how indexer will choose representation

Results

Characteristics of indexing & indexing languages

- Indexing characteristics
  - Automatic vs. manual
  - Exhaustivity: number of topics indexed
- Index language
  - Controlled vs. uncontrolled vocabulary
  - Specificity: level of precision
- Types of retrieval
  - Exact match (set) vs best match (ranked)

Basic measures of performance: precision and recall

- System: returns documents (or not)
- User: finds document relevant (or not)

- **Precision**: ONLY relevant results
- **Recall**: ALL relevant results

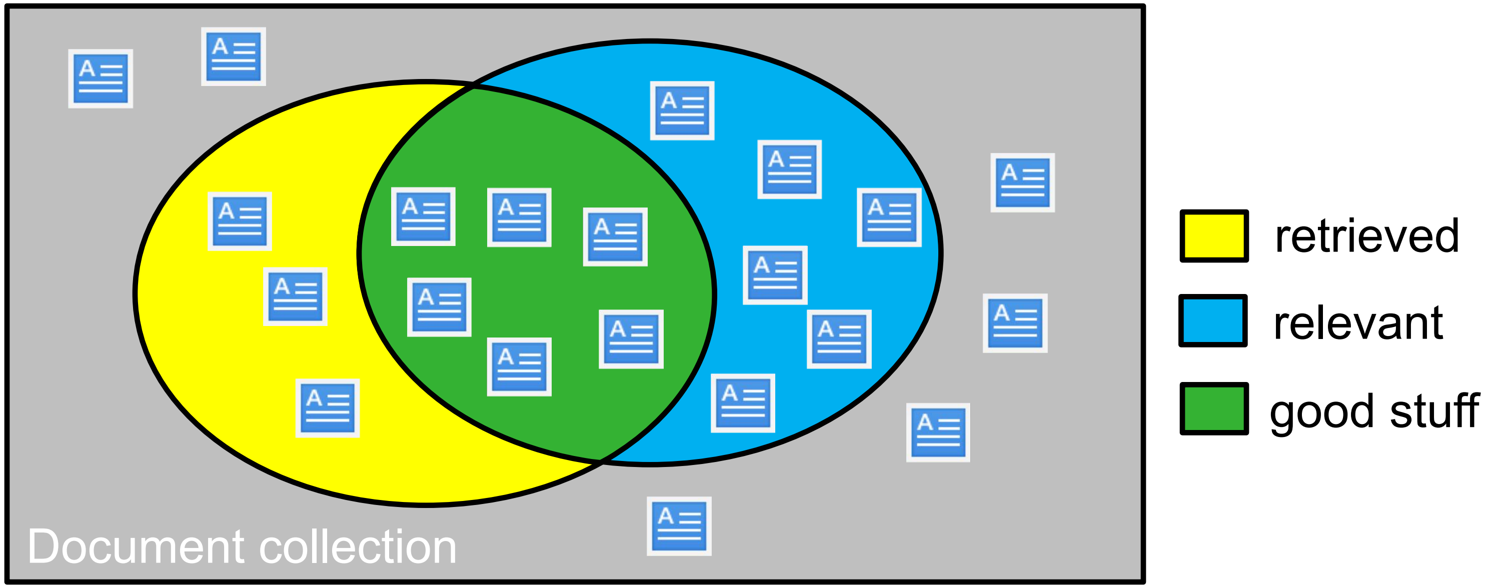
Note: for ***ranked*** retrieval other (related) measures exist!

Basic measures of performance: precision and recall

		System says:	
		Match	No match
User says:	Relevant	True positives *(#TP)*	False negatives *(#FN)*
	Not relevant	False positives *(#FP)*	True negatives *(#TN)*

$$Recall = \frac{\#TP}{\#TP + \#FN} \qquad Precision = \frac{\#TP}{\#TP + \#FP}$$

Basic measures of performance: precision and recall



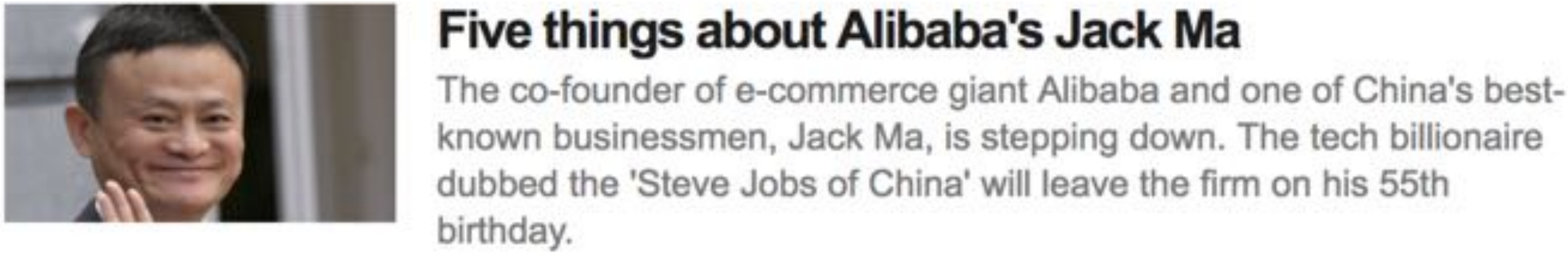
- **Precision**: fraction of relevant retrieved documents 🟩 / 🟨
  - Only relevant
- **Recall**: fraction of retrieved relevant documents 🟩 / 🟦
  - All relevant

Quiz

- How can you achieve a recall of 1?
- How can you achieve a precision of 1?
- Can you measure the precision/recall of a Google search result?
- Is precision or recall important for:
  - News search?
  - Patent search?
  - Product search?

Let’s do some *crowdexing*

- Visit
  - http://dolf.trieschnigg.nl/crowdexing
  - http://goo.gl/x62XPq



Five things about Alibaba's Jack Ma

The co-founder of e-commerce giant Alibaba and one of China's best-known businessmen, Jack Ma, is stepping down. The tech billionaire dubbed the 'Steve Jobs of China' will leave the firm on his 55th birthday.

Some discussion questions

- What terms to use: words, phrases, entities?
- Which terms to include?
- Are all terms equally important?
- How to deal with numbers?
- How to deal with word variations?

Some observations (hopefully ;-))

- You chose an indexing unit, with a certain **specificity**
- You made a **selection** of words to include, resulting in a certain **exhaustivity**
  - Probably you **don't agree**
- Some terms are **more important** than others
- Important information is **implicit**
- Terms can be **ambiguous**
- How **consistent** do you think you are?

Let's automate this

- Extract index terms automatically: ***tokenization***

Copyright 2001 by Randy Glasbergen.
www.glasbergen.com

"The new automated ordering system has really speeded up our business. We're losing customers faster than ever."

Tokenization example

Get indexing terms from text automatically

1. Lowercase text
   "US" is the same as "us"
2. Extract *words*
   "Hepatitus-A"
3. Stopword removal
   "To be or not to be"
4. Stemming
   University → Univers
   Universe → Univers

Five things about Alibaba's Jack Ma The co-founder of e-commerce giant Alibaba and one of China's best-known businessmen, Jack Ma, is stepping down. The tech billionaire dubbed the 'Steve Jobs of China' will leave the firm on his 55th birthday.

Tokenization

Get indexing terms from text automatically

1. **Lowercase text**
   "US" is the same as "us"
2. Extract *words*
   "Hepatitus-A"
3. Stopword removal
   "To be or not to be"
4. Stemming
   University → Univers
   Universe → Univers

five things about alibaba's jack ma the co-founder of e-commerce giant alibaba and one of china's best-known businessmen, jack ma, is stepping down. the tech billionaire dubbed the 'steve jobs of china' will leave the firm on his 55th birthday.

Tokenization

Get indexing terms from text automatically

1. Lowercase text
"US" is the same as "us"
2. **Extract *words***
"Hepatitus-A"
3. Stopword removal
"To be or not to be"
4. Stemming
University → Univers
Universe → Univers

five things about alibaba 's
jack ma the co-founder of e-commerce giant
alibaba and one of china 's best-known businessmen ,
jack ma , is stepping down . the tech billionaire
dubbed the 'steve jobs of china
' will leave the firm on his 55th
birthday .

Tokenization

Get indexing terms from text automatically

1. Lowercase text
   "US" is the same as "us"
2. Extract *words*
   "Hepatitus-A"
3. **Stopword removal**
   "To be or not to be"
4. Stemming
   University → Univers
   Universe → Univers

five things ~~about~~ alibaba 's
jack ~~ma~~ ~~the~~ co-founder ~~of~~ e-commerce giant
alibaba ~~and~~ one ~~of~~ china 's best-known businessmen ,
jack ~~ma~~ , ~~is~~ stepping ~~down~~ . ~~the~~ tech billionaire
dubbed ~~the~~ 'steve jobs ~~of~~ china ' ~~will~~ leave ~~the~~ firm ~~on~~ ~~his~~ 55th birthday .

Tokenization

Get indexing terms from text automatically

1. Lowercase text
   "US" is the same as "us"
2. Extract *words*
   "Hepatitus-A"
3. Stopword removal
   "To be or not to be"
4. **Stemming**
   University → Univers
   Universe → Univers

five thing~~s~~ alibaba 's jack co-found~~er~~ e-commerc~~e~~ giant alibaba one china 's best-known businessmen , jack , step~~ping~~ . tech billionair~~e~~ dub~~bed~~ 'steve job~~s~~ china ' leav~~e~~ firm 55th birthday .

Tokenization

Get indexing terms from text automatically

1. Lowercase text
"US" is the same as "us"
2. Extract *words*
"Hepatitus-A"
3. Stopword removal
"To be or not to be"
4. **Stemming**
University → Univers
Universe → Univers

five thing alibaba 's jack co-found e-commerc giant alibaba one china 's best-known businessmen , jack , step . tech billionair dub 'steve job china ' leav firm 55th birthday .

Case study 1: biomedicine

What is biomedicine?

developmental biology

oncology

cell biology

biology

biochemistry

botany

bioinformatics

chemistry

Biomedicine/ Life sciences

computational neuroscience

immunology

biomechanics

genomics

zoology

medicine

biophysics

microbiology

biotechnology

ecology

physics

food science

genetics

systems biology

immunogenetics

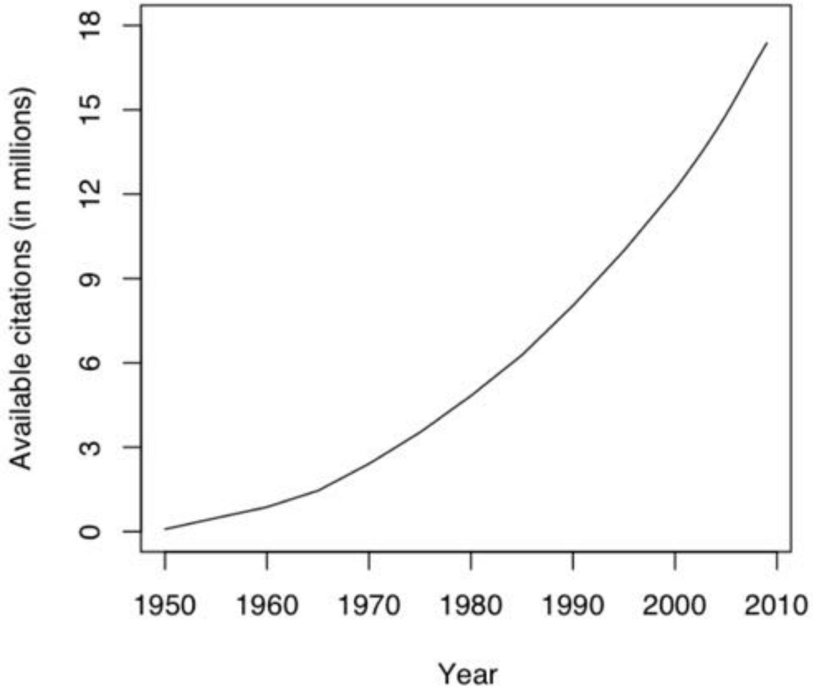
environmental science

What is biomedicine?

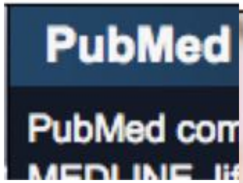
- A large number of related disciplines
- “Studying the structure, function, growth, origin, evolution or distribution of living organisms and their natural environments”

They like to publish

- MEDLINE:
  - A bibliographic database
  - Exponential growth
  - Manually indexed (MeSH)
  - 2013 statistics
    - 19 mln references
    - ± 5,600 journals
  - 2010: 700,000 additions

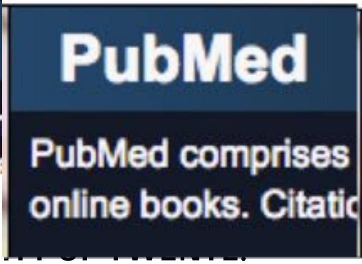


2016 update: 2017 update: 2018 update:



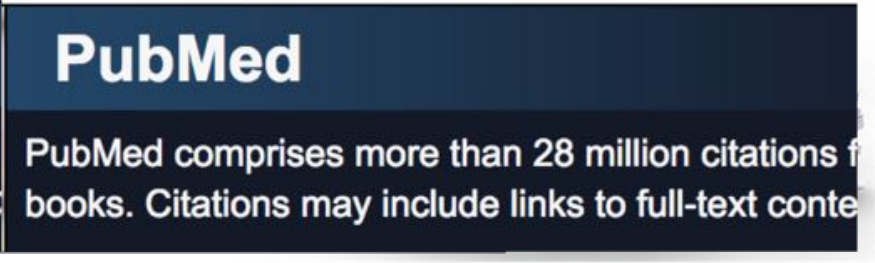
PubMed

PubMed com
MEDLINE, li



PubMed

PubMed comprises
online books. Citatio

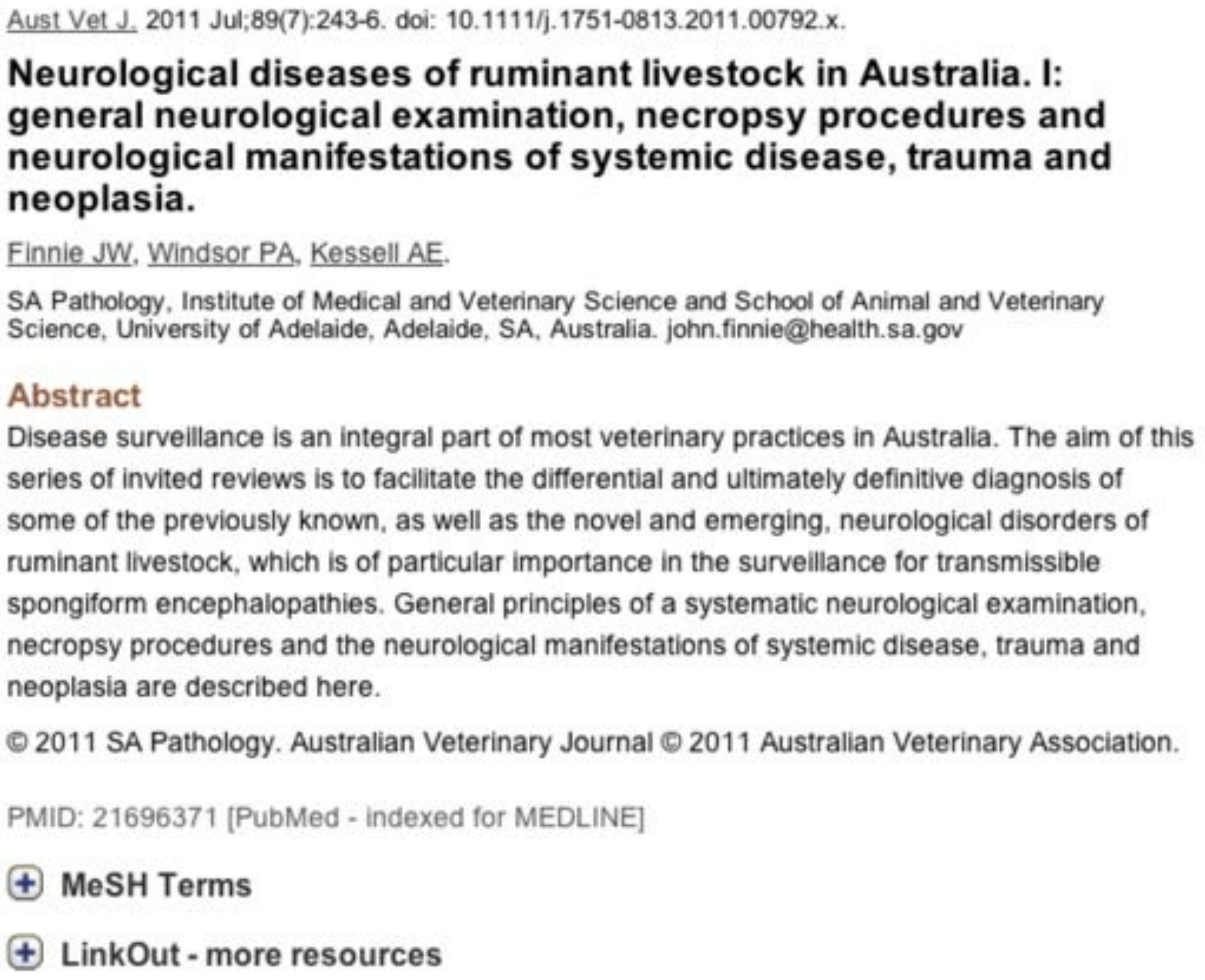


PubMed

PubMed comprises more than 28 million citations f
books. Citations may include links to full-text conte

A sample MEDLINE entry

- Authors & Affiliations
- Title
- Journal
- Publication date
- **Abstract**
- **MeSH terms**



Aust Vet J. 2011 Jul;89(7):243-6. doi: 10.1111/j.1751-0813.2011.00792.x.

Neurological diseases of ruminant livestock in Australia. I: general neurological examination, necropsy procedures and neurological manifestations of systemic disease, trauma and neoplasia.

Finnie JW, Windsor PA, Kessell AE.

SA Pathology, Institute of Medical and Veterinary Science and School of Animal and Veterinary Science, University of Adelaide, Adelaide, SA, Australia. john.finnie@health.sa.gov

Abstract

Disease surveillance is an integral part of most veterinary practices in Australia. The aim of this series of invited reviews is to facilitate the differential and ultimately definitive diagnosis of some of the previously known, as well as the novel and emerging, neurological disorders of ruminant livestock, which is of particular importance in the surveillance for transmissible spongiform encephalopathies. General principles of a systematic neurological examination, necropsy procedures and the neurological manifestations of systemic disease, trauma and neoplasia are described here.

© 2011 SA Pathology. Australian Veterinary Journal © 2011 Australian Veterinary Association.

PMID: 21696371 [PubMed - indexed for MEDLINE]

Quiz

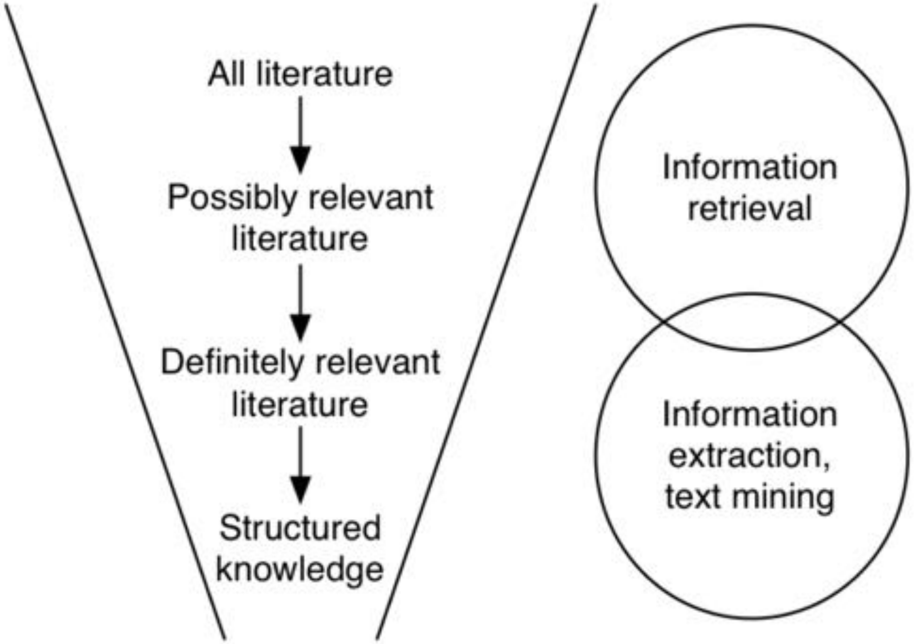
- What is more important in this domain? Precision or recall?

“A month in the laboratory can save an hour in the library”
F. Westheimer (1912-2007), professor of Chemistry at Harvard University

Information retrieval in the text mining landscape

- Information retrieval
  - Finding information
    - *Find information about P53*
- Information extraction
  - Extracting facts
    - *Which proteins interact with P53?*
- Knowledge discovery
  - Discovering new knowledge
  - E.g. combining complementary but disjoint literatures (Swanson)
    - *Fish oil ⇔ blood viscosity*

      blood viscosity ⇔ Raynaud's disease


All literature
Possibly relevant literature
Definitely relevant literature
Structured knowledge
Information retrieval
Information extraction, text mining

[Hersh, 2009]

Terminology: a challenge for biomedical IR

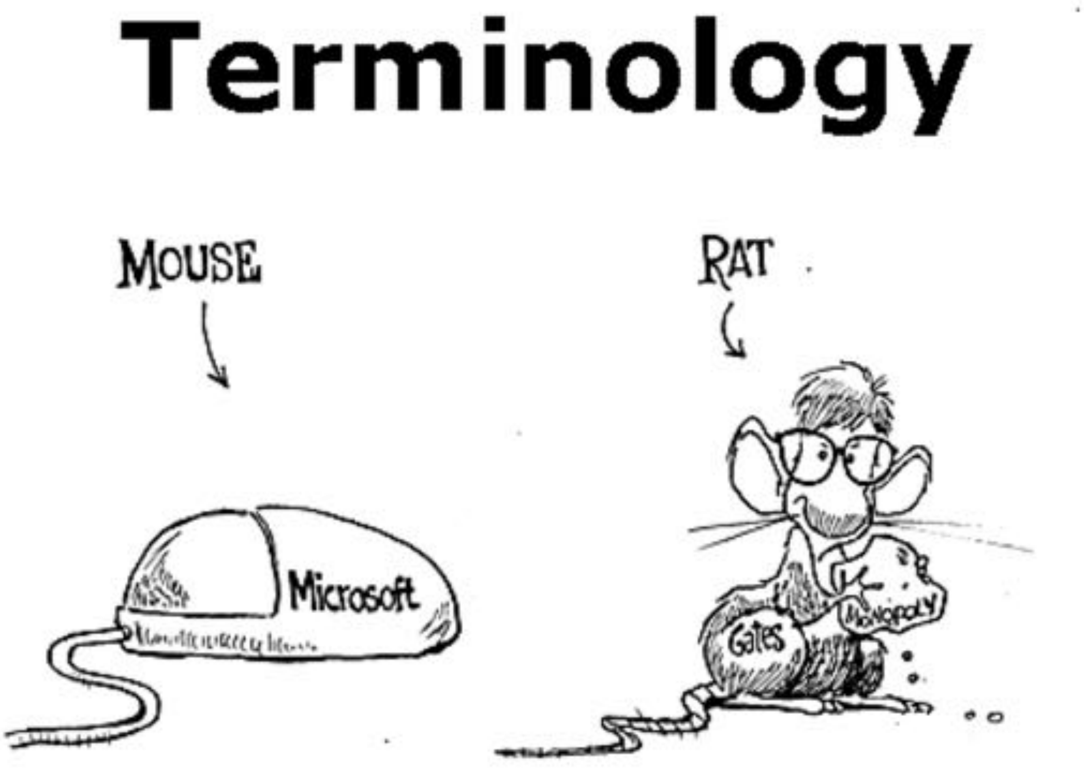
- Biomedical **concepts** are represented by **terms**
- What is a concept?
  - “an abstract idea, a general notion” ~ something interesting
- Examples of biomedical concepts
  - Diseases
  - Organisms
  - Genes
  - Proteins
  - Chemicals
  - …

Mad cow disease

“mad cow disease”
“BSE”
“Bovine spongiform encephalopathy”

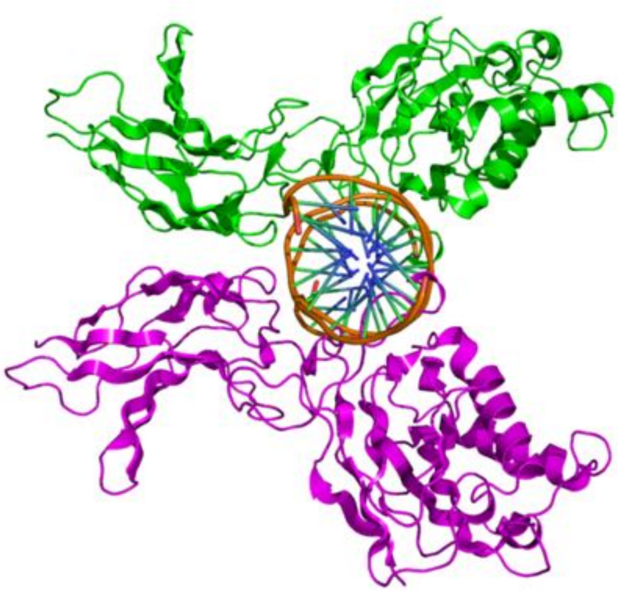
Characteristics of biomedical terminology

- Complex
- Inconsistent
- Many synonyms
- Ambiguous



Biomedical terminology is complex

- Many compound terms
  - *nuclear factor kappa-light-chain-enhancer of activated B cells*
- 85% of the terms consist of more than one word (Nenadic et al, 2005)
- Frequent use of ad hoc abbreviations
  - *TRADD binds to the TNF receptor-associated factor 2 (TRAF-2) that recruits NF-kB-inducible kinase (NIK).*



Biomedical terminology is inconsistent

- 75% of the authors do not use official gene symbol or full gene names (Chen et al., 2005)
- Frequent spelling variation:
  - *NF-kB, nfkb, NF kappa B*
  - *syt4, syt iv*
- Fast changing terminology:
  - How many synonyms of Mexican flu can you think of?

novel influenza A (H1N1), 2009 H1N1 flu, new influenza A virus, pandemic H1N1/09 virus, novel H1N1 virus, A/California/07/2009 (H1N1), H1N1 influenza, H1N1 Virus, Mexican Virus, swine influenza, SI, Pig Flu, Swine-Origin Influenza A H1N1 Virus, Influenza A Virus, H1N1 Subtype, ...

Biomedical terminology is inconsistent

“Biologists would rather share their toothbrush than a gene name”

Michael Ashburner, professor of biology at the University of Cambridge

Biomedical terminology contains many synonyms

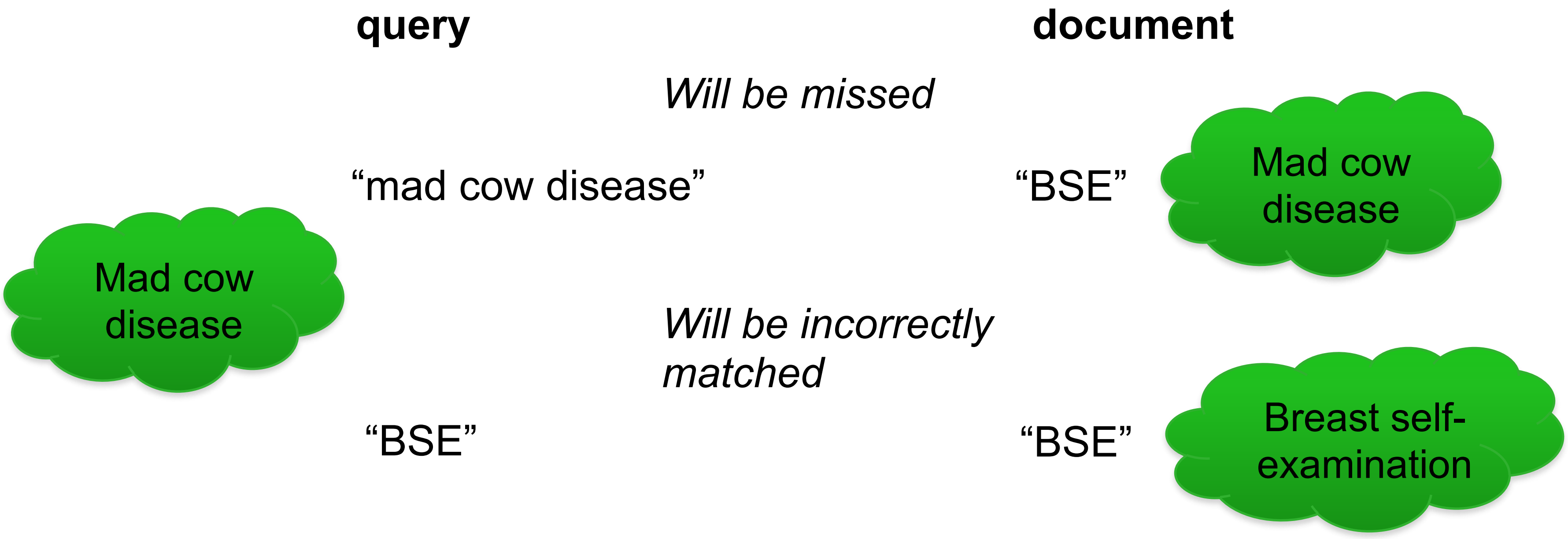
- Nuclear Factor-kappa B
  - Immunoglobulin Enhancer-Binding Protein
  - Ig-EBP-1, Ig EBP 1, IgEBP1
  - NF-kB, NFkappaB, NF-kappa-B, NF-kappa beta
  - Transcription Factor NF kB
  - NF kapa beta

Biomedical terminology is highly ambiguous

- Abbreviations: PSA
  - *prostate specific antigen*
  - *psoriasis arthritis*
  - *poultry science administration*
  - … (100 more)
- Use of general English terms
  - *white* protein
  - *big brain* protein
  - *hr*

Quiz

- What is the effect of these characteristics? (on precision/recall)
- How can an IR system deal with these characteristics?



query

document

Will be missed

“mad cow disease”

“BSE”

Mad cow disease

Mad cow disease

Will be incorrectly matched

“BSE”

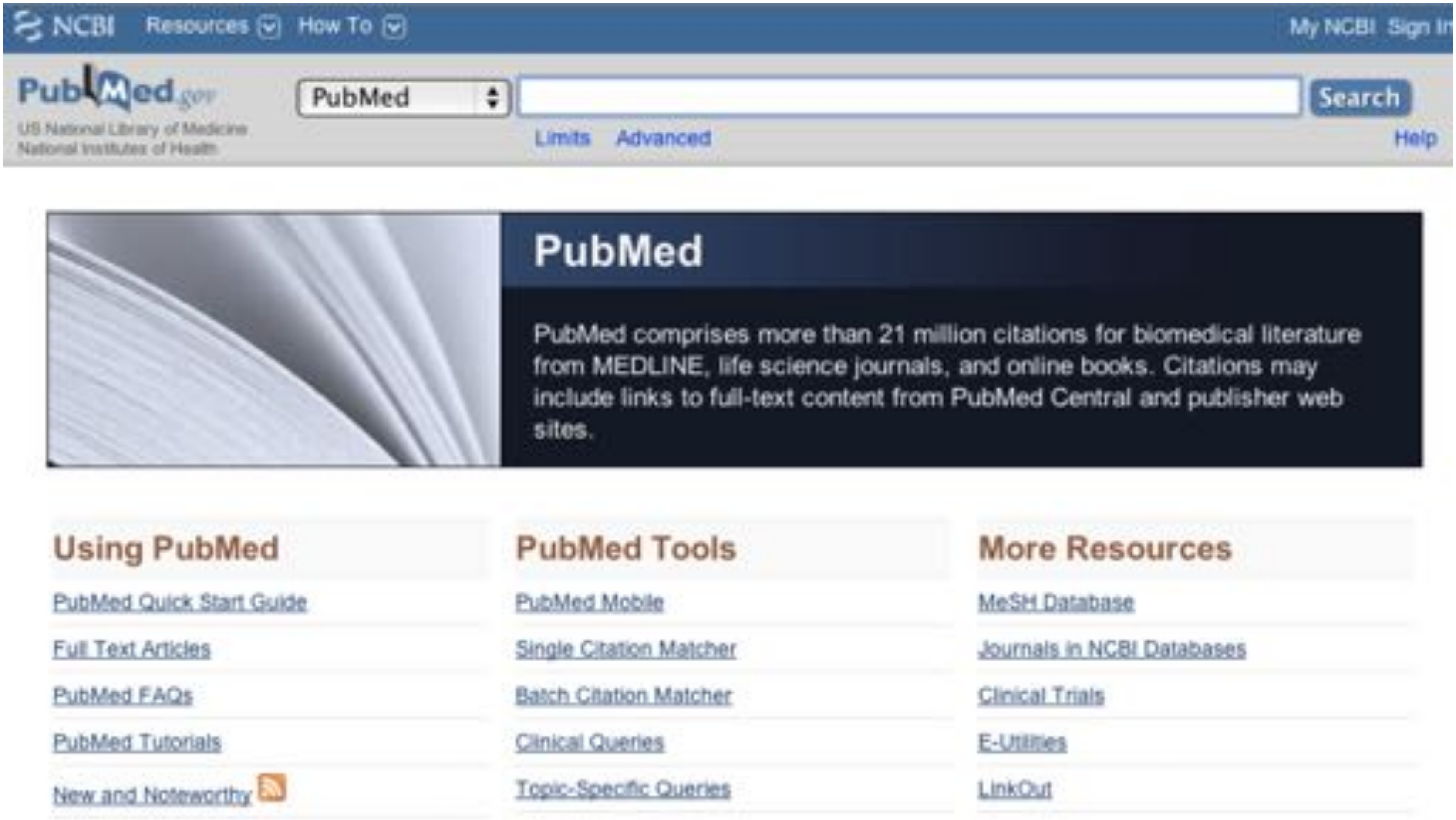
“BSE”

Breast self-examination

Quiz

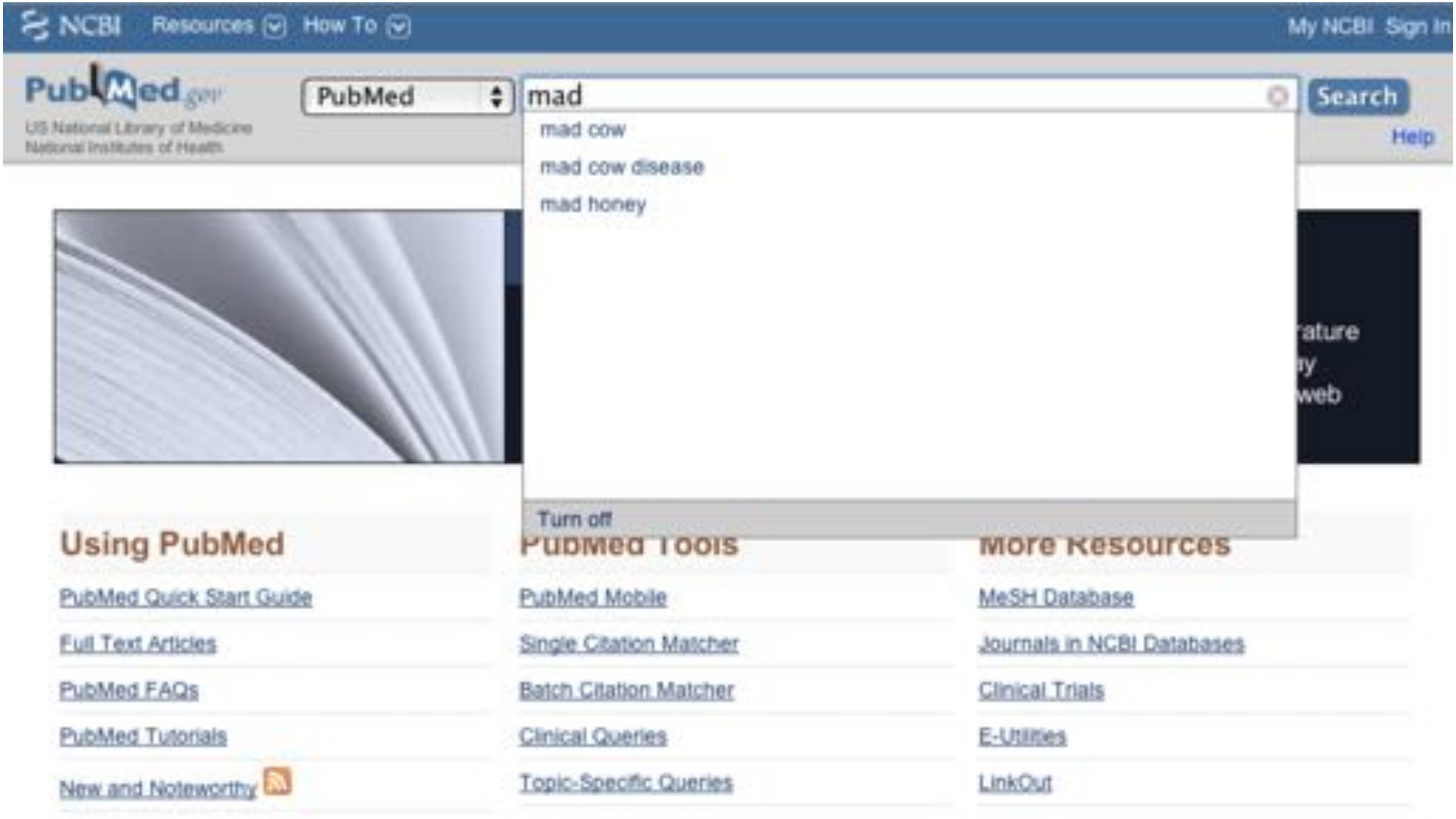
- What is the effect of these characteristics?
  - **Vocabulary mismatch** between query and (relevant) documents
  - Missing synonyms: low recall
  - Using ambiguous terms: low precision
- How can an IR system deal with these characteristics?
  - Incorporate **domain knowledge**, for instance
    - Sophisticated lexical analysis
    - Query/document expansion
    - Concept representations
    - ….

Biomedical Search in Practice: PubMed



PubMed comprises more than 21 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full-text content from PubMed Central and publisher web sites.

Let’s search PubMed



mad

mad cow

mad cow disease

mad honey

Using PubMed

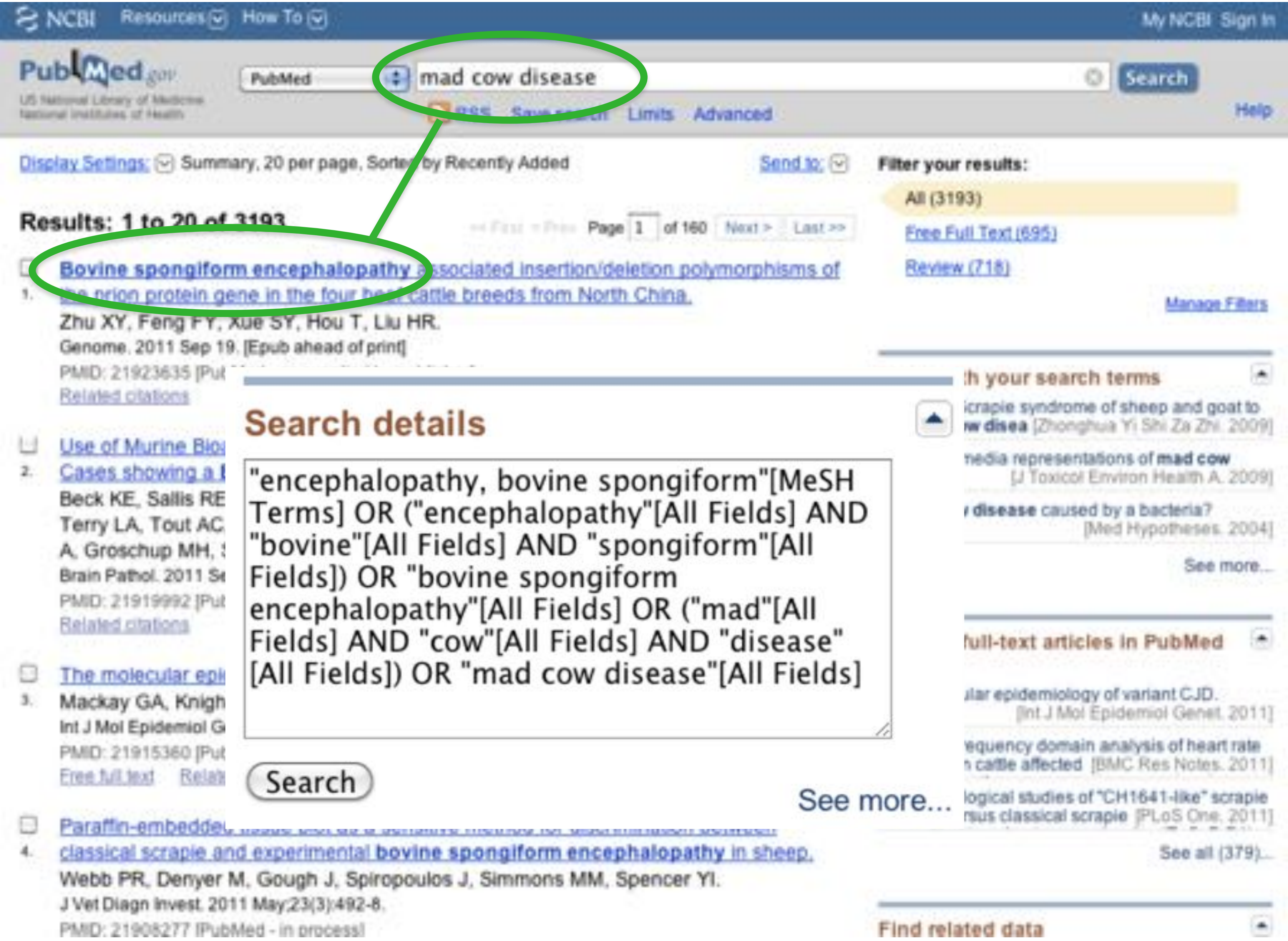
- PubMed Quick Start Guide
- Full Text Articles
- PubMed FAQs
- PubMed Tutorials
- New and Noteworthy

PubMed Tools

- PubMed Mobile
- Single Citation Matcher
- Batch Citation Matcher
- Clinical Queries
- Topic-Specific Queries

More Resources

- MeSH Database
- Journals in NCBI Databases
- Clinical Trials
- E-Utilities
- LinkOut



mad cow disease

Summary, 20 per page, Sorted by Recently Added

Results: 1 to 20 of 3193

1. Bovine spongiform encephalopathy associated insertion/deletion polymorphisms of the prion protein gene in the four beef cattle breeds from North China.
Zhu XY, Feng FY, Xue SY, Hou T, Liu HR.
Genome. 2011 Sep 19. [Epub ahead of print]
PMID: 21923635 [Pu

2. Use of Murine Bio
Cases showing a I
Beck KE, Sallis RE
Terry LA, Tout AC
A, Groschup MH,
Brain Pathol. 2011 Se
PMID: 21919992 [Pu

3. The molecular epi
Mackay GA, Knigh
Int J Mol Epidemiol G
PMID: 21915360 [Pu

4. Paraffin-embedded tissue blot as a sensitive method for discrimination between classical scrapie and experimental bovine spongiform encephalopathy in sheep.
Webb PR, Denyer M, Gough J, Spiropoulos J, Simmons MM, Spencer YI.
J Vet Diagn Invest. 2011 May;23(3):492-8.
PMID: 21908277 [PubMed - in process]

Search details

"encephalopathy, bovine spongiform"[MeSH Terms] OR ("encephalopathy"[All Fields] AND "bovine"[All Fields] AND "spongiform"[All Fields]) OR "bovine spongiform encephalopathy"[All Fields] OR ("mad"[All Fields] AND "cow"[All Fields] AND "disease"[All Fields]) OR "mad cow disease"[All Fields]

...h your search terms
...crapie syndrome of sheep and goat to ...w disea [Zhonghua Yi Shi Za Zhi. 2009]
...media representations of mad cow [J Toxicol Environ Health A. 2009]
...disease caused by a bacteria? [Med Hypotheses. 2004]

...full-text articles in PubMed
...ular epidemiology of variant CJD. [Int J Mol Epidemiol Genet. 2011]
...equency domain analysis of heart rate ...n cattle affected [BMC Res Notes. 2011]
...logical studies of "CH1641-like" scrapie ...rsus classical scrapie [PLoS One. 2011]

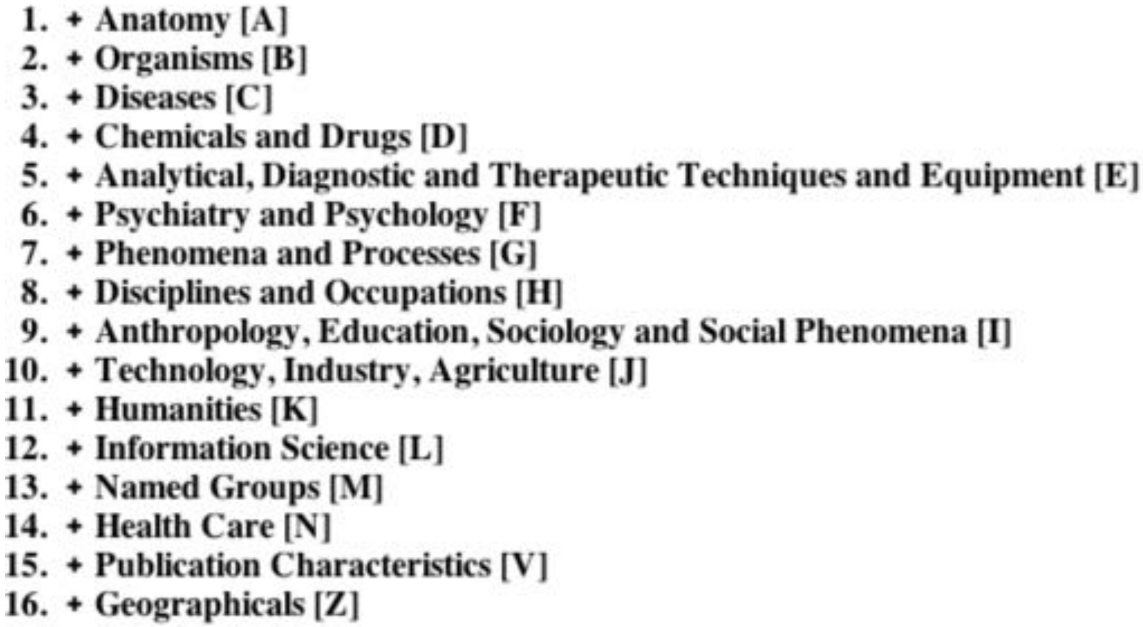
PubMed

- Searches the MEDLINE database
- Boolean matching
- It uses multiple indexing vocabularies:
  - Manual controlled vocabulary index (MeSH) &
  - Automatic uncontrolled vocabulary index (free text)
- By default, sorted by publication date (newest first)
- Automatic query mapping and expansion

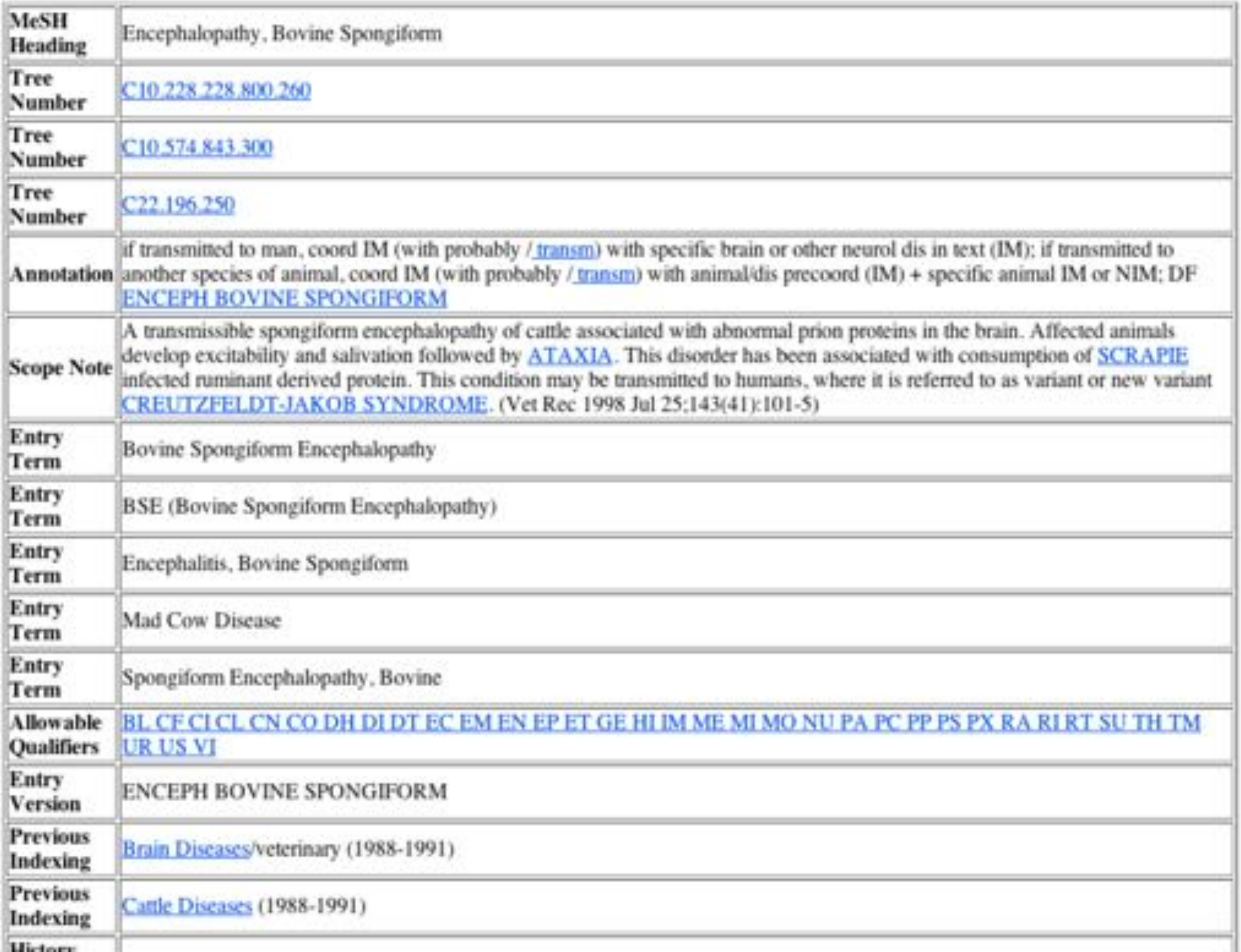
MeSH

MEDICAL SUBJECT HEADINGS

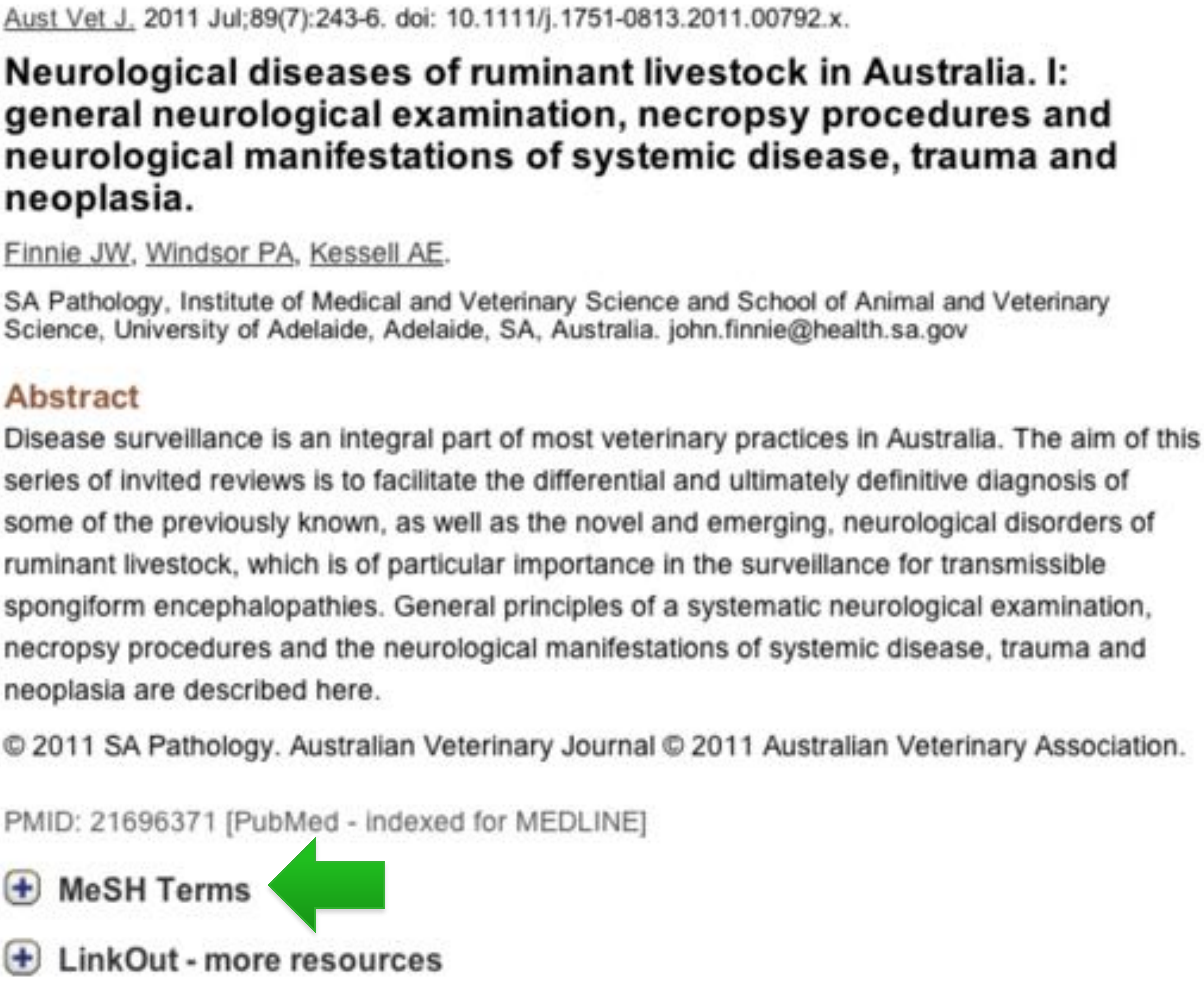
- A controlled vocabulary for indexing biomedical documents
- 24,000 main descriptors + qualifiers
- Hierarchically organized (DAG)



1. + Anatomy [A]
2. + Organisms [B]
3. + Diseases [C]
4. + Chemicals and Drugs [D]
5. + Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]
6. + Psychiatry and Psychology [F]
7. + Phenomena and Processes [G]
8. + Disciplines and Occupations [H]
9. + Anthropology, Education, Sociology and Social Phenomena [I]
10. + Technology, Industry, Agriculture [J]
11. + Humanities [K]
12. + Information Science [L]
13. + Named Groups [M]
14. + Health Care [N]
15. + Publication Characteristics [V]
16. + Geographicals [Z]



MeSH Heading	Encephalopathy, Bovine Spongiform
Tree Number	C10.228.228.800.260
Tree Number	C10.574.843.300
Tree Number	C22.196.250
Annotation	if transmitted to man, coord IM (with probably / transm) with specific brain or other neurol dis in text (IM); if transmitted to another species of animal, coord IM (with probably / transm) with animal/dis precoord (IM) + specific animal IM or NIM; DF ENCEPH BOVINE SPONGIFORM
Scope Note	A transmissible spongiform encephalopathy of cattle associated with abnormal prion proteins in the brain. Affected animals develop excitability and salivation followed by ATAXIA. This disorder has been associated with consumption of SCRAPIE infected ruminant derived protein. This condition may be transmitted to humans, where it is referred to as variant or new variant CREUTZFELDT-JAKOB SYNDROME. (Vet Rec 1998 Jul 25;143(41):101-5)
Entry Term	Bovine Spongiform Encephalopathy
Entry Term	BSE (Bovine Spongiform Encephalopathy)
Entry Term	Encephalitis, Bovine Spongiform
Entry Term	Mad Cow Disease
Entry Term	Spongiform Encephalopathy, Bovine
Allowable Qualifiers	BL CF CI CL CN CO DH DI DT EC EM EN EP ET GE HI IM ME MI MO NU PA PC PP PS PX RA RI RT SU TH TM UR US VI
Entry Version	ENCEPH BOVINE SPONGIFORM
Previous Indexing	Brain Diseases/veterinary (1988-1991)
Previous Indexing	Cattle Diseases (1988-1991)
History	



Aust Vet J. 2011 Jul;89(7):243-6. doi: 10.1111/j.1751-0813.2011.00792.x.

Neurological diseases of ruminant livestock in Australia. I: general neurological examination, necropsy procedures and neurological manifestations of systemic disease, trauma and neoplasia.

Finnie JW, Windsor PA, Kessell AE.

SA Pathology, Institute of Medical and Veterinary Science and School of Animal and Veterinary Science, University of Adelaide, Adelaide, SA, Australia. john.finnie@health.sa.gov

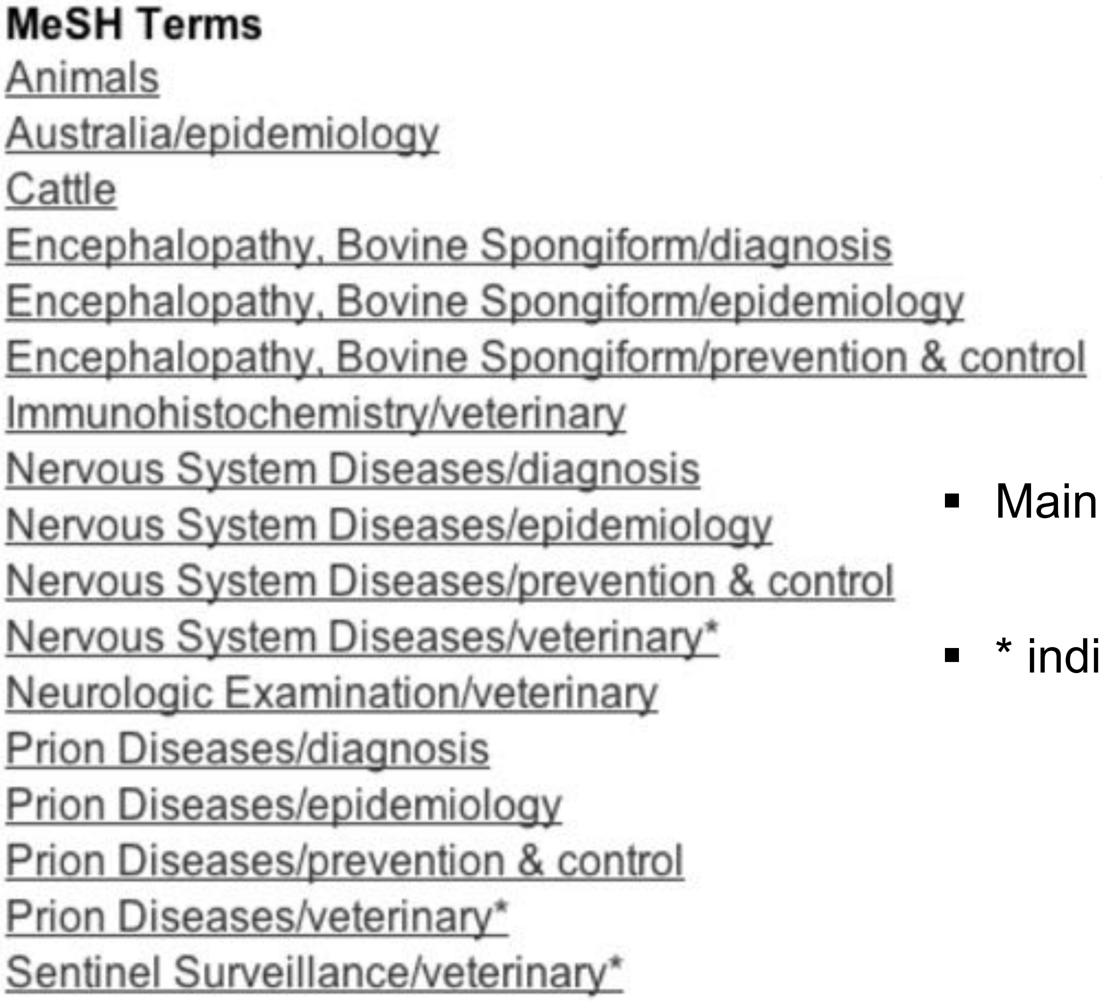
Abstract

Disease surveillance is an integral part of most veterinary practices in Australia. The aim of this series of invited reviews is to facilitate the differential and ultimately definitive diagnosis of some of the previously known, as well as the novel and emerging, neurological disorders of ruminant livestock, which is of particular importance in the surveillance for transmissible spongiform encephalopathies. General principles of a systematic neurological examination, necropsy procedures and the neurological manifestations of systemic disease, trauma and neoplasia are described here.

PMID: 21696371 [PubMed - indexed for MEDLINE]

⊞ MeSH Terms

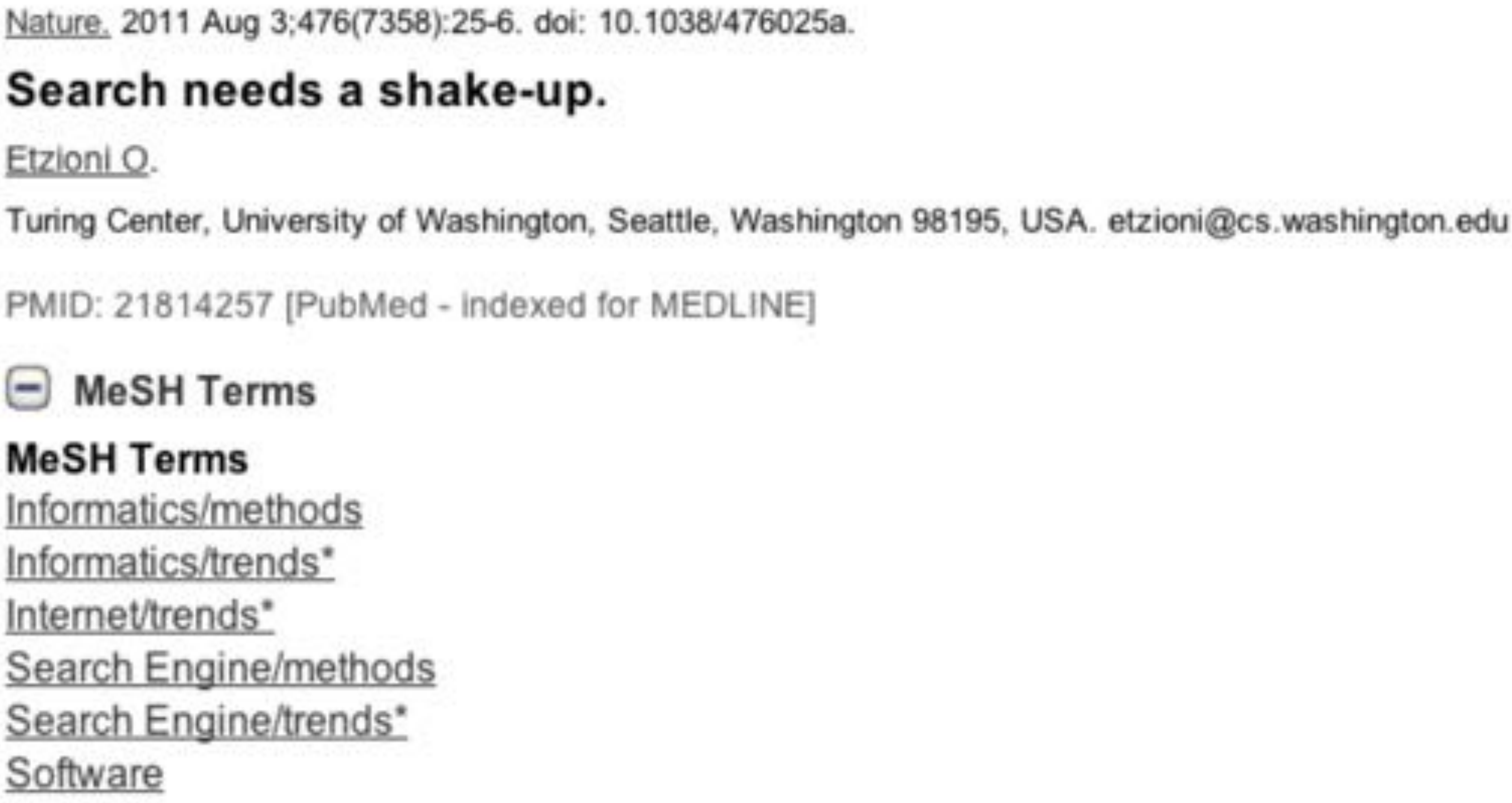
⊞ LinkOut - more resources



MeSH Terms

Animals
Australia/epidemiology
Cattle
Encephalopathy, Bovine Spongiform/diagnosis
Encephalopathy, Bovine Spongiform/epidemiology
Encephalopathy, Bovine Spongiform/prevention & control
Immunohistochemistry/veterinary
Nervous System Diseases/diagnosis
Nervous System Diseases/epidemiology
Nervous System Diseases/prevention & control
Nervous System Diseases/veterinary*
Neurologic Examination/veterinary
Prion Diseases/diagnosis
Prion Diseases/epidemiology
Prion Diseases/prevention & control
Prion Diseases/veterinary*
Sentinel Surveillance/veterinary*

- Main descriptor/qualifier
- * indicates important



Nature. 2011 Aug 3;476(7358):25-6. doi: 10.1038/476025a.

Search needs a shake-up.

Etzioni O.

Turing Center, University of Washington, Seattle, Washington 98195, USA. etzioni@cs.washington.edu

PMID: 21814257 [PubMed - indexed for MEDLINE]

⊟ **MeSH Terms**

MeSH Terms

Informatics/methods
Informatics/trends*
Internet/trends*
Search Engine/methods
Search Engine/trends*
Software

FACTS on MeSH

- Organizing principle: “to conceptually partition the literature”
- Hierarchy: Is-a and part-of relationships
- Yearly updated
- Average: 9 MeSH descriptors per document
- Manually assigned, also based on full-text

Quiz: different styles of indexing

	Manual Controlled vocabulary (MeSH)	Automatic Uncontrolled vocabulary (free text)
Advantages	?	?
Disadvantages	?	?

Aspects: costs, representation quality, consistency, maintainability, effectiveness for searching, user friendliness, exhaustiveness/specificity

Quiz: different styles of indexing

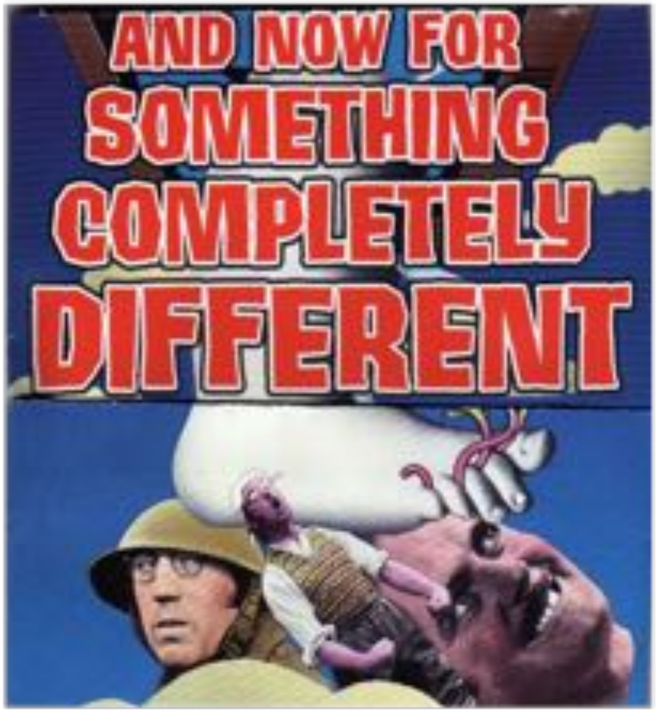
	Manual Controlled vocabulary (MeSH)	Automatic Uncontrolled vocabulary (free text)
Advantages	-Unambiguous -Terms are informative -High level summary	-Fast -Cheap -Trivial to maintain
Disadvantages	-Slow -Expensive -Hard to maintain -Difficult to keep consistent -Difficult to query	-Can be ambiguous -Not as intuitive

Automatic free text vs. manual contr. vocabulary indexing

- Automatic free text indexing is cheap, fast and trivial to maintain
- Controlled vocabulary indexing is easier to understand and unambiguous

➔ They might complement each other

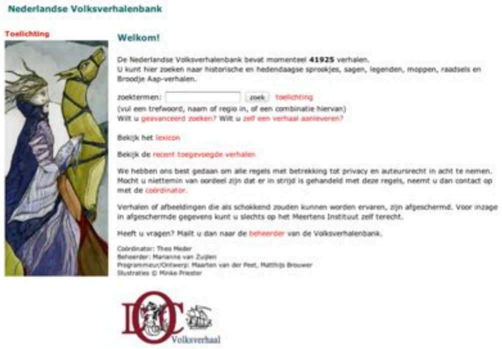
Case study 2: Folktales



The Dutch Folktale Database

- Maintained by the Meertens Institute since 1994
- > 40,000 Dutch folktales, collected since the 19th century
- Subgenres
  - Fairy tales, legends, urban legends
    jokes, riddles, personal narratives
- Languages
  - Dutch, Frisian, Old Dutch, Middle Dutch
    and many Dutch dialects
- Other metadata
  - Summary, **keywords**, story type, motifs
    proper names, storyteller, location etc.
- Online since 2004: www.verhalenbank.nl

instituut
Meertens



Nederlandse Volksverhalenbank

Toelichting

Welkom!

De Nederlandse Volksverhalenbank bevat momenteel **41925** verhalen.
U kunt hier zoeken naar historische en hedendaagse sprookjes, sagen, legenden, moppen, raadsels en Broodje Aap-verhalen.

zoektermen: zoek toelichting
(vul een trefwoord, naam of regio in, of een combinatie hiervan)
Wilt u geavanceerd zoeken? Wilt u zelf een verhaal aanleveren?

Bekijk het lexicon

Bekijk de recent toegevoegde verhalen

We hebben ons best gedaan om alle regels met betrekking tot privacy en auteursrecht in acht te nemen. Mocht u niettemin van oordeel zijn dat er in strijd is gehandeld met deze regels, neemt u dan contact op met de coördinator.

Verhalen of afbeeldingen die als schokkend zouden kunnen worden ervaren, zijn afgeschermd. Voor inzage in afgeschermde gegevens kunt u slechts op het Meertens Instituut zelf terecht.

Heeft u vragen? Mailt u dan naar de beheerder van de Volksverhalenbank.

Coördinator: Theo Meder
Beheerder: Marianne van Zuijlen
Programmeur/Ontwerp: Maarten van der Peet, Matthijs Brouwer
Illustraties © Minke Priester

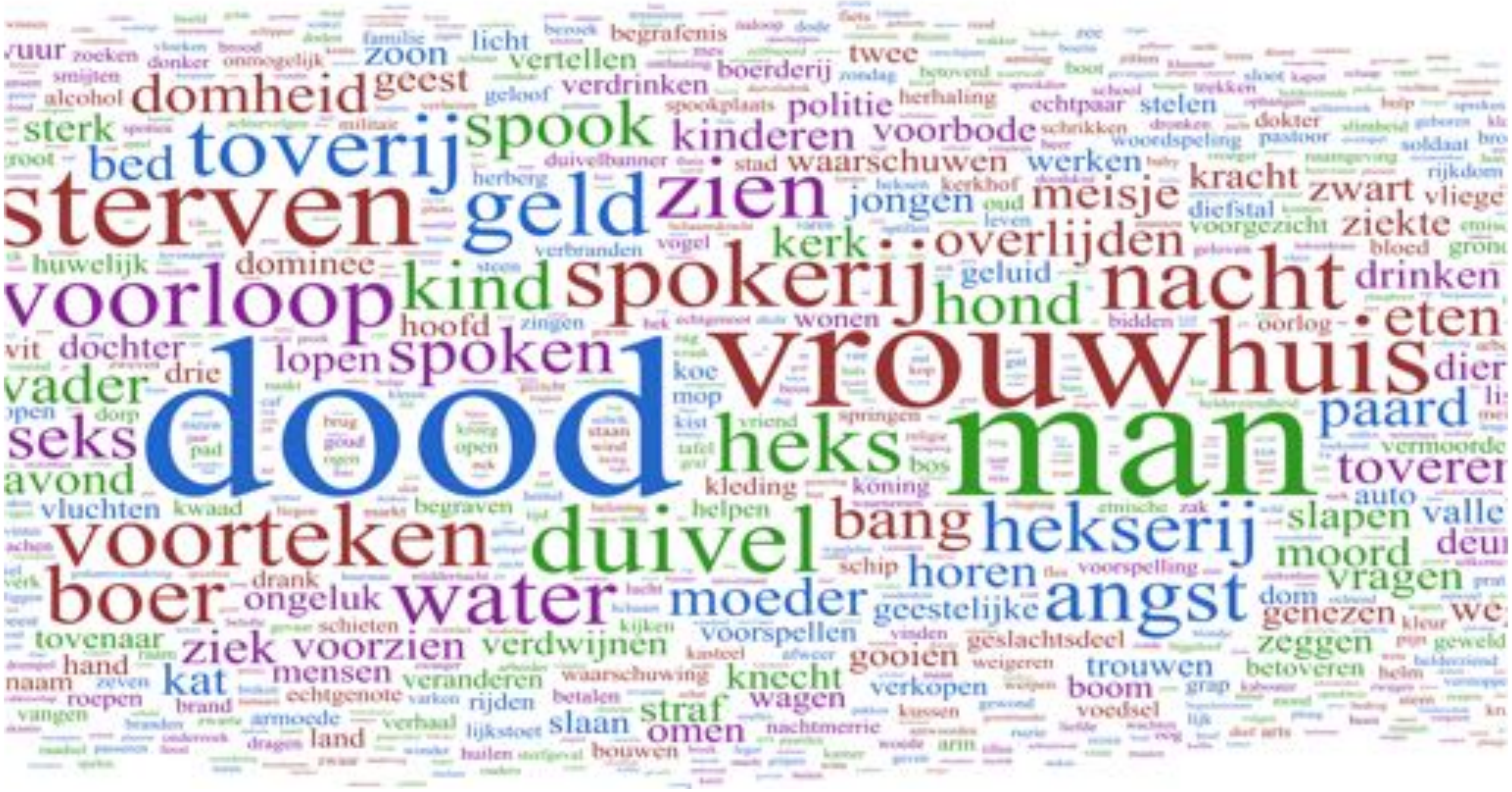
Quiz

- Why do manual indexing in this domain?
- Why use an uncontrolled vocabulary?

Quiz

- Why do manual indexing in this domain?
  - Multilingual content
  - Variety in style (temporal, audience)
  - Assign abstract terms
  - Make a selection of important topics
- Why use an uncontrolled vocabulary?
  - New topics appear frequently (urban legends)
  - Controlled vocabulary is labour-intensive

Manual keywords (1/2)



Manual keywords (2/2)

- Keyword assignment
  - Manual uncontrolled vocabulary indexing
  - Vaguely defined indexing task
  - Carried out by many different annotators
- Statistics (42k docs, 17k Dutch)
  - 15 assigned keywords on average, median 10
  - Mostly single words (90%)
  - 43k unique keywords
  - 65% of keywords appears literally in (Dutch) text

How do the keywords relate to the story text?

- Manual classification of 50 docs, 989 keywords

- Classes fraction
  - Literal 68%
  - Almost literal 12%
  - Synonym 5%
  - Hypernym 2%
  - Typing error <1%
  - Other (more abstract, etc.) 13%

- ➔ 80% can be (almost) literally linked to the text

Do annotators agree?

- Setup
  - 10 annotators, 5 stories each
  - Each story annotated by 2 annotators
  - Judge all story words:
    1) non-relevant; 2) relevant; 3) highly relevant
- Results of measuring inter-annotator agreement
  - Substantial agreement on relevant keywords (κ: 0.62),
    only moderate agreement on highly relevant keywords (κ: 0.48)
  - Reasons for disagreement

    1) verbs and adjectives? 2) overlooked

    3) choice rather than both 4) lack of instructions

Consistency is an issue with manual indexing

Summary

- Different styles of indexing and indexing languages
  - Each with its pros and cons
- Depends on the domain, important factors include
  - Type of information
  - Cost
  - Speed
  - Maintainability
  - Consistency
  - User friendliness