FlexiTerm: Flexible multi–word term recognition

Prof. Irena Spasić
i.spasic@cs.cardiff.ac.uk

Cardiff School of **Computer Science & Informatics**

http://www.cs.cf.ac.uk

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Outline

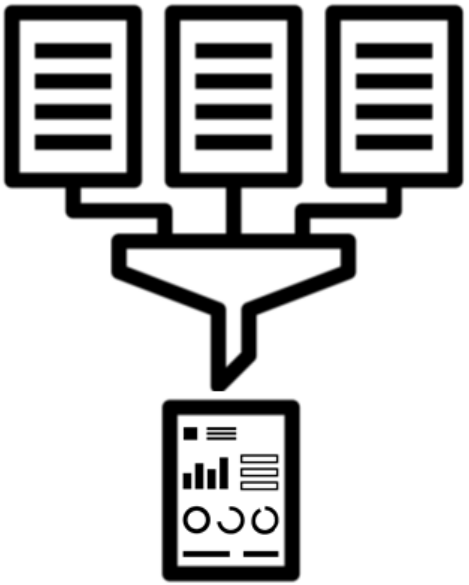
- text analysis in social & life sciences
- multi–word terms
  - termhood
  - unithood
  - variation
- automatic term recognition
  - linguistic approaches
  - statistical approaches
- acronyms as multi–word terms

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Introduction

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Text analysis

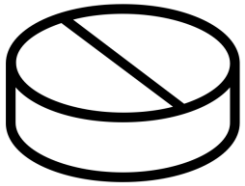


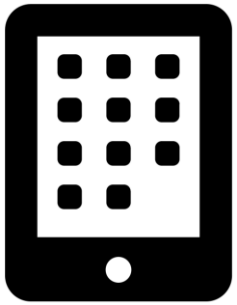
- examples
  - systematic reviews
  - content analysis
  - corpus linguistics
- data driven rather than hypothesis driven
- software support
  - e.g. covidence, NVivo, AntConc
- still a lot of manual labour… reading
- speed reading: skimming & scanning

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Terms

- What are **terms**?
  - means of conveying scientific & technical information
  - linguistic representations of domain-specific concepts
- e.g. tablet

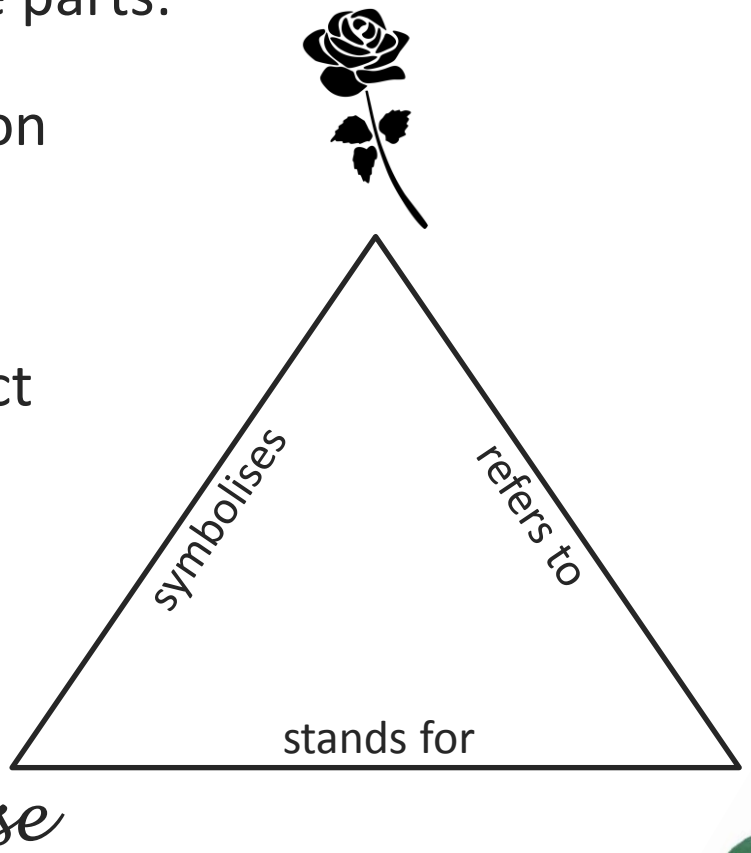




The meaning triangle

- a simple model of semantics
- a sign is broken into three parts:

1. **symbol** representation
2. **concept** abstract idea
3. **referent** specific object



O Romeo, Romeo, wherefore art thou Romeo?
Deny thy father and refuse thy name,
Or, if thou wilt not, be but sworn my love,
And I'll no longer be a Capulet.
'Tis but thy name that is my enemy;
Thou art thyself, though not a Montague.
What's Montague? it is nor hand, nor foot,
Nor arm, nor face, nor any other part
Belonging to a man. O, be some other name!
What's in a name? that which we call a **rose**
By any other name would smell as sweet;
So Romeo would, were he not Romeo call'd,
Retain that dear perfection which he owes
Without that title. Romeo, doff thy name,
And for that name which is no part of thee
Take all myself.

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Multi–word terms

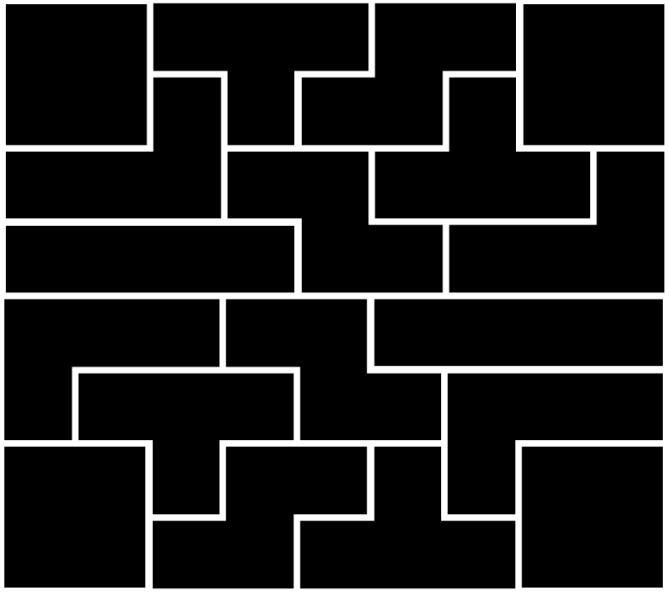
- computer science — recurrent neural network (RNN)
- mathematics — dot product
- biology — stem cell
- chemistry — fatty acid
- medicine — chronic obstructive pulmonary disease (COPD)
- law — reasonable doubt
- economics — quasi-autonomous non-government organisation (QUANGO)
- intelligence — weapon of mass distraction (WMD)

Collocation

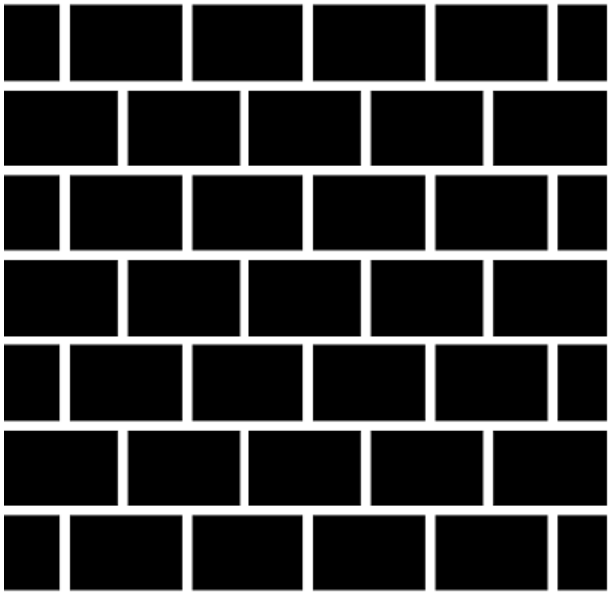
- combination of words that co-occur more often than would be expected by chance

typical collocation	incorrect collocation
strong tea	powerful tea
discharged from hospital	released from hospital
released from prison	discharged from prison
high temperature	tall temperature
piece of cake	part of cake
take the biscuit	have the cookie
dot product	period product
scalar product	N/A
scalar multiplication	N/A

Text representation



- multi-word expressions
- logical segmentation
- latent features



- bag of words or n-grams
- physical segmentation
- surface features

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Problems

- potentially unlimited number of domains
- dynamic nature of some domains
  - computer science: generative adversarial network
  - medicine: swine flu
  - dictionaries are **not** always **up to date**
- user–generated content such as blogs, where lay users use non–standard terminology
  - medicine: full knee replacement ✗
    total knee replacement (TKR) ✓
  - dictionaries are **not** always **suitable**

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Alternatives

- automatic term recognition (ATR)
- recognising terms in text without a dictionary
- potentially distinctive properties
  - syntactic structure
  - frequency distribution
- approaches
  - tagging/parsing + pattern matching
  - counting

Linguistic filtering (Justeson & Katz, 1995)

- preferred phrase structures
- terms are mostly noun phrases containing adjectives, nouns, possessives and prepositions
- **(A | N)$^{+}$ N**
  - e.g. mean/N squared/A error/N
- **(N | A)* N S (N | A)* N**
  - e.g. Zipf/N 's/S law/N
- **(N | A)* N P (N | A)* N**
  - e.g. law/N of/P large/A numbers/N

Cost criteria (Kita et al, 1994)

- collocations are recurrent word sequences
- recurrence is captured by the absolute frequency
- a simple absolute frequency approach does **not** work!
- frequency(sub-sequence) > frequency(sequence)
- e.g. f('in spite') ≥ f('in spite of')
- cost: $K(\alpha) = (|\alpha| - 1) \cdot (f(\alpha) - f(\beta))$
  - α, β ... word sequences, $\beta = u\alpha v$
  - $|\alpha|$ ... length (number of words in α)
  - $f(\alpha)$ ... frequency of α

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Multi–word term recognition

- hybrid solution
  - linguistic filters are used to extract candidate terms
  - ... which are then ranked using cost–like criteria
- C-value (Frantzi & Ananiadou, 1999; Nenadić, Spasić & Ananiadou, 2002)

$$C-value(t)=\begin{cases}\ln|t|\cdot f(t) & , \text{ if } S(t)=\varnothing \\ \ln|t|\cdot\left(f(t)-\frac{1}{|S(t)|}\sum_{s\in S(t)}f(s)\right) & , \text{ if } S(t)\neq\varnothing\end{cases}$$

- e.g. anterior cruciate ligament, posterior cruciate ligament
- the method favours longer, more frequently and independently occurring term candidates

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Term variation

- C–value works well when terms are used consistently, i.e. when they do not vary in structure and content
- however, terms may vary:
  - **orthographic variation**, e.g. posterolateral corner vs. postero–lateral corner vs. postero lateral corner
  - **morphological variation**
    inflection, e.g. lateral meniscus vs. lateral menisci
    derivation, e.g. meniscus tear vs. meniscal tear
  - **syntactic variation**, e.g.
    stone in kidney vs. kidney stone

Term variation

- ≈1/3 of an English scientific corpus accounts for term variants
  - ≈ 59% are semantic variants
  - ≈17% are morphological variants
  - ≈24% are syntactic variants
- frequency–based term recognition methods need to include term normalisation to:
  - associate term variants with one another
  - aggregate their frequencies at the semantic level
  - ... instead of dispersing them across separate variants at the lexical level!

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

FlexiTerm:
Flexible term recognition

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Method overview

- FlexiTerm is an open-source, stand-alone application for automatic term recognition
- similarly to C–value, FlexiTerm performs term recognition in two stages:

1. lexico–syntactic filters are used to select term candidates
2. term candidates are scored using a formula that estimates their collocational stability

- major difference: the flexibility with which term candidates are compared in order to neutralise syntactic, morphological & orthographic variation

Normalisation

- in order to neutralise variation, all term candidates are normalised

1. treat each term candidate as a bag of words
2. remove punctuation (e.g. ' in possessives), numbers and stop words including prepositions (e.g. of)
3. remove any lowercase tokens with ≤2 characters (e.g. *Baker's cyst* vs. *vitamin D*)
4. stem each remaining token

   hypoxia at rest → {*hypoxia, rest*} ← *resting hypoxia*

5. add similar tokens to the bag of words (cont.)

Token similarity

- many types of morphological variation are effectively neutralised with stemming
  - e.g. *transplant* & *transplant~~ation~~* will be reduced to the same stem
- exact string matching will not link orthographic variants
  - e.g. *haemorrhag~~e~~* & *hemorrhag~~e~~* are stemmed to *haemorrhag* & *hemorrhag* respectively
- easily identified using lexical similarity (edit distance)
- phonetic similarity is also important in dealing with new phenomena such as SMS language, e.g. *l8 ~ late*

Syntactic variation

- termhood formula:

$$C-value(t)=\begin{cases}\ln|t|\cdot f(t) & ,\ \text{if } S(t)=\varnothing \\ \ln|t|\cdot\left(f(t)-\frac{1}{|S(t)|}\sum_{s\in S(t)}f(s)\right) & ,\ \text{if } S(t)\neq\varnothing\end{cases}$$

- term candidate:

Method	Representation	Nestedness
C–value	string	substring
FlexiTerm	bag of words	subset

order does **not** matter! → solves the problem of syntactic variation!

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Data

Data set	Topic	Document type	Source
1	molecular biology	abstract	PubMed
2	COPD	abstract	PubMed
3	COPD	blog post	open Web
4	obesity, diabetes	discharge summary	i2b2
5	knee MRI scan	imaging report	NHS

Evaluation

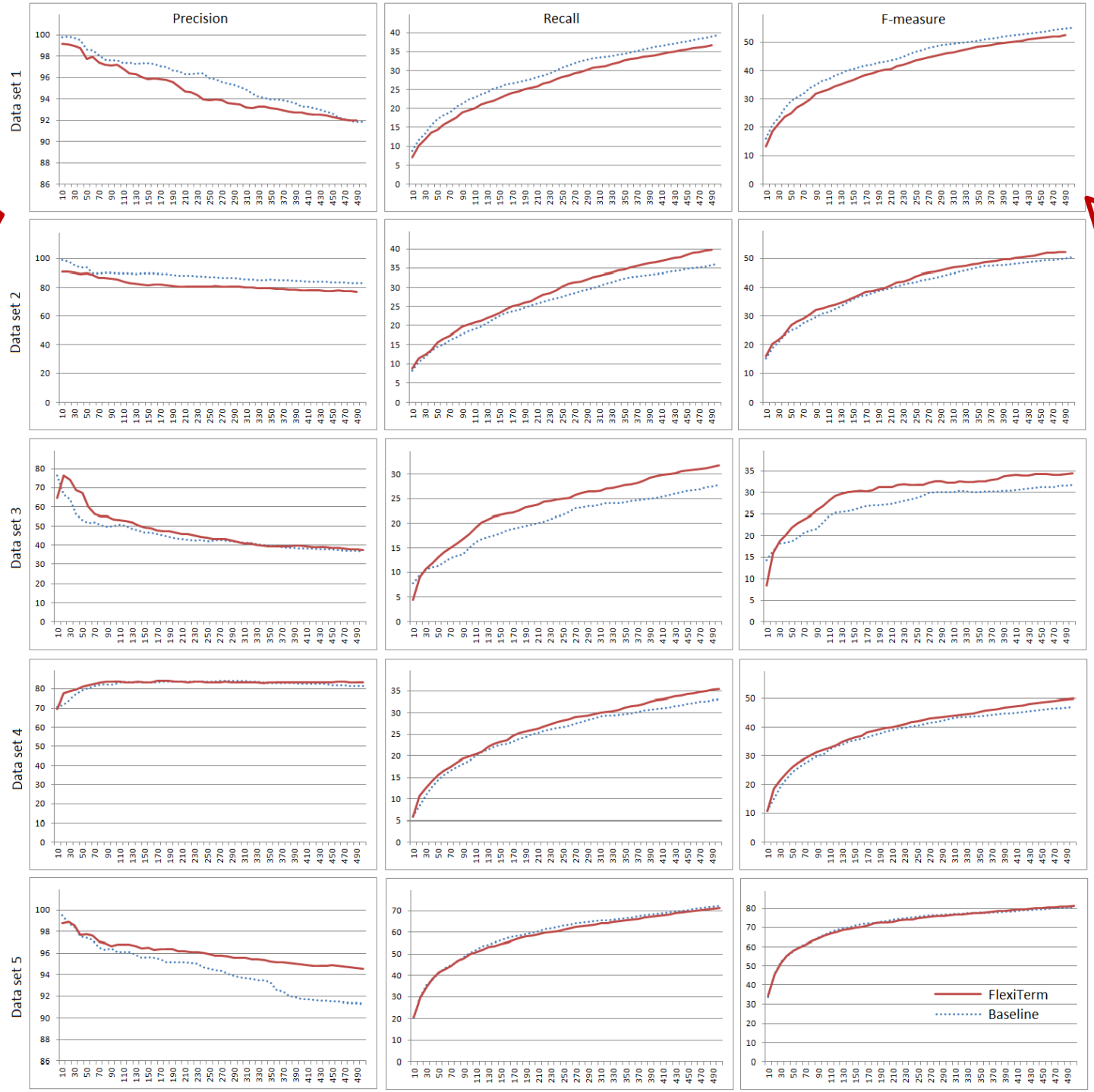
- What counts as a correctly recognised term?!?
- e.g. protein kinase C activation pathway
  - protein C0033684
  - protein kinase C0033640
  - protein kinase C C1259877
  - activation C1879547
  - pathway C1705987
  - protein activation pathway C1514528
  - protein kinase C activation pathway C1514554

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Evaluation

- token-level evaluation
- each token recognised or annotated as part of a term is classified as a true/false positive or false negative
- overlap between automatically recognised terms and manually annotated ones
- precision $P = TP / (TP + FP)$
- recall $R = TP / (TP + FN)$
- F-measure $F = 2PR / (P + R)$

C-value uses GENIA tagger



C-value does not include complex NPs

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Data set 1

Rank	FlexiTerm	TerMine
1	transcription factor transcription factors transcriptional factors	t cell
2	nf-kappa b	transcription factor
3	gene expression expression of genes	nf-kappa b
4	transcriptional activity activator of transcription transcriptional activation activating transcription activators of transcription transcription activation transcriptional activator	gene expression
5	nf-kappab activation nf-kappab activity	cell line
6	human t cells human cells	t lymphocyte
7	cell lines cell line	human monocyte
8	human monocytes	dna binding
9	activation of nf-kappa b nf-kappa b activation nf-kappa b activity	tyrosine phosphorylation
10	protein kinase	b cell

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Data set 2

Rank	FlexiTerm	TerMine
1	chronic obstructive pulmonary disease	chronic obstructive pulmonary disease
2	patients with copd copd patients	obstructive pulmonary disease
3	pulmonary disease	pulmonary disease
4	acute exacerbation acute exacerbations	copd patient
5	copd exacerbation copd exacerbations exacerbations of copd exacerbation of copd	acute exacerbation
6	patients with chronic obstructive pulmonary disease patients with chronic obstructive pulmonary diseases	severe copd
7	lung function	copd exacerbation
8	exacerbations of chronic obstructive pulmonary disease chronic obstructive pulmonary disease exacerbations exacerbation of chronic obstructive pulmonary disease	lung function
9	quality of life	airway inflammation
10	airway inflammation	exercise capacity

Data set 3

Rank	FlexiTerm	TerMine
1	pulmonary rehab pulmanory rehab	pulmonary rehab
2	breathe easy	breathe easy
3	vitamin d	vitamin d
4	lung transplantation lung transplant lung transplants lung transplantations	lung function
5	breathe easy groups breath easy groups breathe easy group	severe copd
6	chest infection chest infections	blood pressure
7	quality of life	lung disease
8	blood pressure	lung transplant
9	lung function	chest infection
10	rehab room	rehab room

Data set 4

Rank	FlexiTerm	TerMine
1	hospital course course of hospitalization	hospital course
2	chest pain	present illness
3	shortness of breath	chest pain
4	coronary artery coronary arteries	coronary artery
5	present illness	blood pressure
6	blood pressure blood pressures	ejection fraction
7	coronary artery disease	coronary artery disease
8	congestive heart failure	myocardial infarction
9	myocardial infarction	congestive heart failure
10	ejection fraction	cardiac catheterization

Data set 5

Rank	FlexiTerm	TerMine
1	mri knee	collateral ligament
2	collateral ligaments	medial meniscus
3	medial meniscus medial mensicus	lateral meniscus
4	lateral meniscus	hyaline cartilage
5	hyaline cartilage	posterior horn
6	posterior horn	femoral condyle
7	joint effusion	joint effusion
8	mri rt knee mri knee rt	mri lt knee
9	mri lt knee mri knee lt	lateral femoral condyle
10	lateral femoral condyle	medial femoral condyle
11	postero-lateral corner posterolateral corner	18 55!
14	infrapatellar fat pad infra-patella fat pad infra-patellar fat pad	20 281! 281!

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

FlexiTerm 2.0:
Acronyms as multi–word terms

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Acronyms

- another type of variation associated with multi–word terms
- multiple words are blended into a single token by taking the initial letters of:
  - words, e.g. chronic obstructive pulmonary disease (COPD)
  - morphemes, e.g. inhaled corticosteroids (ICS)
- the number of acronyms in PubMed is increasing by 11K per annum
- handy proxies for multi–word terms, so should be treated as multi–word terms themselves

Issues

- acronyms are a highly productive type of term variation
- e.g.
  - chronic obstructive pulmonary disease
  - COPD
  - COPD patients
  - patients with chronic obstructive pulmonary disease
- termhood formula:

$$C-value(t)=\begin{cases}\ln|t|\cdot f(t) & \text{, if } S(t)=\varnothing \\ \ln|t|\cdot\left(f(t)-\frac{1}{|S(t)|}\sum_{s\in S(t)}f(s)\right) & \text{, if } S(t)\neq\varnothing\end{cases}$$

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Solution

- mapping acronyms to their full forms would resolve these issues
- prerequisite: an acronym recognition method to extract acronym–definition pairs from a corpus
- cannot be done by post–processing FlexiTerm results
- acronym recognition needs to be fully integrated into the multi–word term recognition process
  - after the selection of multi–word term candidates
  - before termhood calculation

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Two types of acronyms

1. **explicit** (or **local**) acronyms
   - defined in a text document following scientific writing conventions
   - e.g. scientific papers
   - ... chronic obstructive pulmonary disease (COPD) ...
2. **implicit** (or **global**) acronyms
   - appear in a text document without their definitions
   - e.g. clinical narratives
   - … ACL … anterior cruciate ligament … ACL …

Explicit acronyms

- the prevalence of acronyms in biomedicine gave rise to proliferation of acronym recognition methods
- focus on extracting acronyms from the literature
- rely on scientific writing conventions
  - acronym should be defined the first time it is used
  - the full form followed by the acronym, written in uppercase, within parentheses
- pattern matching used to identify potential acronym–definition pairs followed by heuristic alignment of the two
- we re-used one such method (Schwartz & Hearst, 2003)

Implicit acronyms

- not explicitly defined in a document
- commonly found in clinical narratives as widely accepted synonyms of the corresponding terms, e.g.
  - STD vs. sexually transmitted disease
- such acronyms are known globally and, hence, are described in relevant dictionaries
- few methods focus on implicit acronym recognition in clinical narratives incorporate such dictionaries
- not appropriate for FlexiTerm as a data–driven, domain–independent method

Implicit acronyms

- a simple heuristic approach favours precision over recall

1. identify potential acronyms using their orthographic properties and frequency of occurrence
   - must start with an uppercase letter
   - must not contain a lowercase letter
   - must not end with a period
   - at least three characters long
   - frequency of occurrence above a threshold
2. compare acronyms against term candidates

- in the future, we will explore distributional semantics

FlexiTerm 2.0

1. extract term candidates using lexico–syntactic filters
2. **process acronyms**
   a. extract acronyms and their full forms (term candidates from step 1)
   b. add acronyms to the list of term candidates
   c. expand all acronym mentions to full forms
3. normalise term candidates as before
4. score term candidates using the C–value formula

$$C-value(t)=\begin{cases}\ln|t|\cdot f(t) & \text{, if } S(t)=\varnothing \\ \ln|t|\cdot\left(f(t)-\frac{1}{|S(t)|}\sum_{s\in S(t)}f(s)\right) & \text{, if } S(t)\neq\varnothing\end{cases}$$

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Performance improvement

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

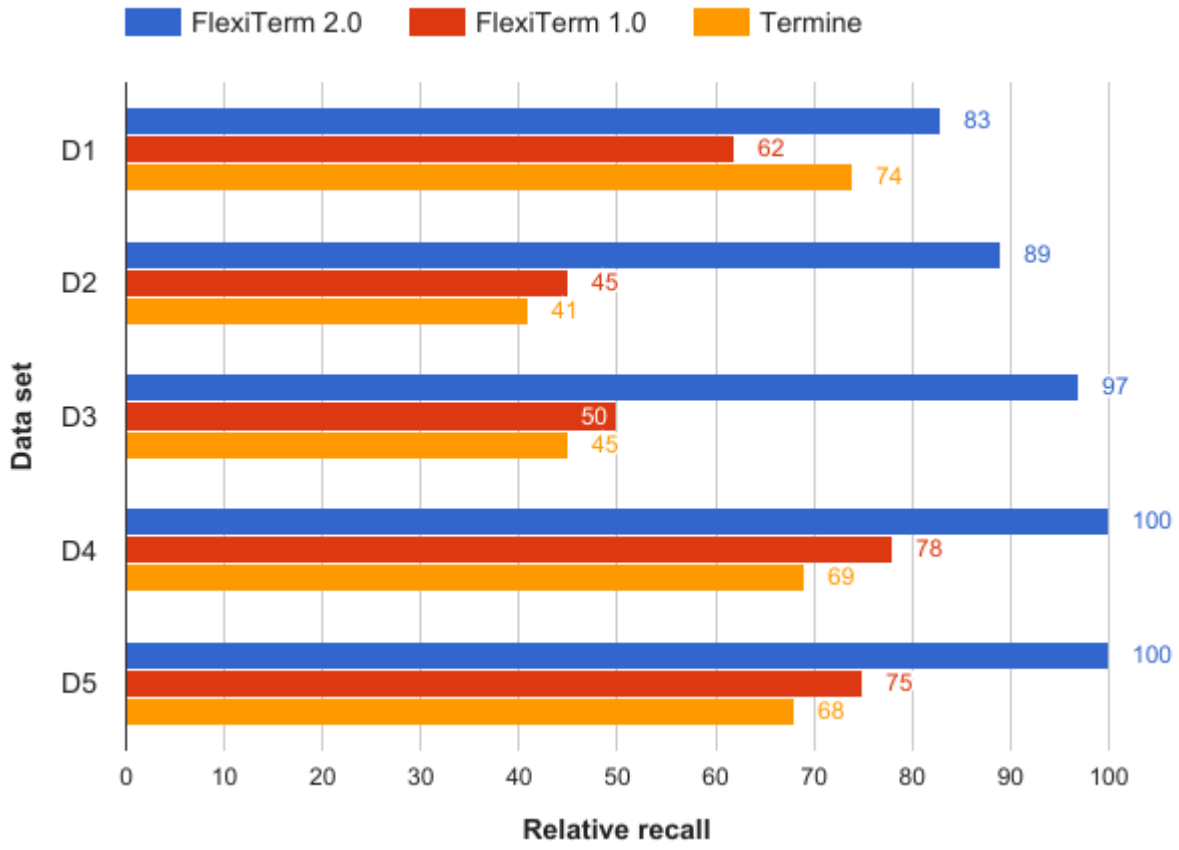
Application context

- by addressing acronyms in addition to morphological, orthographic and syntactic variation, we wanted to improve term conflation
  - grouping all variants of the same term
- one of the most prominent applications of term conflation is information retrieval
  - a process of selecting documents relevant to a user's information need expressed using a search query
- term conflation can support query expansion
  - adding synonyms and other closely related words to the search query

Evaluation measures

- precision & recall
- calculating recall requires manually annotating the whole document collection
- impractical in many cases
- relative recall compares multiple systems by only considering relevant documents retrieved by any given system
- only the retrieved documents need to be manually inspected

Relative recall



CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

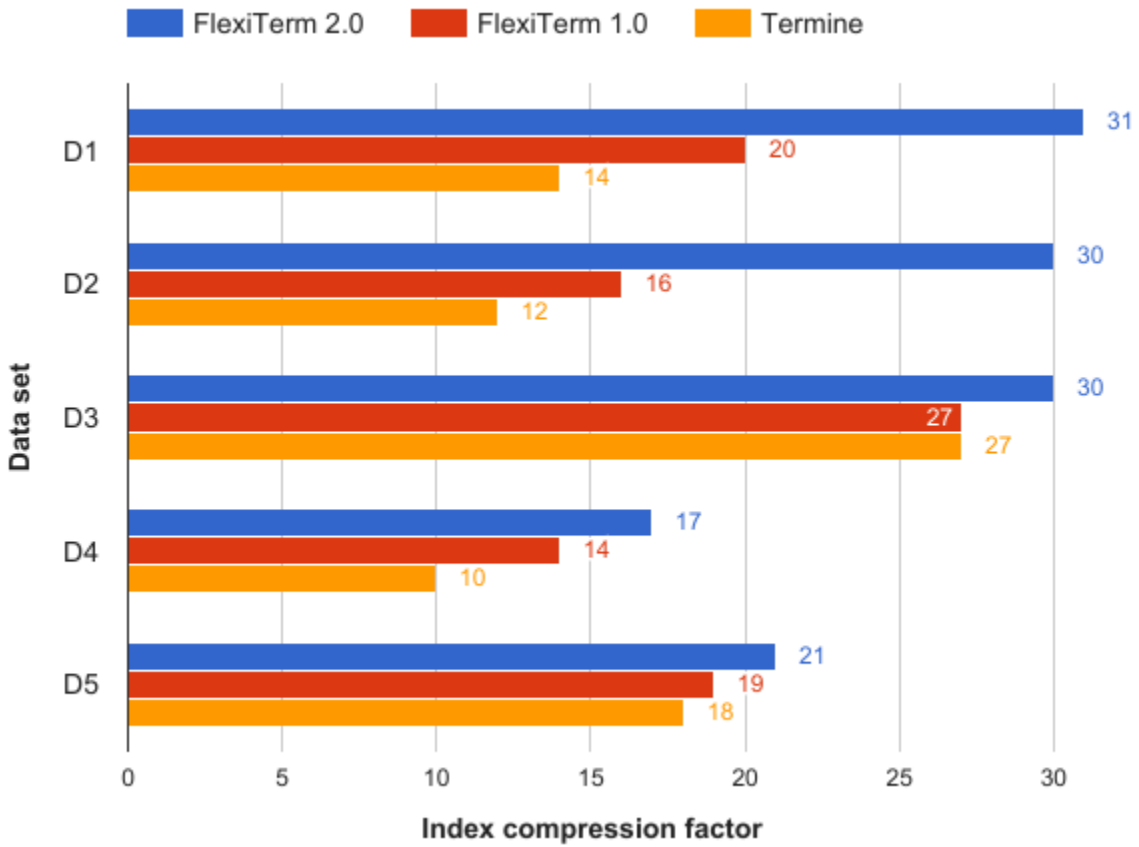
Evaluation measures

- in the context of information retrieval, we can also measure the extent to what a term–based index would be compressed by conflation of term variants
- analogous to the idea of index compression factor
- the fractional reduction in index size achieved through stemming

$$ICF = (w - s)/s$$

- w = # of distinct words, s = # of distinct stems
- w = # of distinct term variants, s = # of distinct terms (i.e. their normalised representatives)

Index compression factor



CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Data set 1

Rank	FlexiTerm 2.0	FlexiTerm 1.0	TerMine
1	TNF–alpha TNF tumor necrosis factor alpha TNF alpha	transcription factor transcription factors transcriptional factors	t cell
2	tumor necrosis factor	NF–kappa B	transcription factor
3	NF–kappa B NF–kappaB NF kappa B nuclear factor–kappa B nuclear factor kappa B nuclear factor kappaB nuclear factor–kappaB	gene expression expression of genes	NF–kappa B
4	transcription factor transcription factors transcriptional factors	transcriptional activity activator of transcription transcriptional activation activating transcription activators of transcription transcription activation transcriptional activator	gene expression
5	gene expression expression of genes	NF–kappaB activation NF–kappaB activity	cell line
6	activation of NF–kappa B NF–kappa B activation NF–kappa B activity	human T cells human cells	T lymphocyte

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Data set 1

Acronym	Full form	Frequency	Term rank	Previous term rank
NF–kappaB	nuclear factor–kappaB	36	3	31
TNF–alpha	tumor necrosis factor alpha	34	1	13
TNF	tumor necrosis factor alpha	24	1	13
CBF	core binding factor	19	10	N/A
GM–CSF	granulocyte–macrophage colony–stimulating factor	15	12	44
GR	glucocorticoid receptor	12	7	23
PMA	phorbol myristate acetate	12	19	57
AR	androgen receptor	11	36	58
HIV	human immunodeficiency virus	11	17	40
IFN–gamma	interferon gamma	11	47	N/A

Data set 2

Rank	FlexiTerm 2.0	FlexiTerm 1.0	TerMine
1	COPD chronic obstructive pulmonary disease	chronic obstructive pulmonary disease	chronic obstructive pulmonary disease
2	pulmonary disease	patients with COPD COPD patients	obstructive pulmonary disease
3	patients with COPD COPD patients patients with chronic obstructive pulmonary disease patients with chronic obstructive pulmonary diseases	pulmonary disease	pulmonary disease
4	COPD exacerbation COPD exacerbations exacerbations of chronic obstructive pulmonary disease exacerbations of COPD chronic obstructive pulmonary disease exacerbations exacerbation of chronic obstructive pulmonary disease exacerbation of COPD	acute exacerbation acute exacerbations	COPD patient
5	patients with milder disease	COPD exacerbation COPD exacerbations exacerbations of COPD exacerbation of COPD	acute exacerbation
6	AECOPD acute exacerbation of COPD acute exacerbations of chronic obstructive pulmonary disease acute exacerbations of COPD	patients with chronic obstructive pulmonary disease patients with chronic obstructive pulmonary diseases	severe COPD
7	acute exacerbation acute exacerbations	lung function	COPD exacerbation
8	QoL quality of life	exacerbations of chronic obstructive pulmonary disease chronic obstructive pulmonary disease exacerbations exacerbation of chronic obstructive pulmonary disease	lung function

Data set 2

Acronym	Full form	Frequency	Term rank	Previous term rank
COPD	chronic obstructive pulmonary disease	406	1	1
PR	pulmonary rehabilitation	26	9	27
QoL	quality of life	15	8	9
AECOPD	acute exacerbations of chronic obstructive pulmonary disease	14	6	14
OR	odd ratio	13	15	N/A
ICS	inhaled corticosteroids	10	30	35
BAL	bronchial lavage	9	42	N/A
FRC	functional residual capacity	9	21	N/A
HI	high–intensity group	9	33	N/A
CB	chronic bronchitis	8	14	24

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Data set 3

Rank	FlexiTerm 2.0	FlexiTerm 1.0	TerMine
1	COPD chronic obstructive pulmonary disease COPD disease	pulmonary rehab pulmanory rehab	pulmonary rehab
2	chronic disease	breathe easy	breathe easy
3	pulmonary rehab pulmanory rehab	vitamin D	vitamin D
4	breathe easy	lung transplantation lung transplant lung transplants lung transplantations	lung function
5	vitamin D	breathe easy groups breath easy groups breathe easy group	severe COPD
6	lung transplantation lung transplant lung transplants lung transplantations	chest infection chest infections	blood pressure
7	COPD blog	quality of life	lung disease
8	breathe easy groups breath easy groups breathe easy group	blood pressure	lung transplant
9	chest infection chest infections	lung function	chest infection
10	quality of life	rehab room	rehab room

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Data set 3

Acronym	Full form	Frequency	Term rank	Previous term rank
COPD	chronic obstructive pulmonary disease	103	1	N/A
UBE	upper body ergometer	3	13	N/A

Data set 4

Rank	FlexiTerm 2.0	FlexiTerm 1.0	TerMine
1	hospital course course of hospitalization	hospital course course of hospitalization	hospital course
2	chest pain	chest pain	present illness
3	congestive heart failure CHF sx of CHF	shortness of breath	chest pain
4	coronary artery coronary arteries	coronary artery coronary arteries	coronary artery
5	shortness of breath	present illness	blood pressure
6	blood pressure blood pressures	blood pressure blood pressures	ejection fraction
7	present illness	coronary artery disease	coronary artery disease
8	heart failure	congestive heart failure	myocardial infarction
9	coronary artery disease	myocardial infarction	congestive heart failure
10	RCA right coronary artery	ejection fraction	cardiac catheterization

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Data set 4

Acronym	Full form	Frequency	Term rank	Previous term rank
CHF	congestive heart failure	27	3	8
DVT	deep venous thrombosis	19	14	76
RCA	right coronary artery	15	10	18
PTCA	percutaneous transhepatic coronary angioplasty	11	24	73
ETT	exercise tolerance test	10	17	37
SVG	saphenous vein graft	9	25	56
PND	paroxysmal nocturnal dyspnea	7	30	56
CCU	cardiac care unit	6	36	60
COPD	chronic obstructive pulmonary disease	6	53	N/A
UTI	urinary tract infection	6	21	31

Data set 5

Rank	FlexiTerm 2.0	FlexiTerm 1.0	TerMine
1	cruciate ligaments	MRI KNEE	collateral ligament
2	ACL anterior cruciate ligament	collateral ligaments	medial meniscus
3	collateral ligaments	medial meniscus medial mensicus	lateral meniscus
4	MRI KNEE	lateral meniscus	hyaline cartilage
5	PCL posterior cruciate ligament	hyaline cartilage	posterior horn
6	medial meniscus medial mensicus	posterior horn	femoral condyle
7	lateral meniscus	joint effusion	joint effusion
8	MCL medial collateral ligament	MRI RT KNEE MRI KNEE RT	MRI LT KNEE
9	hyaline cartilage	MRI LT KNEE MRI KNEE LT	lateral femoral condyle
10	posterior horn	lateral femoral condyle	medial femoral condyle

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Data set 5

Acronym	Full form	Frequency	Term rank	Previous term rank
ACL	anterior cruciate ligament	97	2	N/A
PCL	posterior cruciate ligament	57	5	N/A
MCL	medial collateral ligament	35	8	17
LCL	lateral collateral ligament	3	33	36

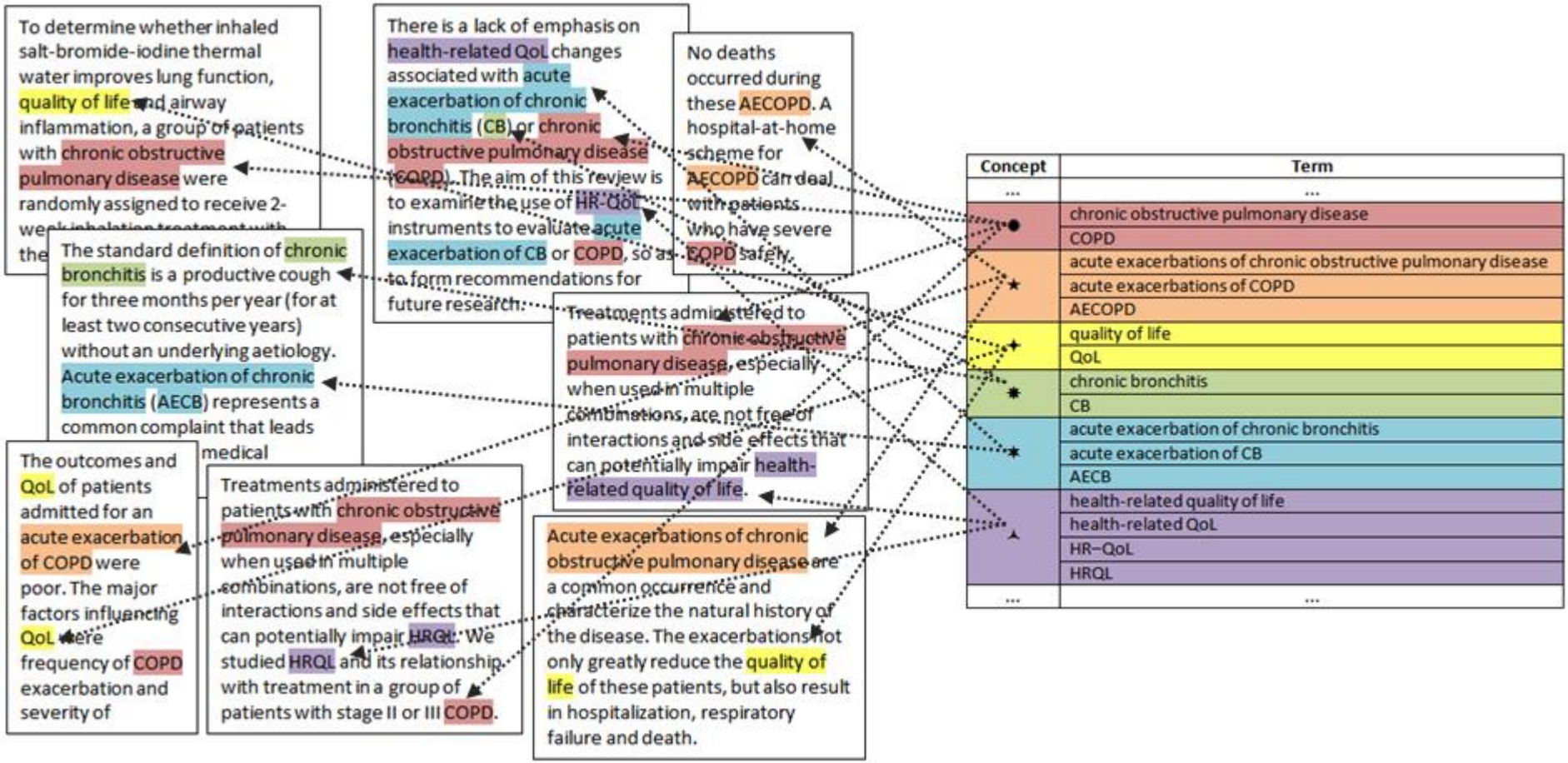
Conclusion

- acronyms significantly improve the performance of multi-word term recognition in terms of:
  - recall
    - from false negatives to true positives
  - term conflation
    - concepts as latent variables
    - statistical analysis, e.g. topic modelling
  - ranking
    - implications for content analysis

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Further information

https://users.cs.cf.ac.uk/I.Spasic/flexiterm/



Thank you! Questions?

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD

Title

CARDIFF UNIVERSITY
PRIFYSGOL CAERDYDD