Discovery of listening experiences

Phase two (June 2018)

Enrico Daga
enrico.daga@open.ac.uk
@enridaga

Outline

- Discovering listening experiences
- Benchmark development
- Learning listening experiences
- Experiments results
- Next steps

Problem / *Objective*

How to identify accounts of listening experiences from texts?

An automatic bookmarks generator for texts identifying candidates Listening Experiences

Approach

We hypothesised that LEs are a subset of the texts talking about music.

Phase 1:

To develop a dictionary of terms whose occurrence in a text could signify a discourse about music.

To show that this dictionary represents well Listening Experiences (LE) in the database.

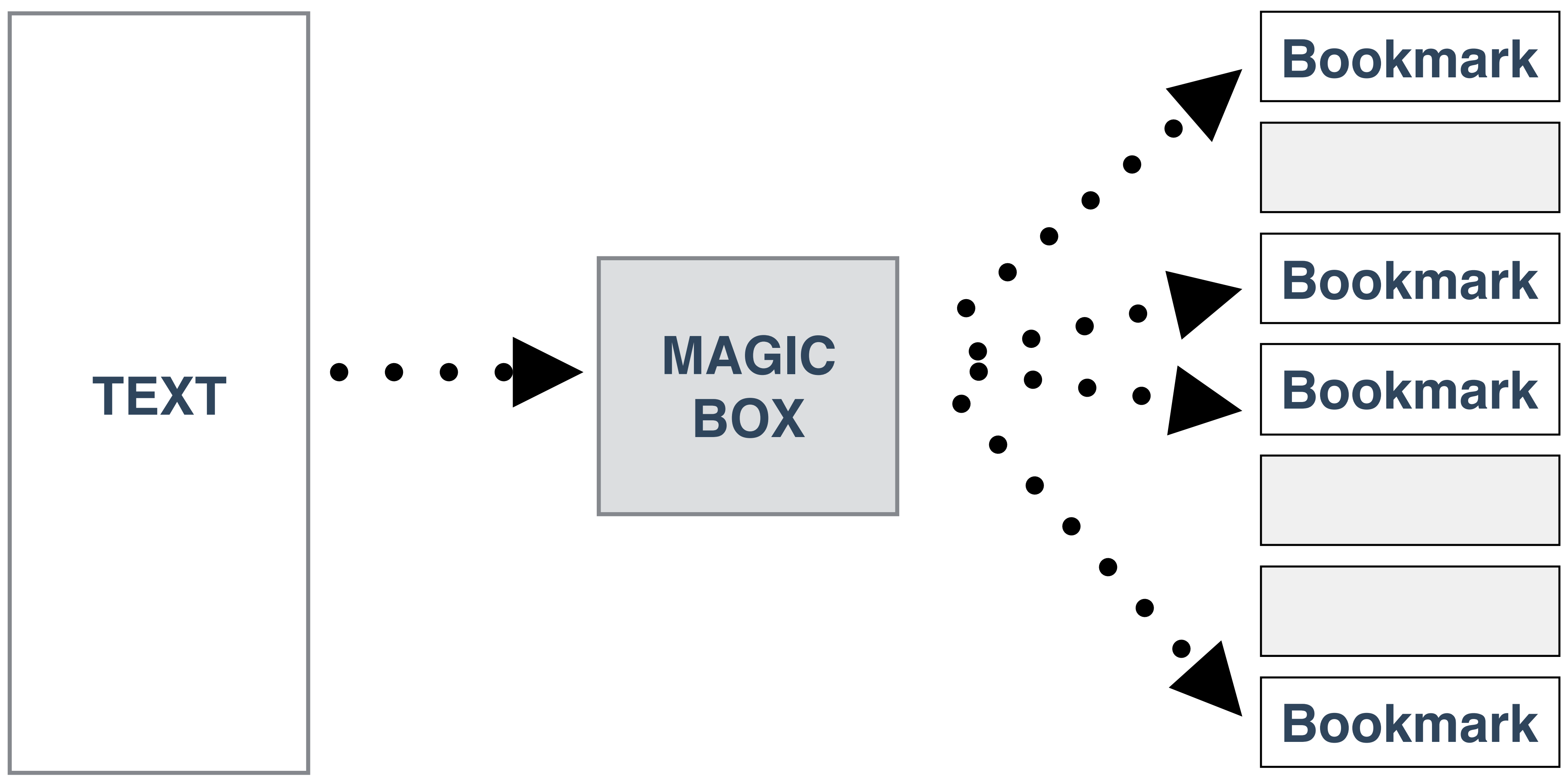
Phase 2 (in progress):

To develop a approach using the dictionary in combination with *features* of LE and evaluate it on a gold standard of LE and associated sources.

Phase 3:

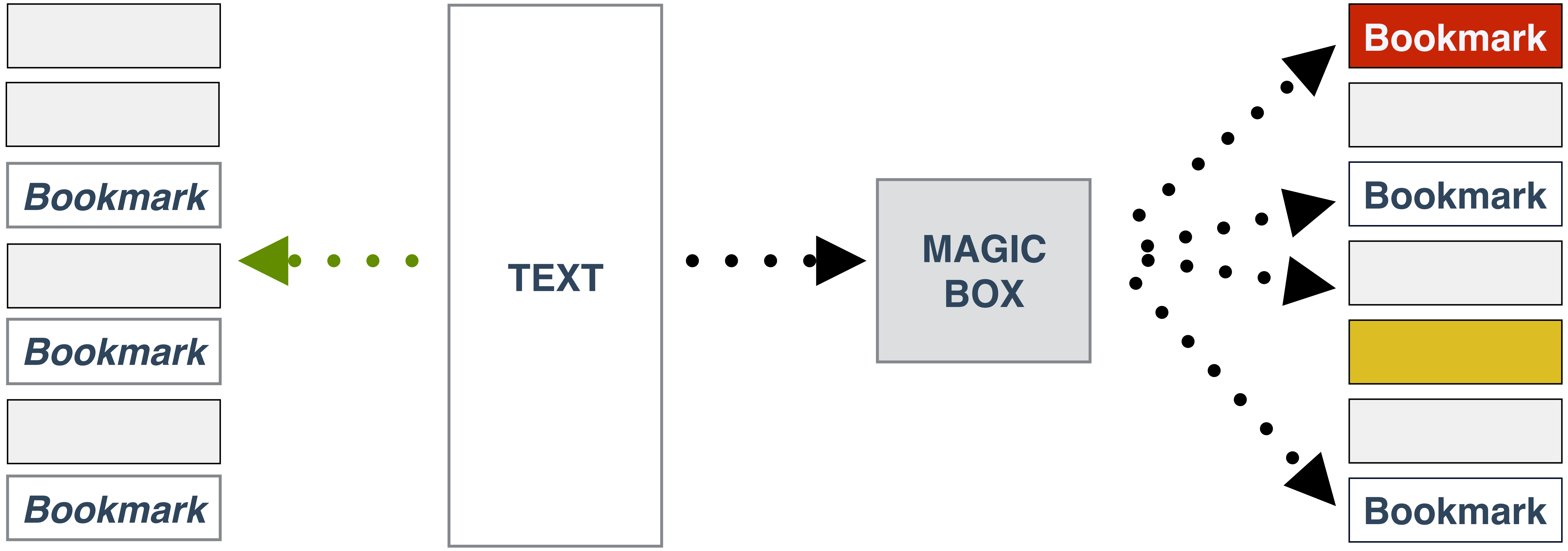
To develop a system that generate annotations of texts and evaluate it with a user study.

Discovery Listening Experiences in texts



- Objective: point the reader to the areas in the text where there is a high chance of finding a LE
- What should go in the magic box?
- How do we know that (whatever we do) is working well?

Discovery Listening Experiences in texts



- We need a benchmark!
- A collection of texts for which we already know where LEs are!

Benchmark development

- We can collects sources from the Listening Experiences Database
  - With a full text available
- However, we still don’t know where the excerpts are located in the text!
- Re-finding them manually and write down the begin/end offsets is not an option …
- We could search for the excerpt in the text!


This very handsome woman [Grassini]_was in every thing the direct contrary of her rival_[Mrs. Billington]. With a beautiful form, and a grace peculiarly her own, she was an excellent actress, and her style of singing was exclusively the cantabile, which became heavy

a la longue, and bordered a little on the monotonous: for her voice, which it was said had been a high soprano, was by some accident reduced to a low and confined contralto. She had entirely lost all its upper tones, and possessed little more than one octave of good natural notes; if she attempted to go higher, she produced only a shriek, quite unnatural, and almost painful to the ear. Her first appearance was in La Vergine del Sole, an opera of Mayer’s, well suited to her peculiar talents; but her success was not very decisive as a singer […]

*This very handsome woman was in every_*
*thing the direct contrary of her rival. With_*
*a beautiful form, and a grace peculiarly her_*

_

_

_

92 GRASSINI._

_

*own, she was an excellent actress, and her_*
style of singing was exclusively the cantabile,_
*which became heavy d la tongue, and bordered_*
a little on the monotonous_: for her voice,_
which it was said had been a high soprano,_
*was by some accident reduced to a low and_*
*confined contralto. She had entirely lost all_*
*its upper tones, and possessed little more than_*
*one octave of good natural notes_; if she_*
*attempted to go higher, she produced only a_*
*shriek, quite unnatural, and almost painful to_*
*the ear. Her first appearance was in La_*
*Vergine del Sole, an opera of Mayer's, well_*
*suited to her peculiar talents ; but her success_*
was not very decisive as a singer, though her

Musical Reminiscences: Containing an Account of Italian Opera in England, From 1773. The Fourth Edition, Continued to the Present Time, and Including The Festival in Westminster Abbey.

```
Excerpts[] = get_led(Source)
Text = text(Source)
best[t,b,e,s] ; // text, begin, end, score
Foreach excerpt in Excerpts[]:
   words[] = tokenize(excerpt)
   words[] = sortByLengthDesc(words[]) // Longest on top
   Foreach word in words[]:
     occurrences[][b,e] = find(word, Text)
     position[b,e]  = find(word, excerpt)
     Foreach occurrence[b,e] in occurrences[][b,e]:
        begin = occurrence.b - position.b
        end   = occurrence.e + len(excerpt) - position.e
        possible = substring(Source, begin, end)
        score = levenshtein(excerpt, possible)
        if(score < best[s])
          best[t,b,e,s] = [possible, begin, end, score]
        fi
     End
   End
End
return best
```

Benchmark development

- Collected sources from the Listening Experiences Database
    - With a full text available
- Reached approximately 10% of the total set
- 17 books, 1098 experiences
- We generated a collection of bookmarks with the algorithm described to be used to test possible approaches

https://docs.google.com/spreadsheets/d/17wZS509Rkp5sNQ18x48AAiUwnQXKecgUIUzoiooFUo8

RECOLL	Recollections of an old musician by Ryan, Thomas, 1827-1903	92
GLOVER	Jimmy Glover: His Book	7
MUSING	Musings and Memories of a Musician	48
PILGRIM	Westminster Pilgrim; Being a Record of Service in Church, Cathedral and Abbey, College University and Concert Room, with a Few Notes on Sport	24
GRONOW	Anecdotes of celebrities of London and Paris : to which are added the last recollections of Captain Gronow, formerly of the First Foot Guards	5
TWOMEN	Two Men: a Memoir	12
OUSELEY	Memorials of Frederick Arthur Gore Ouseley, baronet, M.A., doctor and professor of music in the University of Oxford; preceptor and canon residentary in the cathedral church of Hereford:	1
MULIFE	My Musical Life	68
THIRTY	Thirty years of musical life in London	111
REOPER	Reminiscences of the Opera	73
WESLEY	The Journal of John Wesley	8
RECMUS	Recent music and musicians as described in the diaries and correspondence of Ignatz Moscheles	72
MASONB	Musical Letters from Abroad	179
MUSREM	Musical Reminiscences: Containing an Account of Italian Opera in England, From 1773. The Fourth Edition, Continued to the Present Time, and Including The Festival in Westminster Abbey.	123
REMKEL1	Reminiscences of Michael Kelly (Vol 1), of the King's Theatre, and Theatre Royal Drury Lane, including a period of nearly half a century; with original anecdotes of many distinguished	68
REMKEL2	Reminiscences of Michael Kelly (Vol 2), of the King's Theatre, and Theatre Royal Drury Lane, including a period of nearly half a century; with original anecdotes of many distinguished	105
MEMMUS	Memories of a musician	102

Benchmark development

- Assumptions:
  - The LED database contains all the Listening Experiences in these books (is that true?)
  - It is enough to identify a text area that overlaps with any bookmark in the benchmark

https://docs.google.com/spreadsheets/d/17wZS509Rkp5sNQ18x48AAiUwnQXKecgUlUzoiooFUo8

Approaches

Musical Heat Annotator:

- based on the hypothesis that LEs are a subset of the texts talking about music.
- the more the text overlaps with **Gutenberg-M** the more it could be a LE

Random Forest Annotator:

- based on the assumption that we can train a classifier abstracting features from LE texts
- A Random Forest Classifier is a Machine Learning approach used successfully for binary texts classification: "*it creates a set of decision trees from randomly selected subset of training set. It then aggregates the votes from different decision trees to decide the final class of the test object.*"

Gutenberg-M: Development of a music dictionary

- Gutenberg corpus (english subset)
- We use NLP to get a vector of terms for each documents (StanfordNLP)
- We calculated TF/IDF of each doc/term pair in Gutenberg
- We collecting the terms in documents classified in the <u>Music shelf</u>
- We sorted them by relevance towards the sub-corpus
- We validated the dictionary against the LED set and the Reuters-21578 corpus (as negative)

From Phase 1

Gutenberg-M: Text to vector (NLP)

- Removing stopwords, keeping Full POS information
- Example: *"So the Rontgens have played you the new Brahms symphony! - another of my few musical joys taken from me! It always happens that when I have been specially counting on something of the sort as regards you, Fate […]"* - LED-1438250799133

```
0  rontgen[NNS]
1  play[VBN]
2  Brahms[NNP]
3  symphony[NN]
4  another[DT]
5  musical[JJ]
6  take[VBN]
7  always[RB]
8  happen[VBZ]
9  specially[RB]
10 count[VBG]
11 something[NN]
12 sort[NN]
13 regard[VBZ]
14 Fate[NNP]
```

From Phase 1

Gutenberg-M: Text to vector (NLP)

CC Coordinating conjunction
CD Cardinal number
DT Determiner
EX Existential there
FW Foreign word
IN Preposition or subordinating conjunction
JJ Adjective
JJR Adjective, comparative
JJS Adjective, superlative
LS List item marker
MD Modal
NN Noun, singular or mass
NNS Noun, plural
NNP Proper noun, singular
NNPS Proper noun, plural
PDT Predeterminer
POS Possessive ending
PRP Personal pronoun
PRP$ Possessive pronoun

RB Adverb
RBR Adverb, comparative
RBS Adverb, superlative
RP Particle
SYM Symbol
TO to
UH Interjection
VB Verb, base form
VBD Verb, past tense
VBG Verb, gerund or present participle
VBN Verb, past participle
VBP Verb, non3rd person singular present
VBZ Verb, 3rd person singular present
WDT Whdeterminer
WP Whpronoun
WP$ Possessive whpronoun
WRB Whadverb

Gutenberg-M: TFIDF

*"In information retrieval, tf–idf or TFIDF, short for term frequency–inverse document frequency, is a numerical statistic that is intended to reflect **how important a word is to a document** in a collection or corpus." [1]*

```
term_freq = term_usages / doc_size
idf = LOG(48790 / num_docs_with_term)
tf_idf = term_freq*idf
```

Highest TF-IDF: 1.5901121823585802
Lowest TF-IDF: 4.032538525747152e-08
Highest TF-IDF in the Music Shelf: 0.0922981613222286
Lowest TF-IDF in the Music Shelf: 7.517321708209822e-07

[1] https://en.wikipedia.org/wiki/Tf-idf


```
Document: Gutenberg-15141
-
Beethoven[NNP]     0.07272755403226193
Symphony[NNP]      0.015139485794100219
Schindler[NNP]     0.007967133189523013
Vienna[NNP]        0.007256378255299395
Haydn[NNP]         0.0071413210885995495
Wagner[NNP]        0.007088376068171141
Breuning[NNP]      0.006717815731235641
Ries[NNP]          0.006111818988630585
Mozart[NNP]        0.0059785964542184945
Lichnowsky[NNP]    0.005846276132915727
quartet[NNS]       0.0054224336619273
Czerny[NNP]        0.005217816538462906
Mass[NNP]          0.005135716029154898
opus[NN]           0.004832913297756952
composer[NN]       0.004442636696952911
Karl[NNP]          0.004326928936343346
Holz[NNP]          0.004142928952284239
Bach[NNP]          0.0037425004179032417
sonata[NNS]        0.0035618383556334826
Bonn[NNP]          0.00355707250098514
symphony[NNS]      0.003447084601144992
music[NN]          0.0032652203770744768
```

From Phase 1

Gutenberg-M: Statistics

- Gutenberg (english): **48790** documents, **79** in the *Music* shelf
- Number of doc/terms occurrences: 1.460.211.421
- Number of distinct terms: 7.183.327
- Number of terms occurring only in 1 doc: 4.405.918
- Number of doc/terms in the Music Shelf: 1.934.581
- Number of distinct terms in the Music Shelf: 89.883
    - **1.25%** of the total of distinct terms in the corpus

From Phase 1

Gutenberg-M: Dictionary

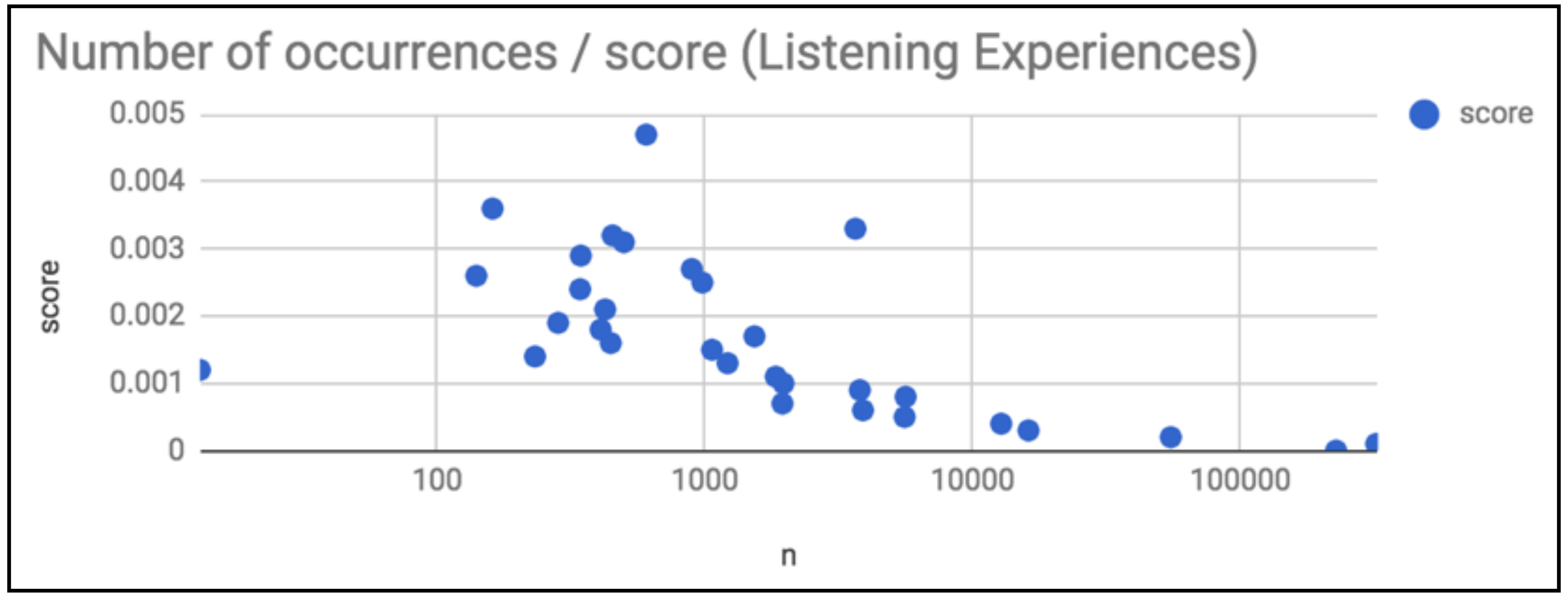
- 89.883 terms ordered by relevance
- Relevance = AVG(TFIDF) of docs in Music Shelf

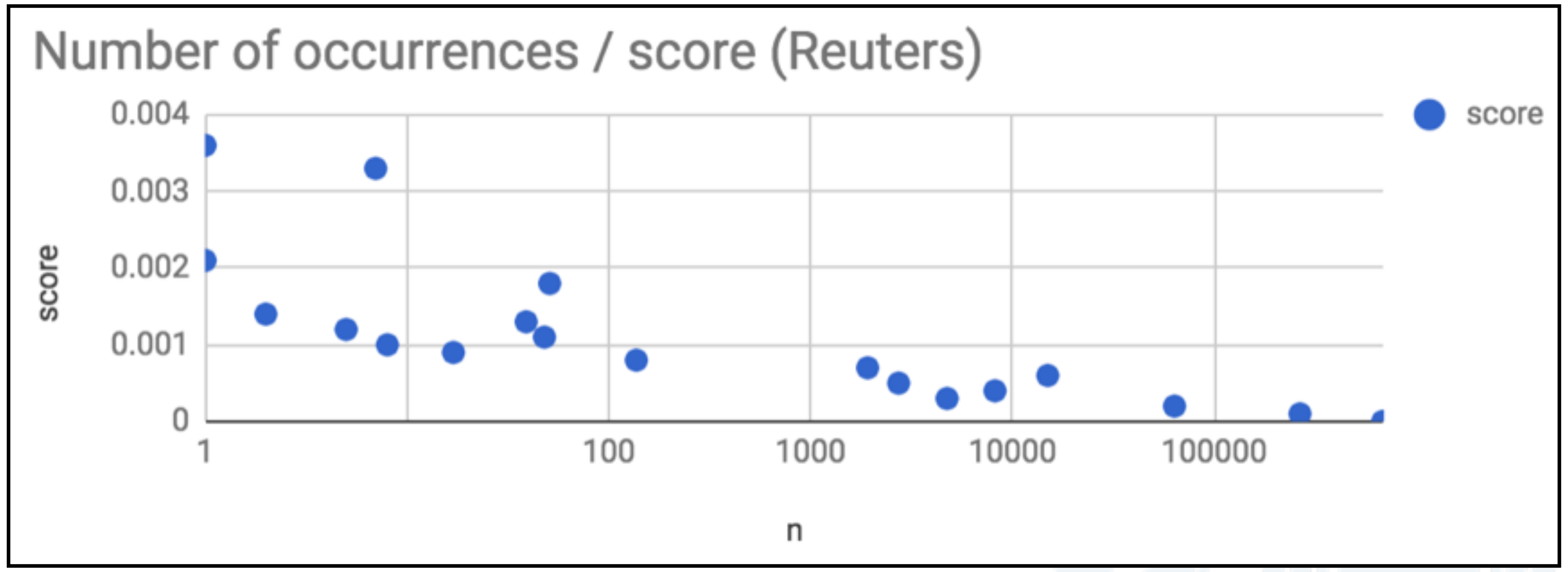
Beethoven[NNP]	0.004708996602	1
vocal[JJ]	0.003577405412	2
music[NN]	0.003279422105	3
Liszt[NNP]	0.003201453413	4
Chopin[NNP]	0.003163986853	5
composer[NN]	0.003115849809	6
Mozart[NNP]	0.002860199248	7
musical[JJ]	0.002722584954	8
Haydn[NNP]	0.002579207714	9
piano[NN]	0.002500942374	10
aria[NN]	0.0006770586871	98
fugue[NN]	0.0006655704232	99
theme[NN]	0.0006590153165	100
accent[NN]	0.000222760115	497
master[NNS]	0.0002227463667	498
Dickens[NNP]	0.0002227386521	499
resonance-chamber[NNS]	0.0002226351367	500
leading-tone[NN]	0.0002224820318	501
florid[JJ]	0.0001438729148	997
sound[VBZ]	0.000143856694	998
score[NNS]	0.0001437556948	999
rondo[NN]	0.0001435829476	1000
sweet[JJ]	0.0001435409753	1001
sense[NN]	0.0001434473773	1002
gesture[NNS]	9.09E-05	1997
hammer[NNS]	9.08E-05	1998
flow[NN]	9.08E-05	1999
sorrow[NN]	9.08E-05	2000
monophonic[JJ]	9.08E-05	2001
saint[NNS]	4.79E-05	4997
move[VBZ]	4.79E-05	4998
moderately[RB]	4.79E-05	4999
Cecilia[NNP]	4.79E-05	5000
Nibelung[NNP]	4.79E-05	5001
mean[VBD]	2.80E-05	9997
aloft[RB]	2.80E-05	9998
o'er[RB]	2.80E-05	9999
unaffected[JJ]	2.80E-05	10000
Stockhausen[NNP]	2.80E-05	10001
indulgent[JJ]	1.42E-05	19997
emulation[NN]	1.42E-05	19998
emerge[VB]	1.42E-05	19999
two-step[NNS]	1.42E-05	20000
Lauriett[NNP]	1.42E-05	20001
unfitness[NN]	5.86E-06	39997
Aryan[NNP]	5.86E-06	39998
Sirens[NNPS]	5.86E-06	39999
MACREADY[NNP]	5.86E-06	40000
fence[VBN]	5.85E-06	40001
offrir[FW]	3.18E-06	59997
postes[FW]	3.18E-06	59998
Dorf[NNP]	3.18E-06	59999
Dewing[NNP]	3.18E-06	60000
legitimise[VBN]	3.18E-06	60001

From Phase 1

Gutenberg-M: Validation

We compared Listening Experiences and the Reuters-21578 corpus [1] (used to benchmark news classification systems, does not include music as category).





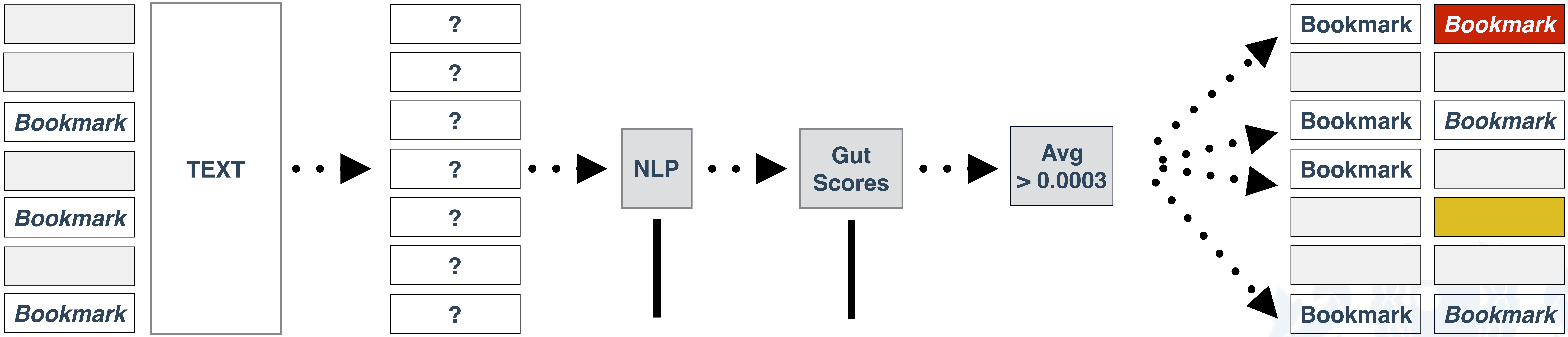
- We matched the vector of each corpus with the music dictionary, and clustered the number of occurrences per score range (log scale in the pictures)
- We calculated a distribution score (sum(scores) / corpus vector length)
  - LE (vector length: 949301) is **0.000**11480226659861874
  - Reuters-21578 (vector length: 1372059) is 4.6368513916777576e-05 (**0.0000**4636851)

The dictionary fits better LEs then Reuters

[1] http://kdd.ics.uci.edu/databases/reuters21578/reuters21578.html

From Phase 1

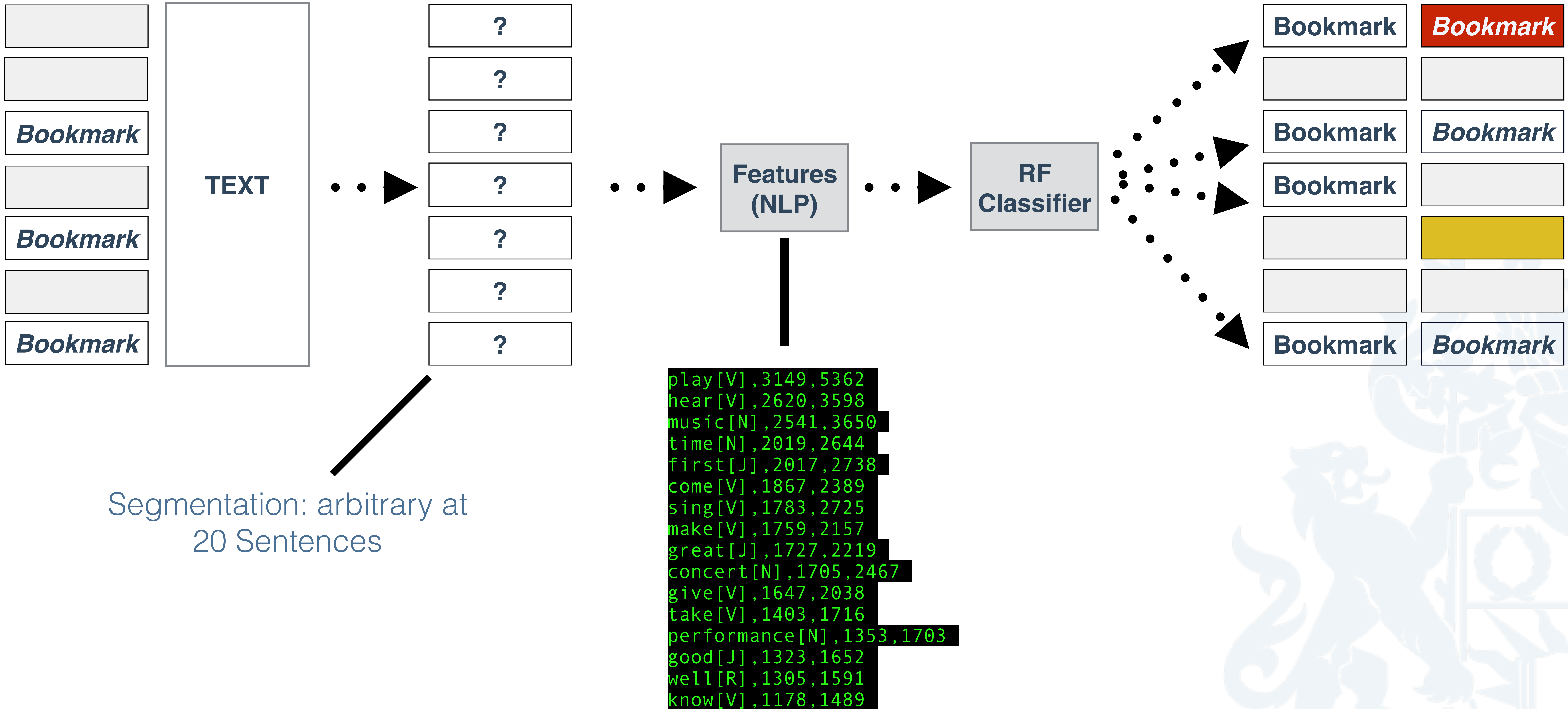
Musical Heat Annotator



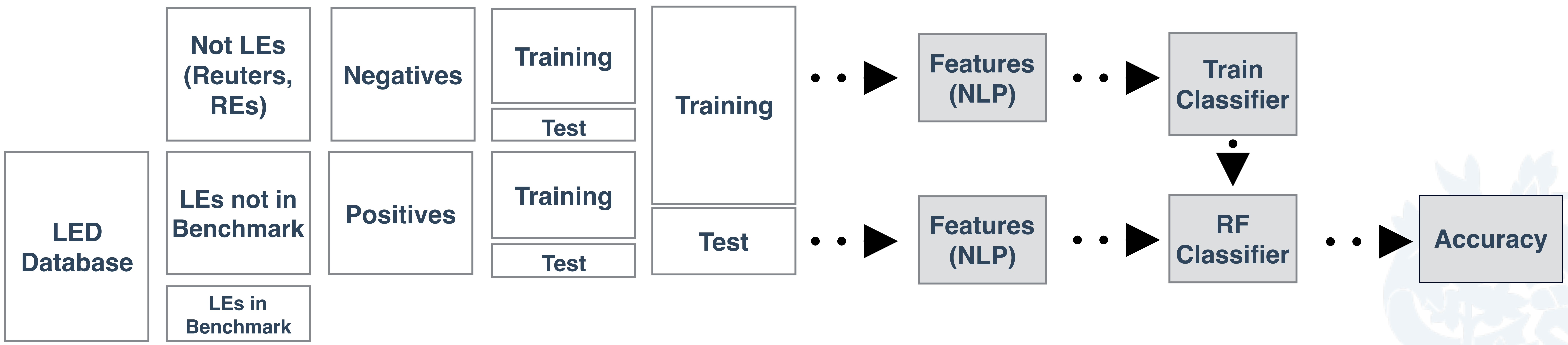
Segmentation: arbitrary at 20 Sentences

0	rontgen[NNS]	0.00072727554032261 93
1	play[VBN]	0.00000796713318952301 3
2	Brahms[NNP]	0.000725637825529939 5
3	symphony[NN]	0.000714132108859954 95
4	another[DT]	0.000070088376068171141
5	musical[JJ]	0.00006717815731235641
6	take[VBN]	0.000005978596454218494 5
7	always[RB]	0.000058462761329157 27
8	happen[VBZ]	0.0000054224336619273
9	specially[RB]	0.000521781653846290 6
10	count[VBG]	0.000513571602915489 8
11	something[NN]	0.000483291329775695 2
12	sort[NN]	0.000444263669695291 1
13	regard[VBZ]	0.0000043269289363433 46
14	Fate[NNP]	0.0000041429289522842 39

Random Forest Annotator



Random Forest Annotator



Random Forest: Training Set

Components:

- Listening Experiences (LE) - 9059 (without the ones in the Benchmark)
- Reuters21578 (Reu) - usually shorter - 19043
- Reading Experiences Database (Red) - 11727

Combinations:

- LE + Reu: 9059 positives + 9059 negatives of the same average length
- LE + Red: 9059 positives + 6161 negatives of the same average length
- LE + Reu + Red (Same size) 9059 positives + 9059 negatives of the same size
- LE + Neg (Reu + Red original form all) 9059 positives and 30771 negatives

Features (Learnt Vocabulary)

• all	Terms from all the training set
• le	Led vocabulary all
• le10k	Led frequency top 10k
• le5k	Led frequency top 5k
• le1k	Led frequency top 1k

How do we measure the performance

Our objective:

- Reduce the number of areas to supervise
- Reduce the number of areas that are not proposed but contain a LE (< F->T)

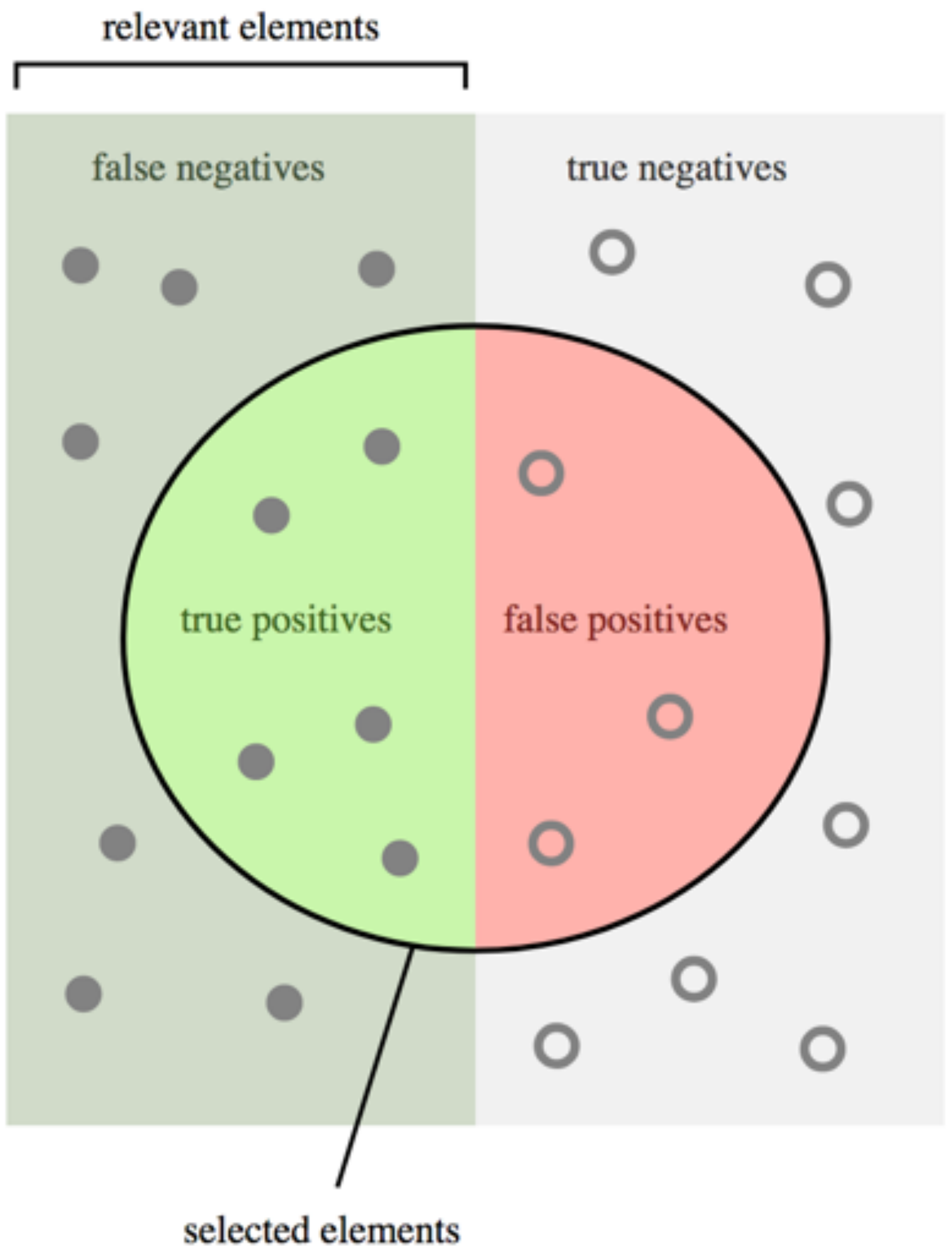
As a Machine Learning problem:

- Accuracy: maximise the correct classification (T->T, F->F)

As a Information Retrieval problem:

- F1: maximise the positives returned (T->T)

As an Information Retrieval problem



How many selected items are relevant?

Precision = 

How many relevant items are selected?

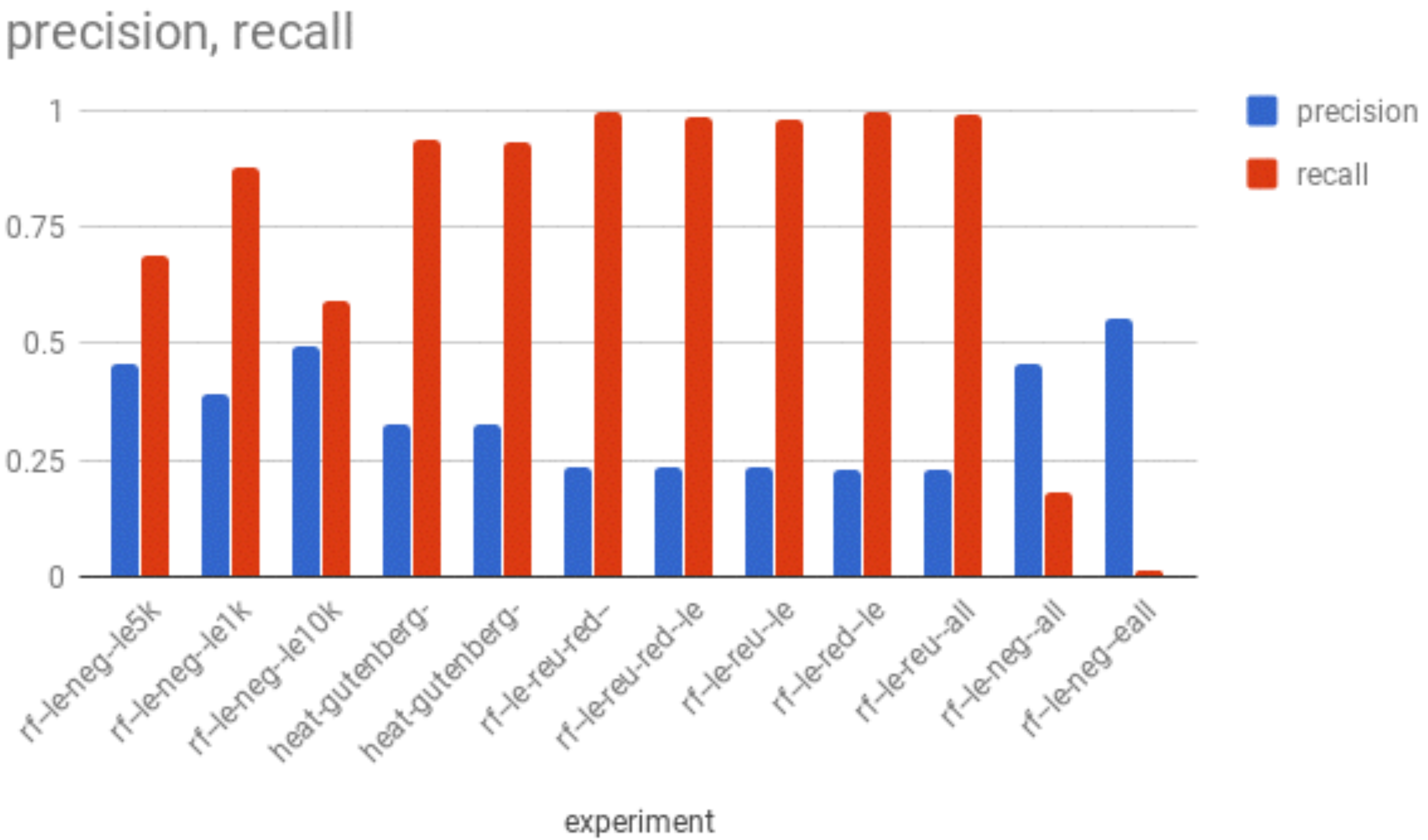
Recall = 

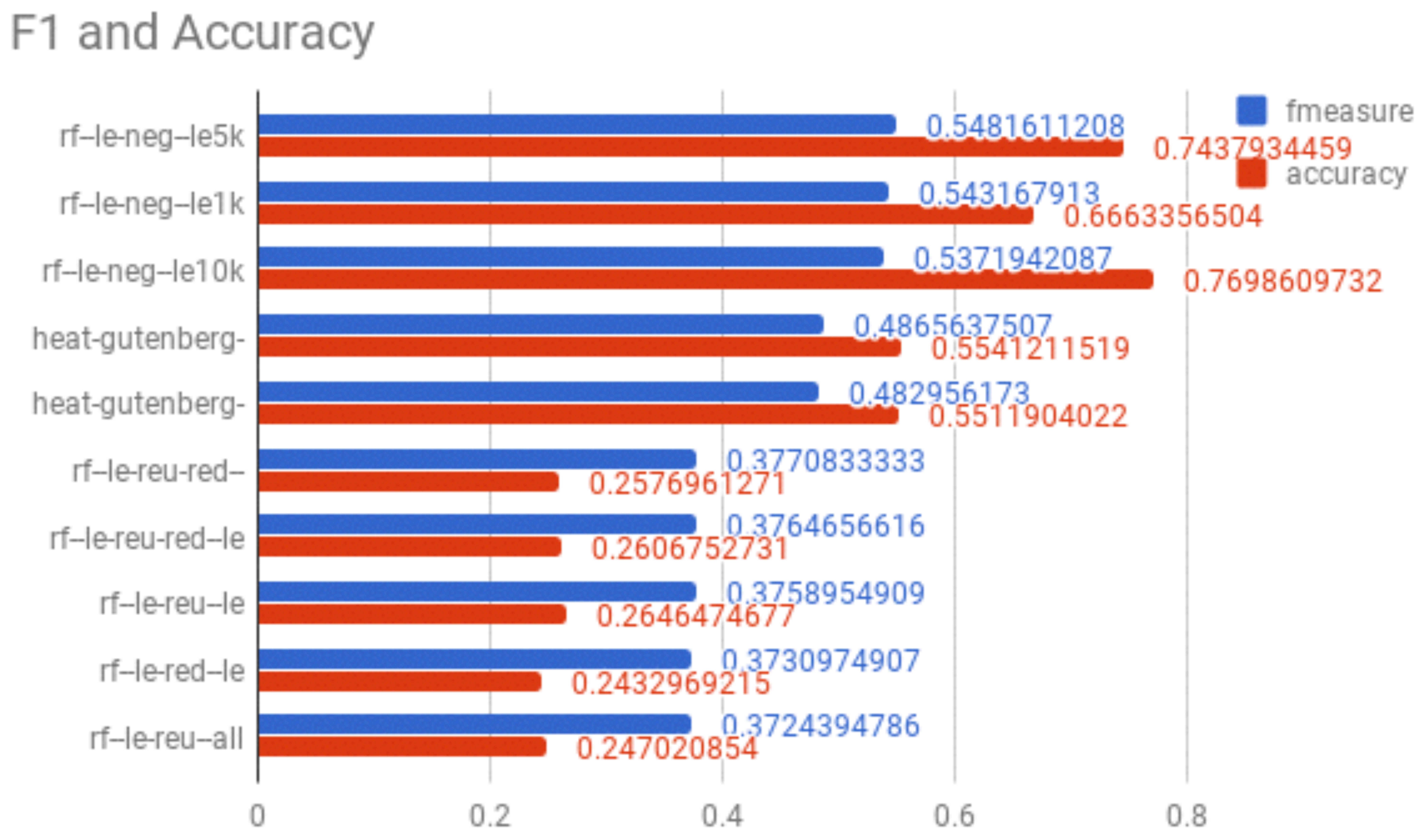
Our objective:

- Maximise recall
  - (maximise the true positives)
- Maximise precision
  - (minimise the false positives)

$$F_1 = \frac{2}{\frac{1}{\mathrm{recall}} + \frac{1}{\mathrm{precision}}} = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}}.$$

experiment	population	relevant	retrieved	positives	precision	recall	fmeasure	accuracy	error
rf--le-neg--le5k	4028	910	1374	626	0.45560407	0.68791208	0.54816112	0.7437934	0.25620655
rf--le-neg--le1k	4028	910	2032	799	0.39320866	0.87802197	0.543167913	0.6663356	0.33366434
rf--le-neg--le10k	4028	910	1093	538	0.49222323	0.59120879	**0.537194208**	**0.7698609**	0.23013902
heat-gutenberg-00003-20	4028	910	2588	851	0.32882534	**0.93516483**	0.486563750	0.554121	0.44587884
heat-gutenberg-00003	16087	3627	10337	3372	0.32620682	0.92969396	0.482956173	0.551190	0.44880959
rf--le-reu-red--le10k	4028	910	3890	905	0.23264781	0.99450549	0.377083333	0.2576961	0.74230387
rf--le-reu-red--le	4028	910	3866	899	**0.23254009**	**0.98791208**	0.376465661	**0.2606752**	0.73932472
rf--le-reu--le	4028	910	3836	892	0.23253388	0.98021978	0.375895490	0.2646474	0.73535253
rf--le-red--le	4028	910	3952	907	0.22950404	0.99670329	0.373097490	0.2432969	0.75670307
rf--le-reu--all	4028	910	3923	900	0.22941626	0.98901098	0.372439478	0.2470208	0.752979146
rf--le-neg--all	4028	910	**354**	162	0.45762711	**0.17802197**	0.25632911	**0.7666335**	0.233366435


precision, recall
1
0.75
0.5
0.25
0
precision
recall
rf--le-neg--le5k
rf--le-neg--le1k
rf--le-neg--le10k
heat-gutenberg-
heat-gutenberg-
rf--le-reu-red-
rf--le-reu-red--le
rf--le-reu--le
rf--le-red--le
rf--le-reu-all
rf--le-neg--all
rf--le-neg--eall
experiment


F1 and Accuracy
fmeasure
accuracy
rf--le-neg--le5k
rf--le-neg--le1k
rf--le-neg--le10k
heat-gutenberg-
heat-gutenberg-
rf--le-reu-red--
rf--le-reu-red--le
rf--le-reu--le
rf--le-red--le
rf--le-reu--all
0.5481611208
0.7437934459
0.543167913
0.6663356504
0.5371942087
0.7698609732
0.4865637507
0.5541211519
0.482956173
0.5511904022
0.3770833333
0.2576961271
0.3764656616
0.2606752731
0.3758954909
0.2646474677
0.3730974907
0.2432969215
0.3724394786
0.247020854
0
0.2
0.4
0.6
0.8

Discussion

- The Heat hypothesis seems confirmed but it is not enough
  - good recall but low precision and accuracy
- A training set with examples unbalanced on the negatives works better
  - *LEs too much similar to non-LEs?*
  - *LEs features are hard to capture …*
- A training set limited to the most frequent features works better
  - *LEs features are very specific? But which are they?*

How to improve

- Analyse false negatives and derive hints about how to improve
- Combine the Heat recall with the Random Forest precision
  - Use Gutenberg-M as feature set for RF?
  - Regenerate Gutenberg-M using a stronger POS abstraction
- Try other ML approaches like Support Vector Machines (SVM)
- Try different segmentation strategies (smaller texts, moving window)
- Explore with unsupervised learning and try to capture LEs features by explaining clusters

Summary

Progress:

- Built a benchmark using the LEs and their original sources
- We experimented on learning LEs from the LED database and reached a **baseline** performance

Perspectives:

- Elaborate on the approach to reach a good performance (almost all LEs + the fewest negatives)
- Develop a system that generate annotations of texts and evaluate it with a user study.