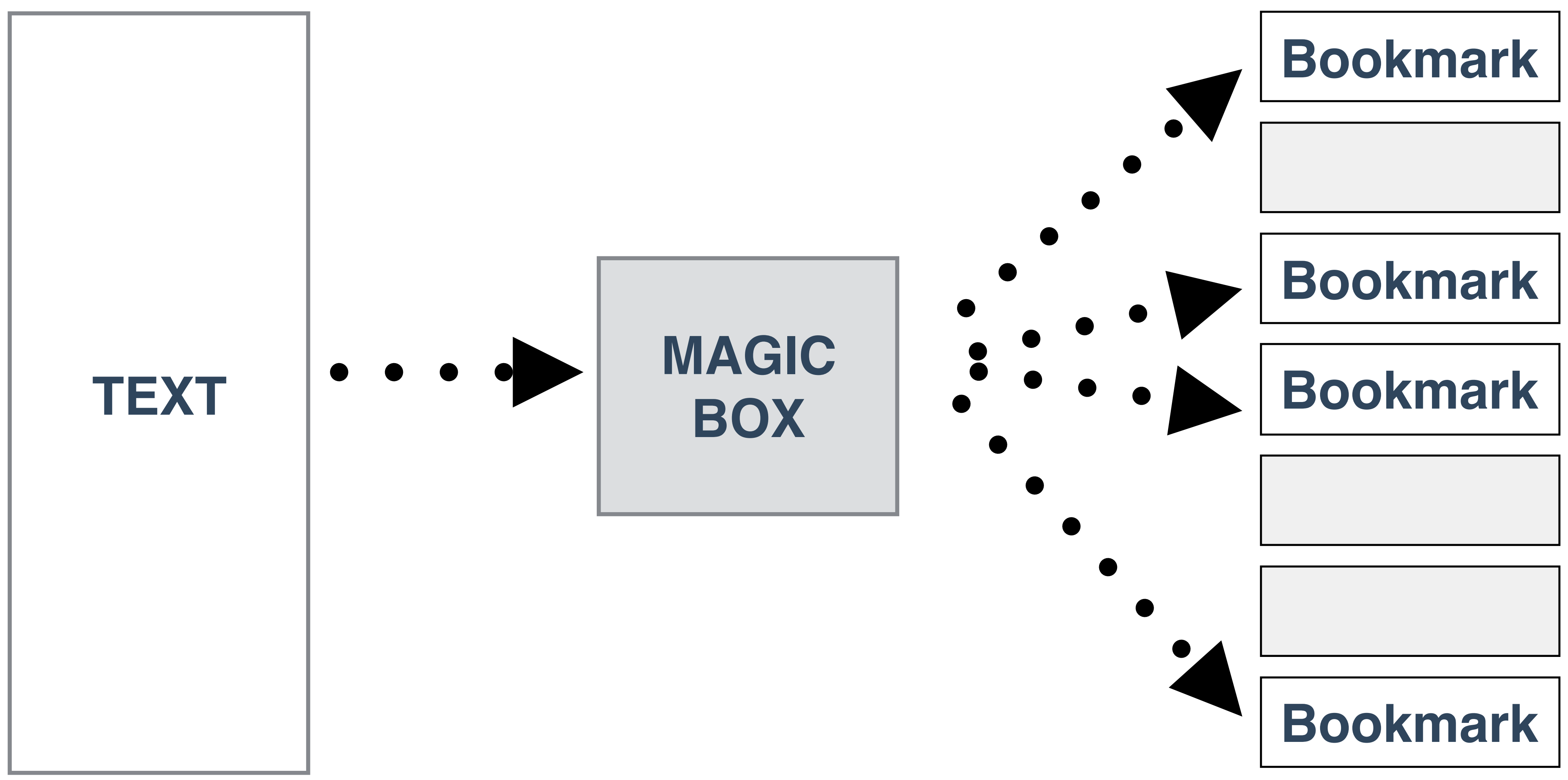
Discovery of listening experiences

Phase *third* (September 2018)

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

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?

Recap: where we left

Phase 1:

We hypothesised that LEs are a subset of the texts talking about music (*Heat hypothesis*). We developed a dictionary of musical terms (Music-Gut) using the Gutenberg English corpus (TFIDF). This dictionary represents well Listening Experiences (LE) in the database.

Phase 2:

We developed and compared a set of approaches (in several variants):

- **Music-Gut**: using the dictionary in combination with *features* of LE and evaluate it on a gold standard of LE and associated sources.
- **Music-Forest**: implementing a Random Forest Classifier trained on LED, RED, and Reuters
- **Music-Emb**: using a dictionary of word-embeddings computed from Gutenberg English **(new!)**
- **Sentiment**: using a dictionary computed from SentiWordNet **(new!)**

Phase 3 (in progress):

To develop a system that generate annotations of texts

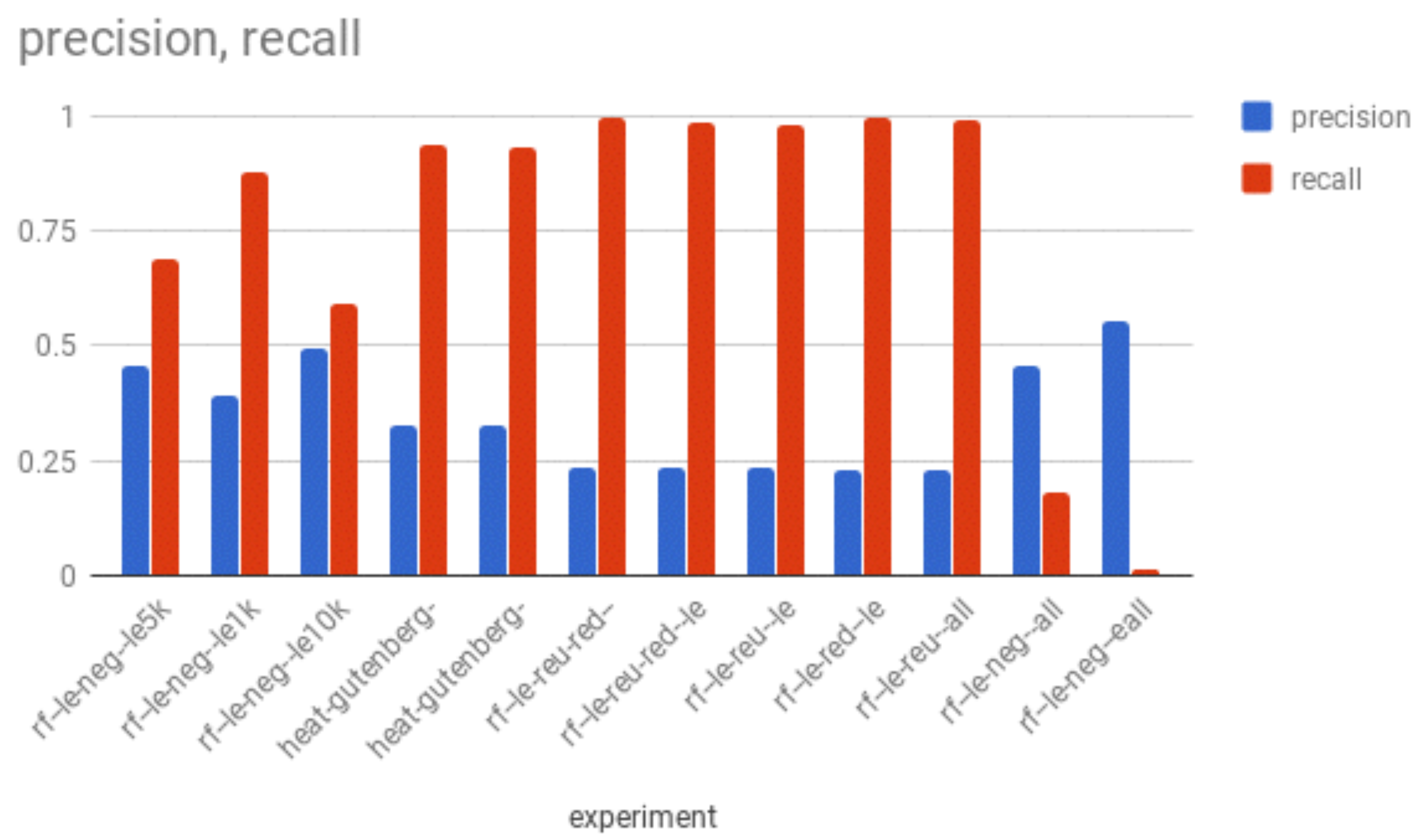
To evaluate it with a user study.

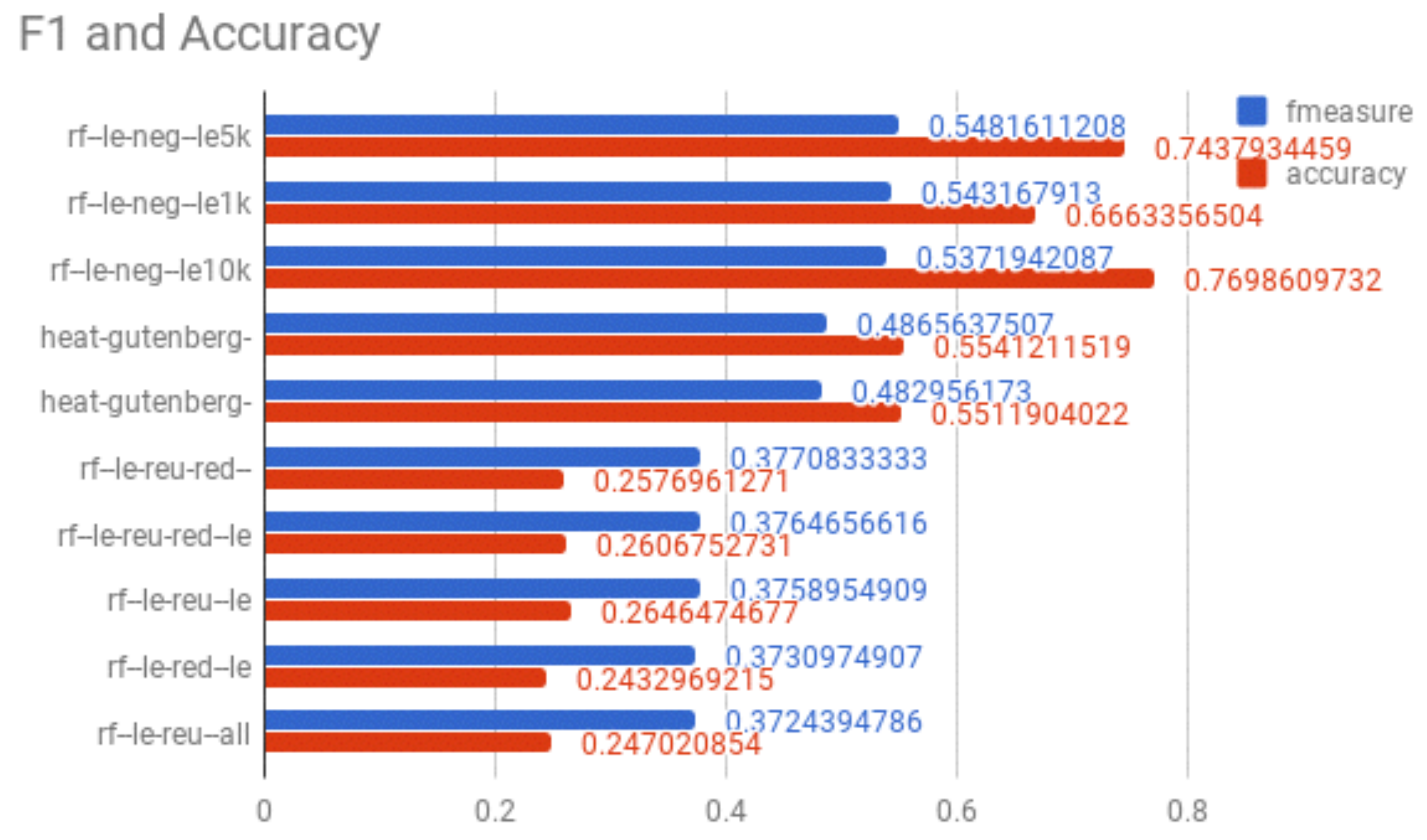
Open issues with past experiments

- The *Heat hypothesis* seems confirmed but **it is not enough**
  - **Music-Gut**, good recall, low precision and accuracy
  - **Music-Forest**, better precision but very low recall

- Evaluated on a benchmark (**B2**)
  - 17 books, 1098 experiences
  - A collection of bookmarks identifying LEs from the LED automatically
  - Assumed that the LED database contains all the Listening Experiences in these books
  - But that assumption was false

- **Music-Gut** v1 too specific POS tags (VVB, not V)





Progress

- We analysed false negatives and derive hints about how to improve: they were mostly valid LEs!!!
  - We evolved **B2** into a manually supervised Gold Standard (**B3**)
- Recomputed **Gutenberg-M** using a stronger POS abstraction
- Introduced a <u>Third</u> approach using Word Embeddings: **Music-Emb**
- Tried several approaches, including *Sentiment Heat*
- Performed a new set of experiments on the more accurate **B3**
- <u>Developed a tool for finding traces of LEs in texts</u>

Recap: where we are

Phase 1:

We hypothesised that LEs are a subset of the texts talking about music (*Heat hypothesis*). We developed a dictionary of musical terms (Music-Gut) using the Gutenberg English corpus (TFIDF). This dictionary represents well Listening Experiences (LE) in the database.

Phase 2:

We developed and compared a set of approaches (in several variants):

- **Music-Gut**: using the dictionary in combination with *features* of LE and evaluate it on a gold standard of LE and associated sources.
- **Music-Forest**: implementing a Random Forest Classifier trained on LED, RED, and Reuters
- **Sentiment Heat**: using a dictionary computed from SentiWordNet **(new!)**
- **Music-Emb**: using dictionaries of word-embeddings computed from Gutenberg English **(new!)**

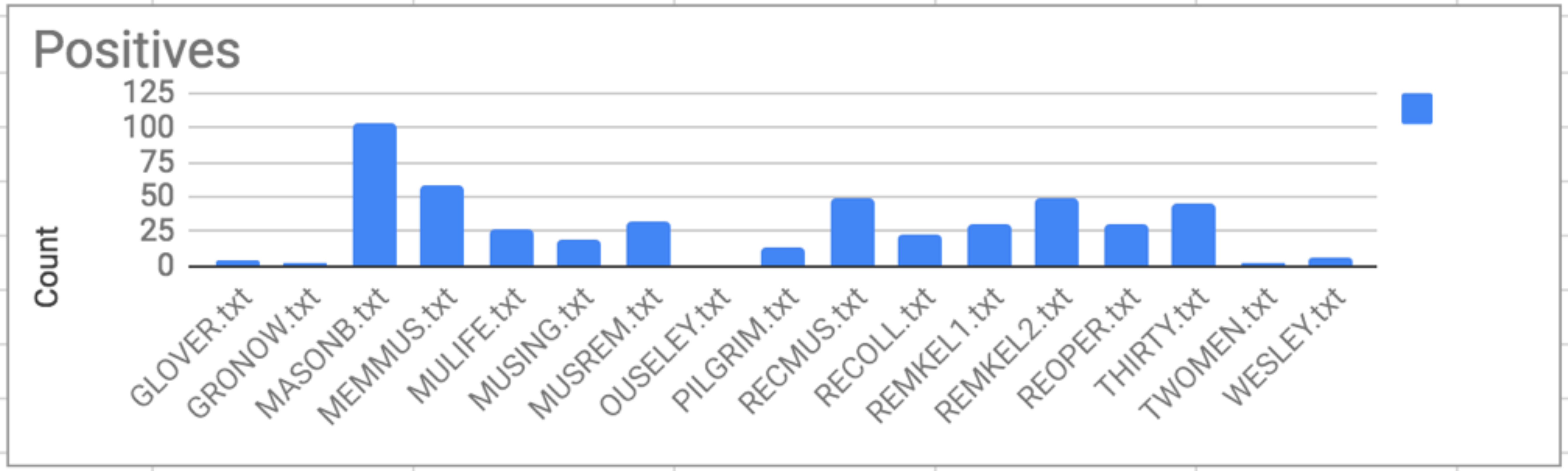
Phase 3 (in progress):

To develop a system that generate annotations of texts

To evaluate it with a user study.

Gold Standard development - B3

- We selected 500 positive samples from B2
- We manually selected a snippet of text of a similar length as negative example
- Resulting in 1000 text snippets of similar length (AVG ~125 words)



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

The building is in good taste and convenient, being in size and form much like one of our larger city churches. It is quite free, however, from all those appearances of finery, or attempts at display or show, which we sometimes see in our American churches, and which are always unbecoming; while, on the other hand, there is nothing of the rudeness or coarseness which is to be seen in some of the Swiss churches. It seats, probably, from 1,200 to 1,500 persons, and was, when we were present, quite full. The centre of the house, below, was occupied by women; and the outside or wall slips, by men. The galleries, on both sides, were occupied exclusively by men. The organ is large, extending nearly across the end of the house; one man (precentor) leads the singing, aided by some twenty girls and boys, whose voices could hardly be heard. **The organ was played in fine church style, with dignity, elevation, and firmness. It is certainly a great relief to hear these German organs (or many of them) played without the least attempt at showing off stops, or at that prettiness which seeks to please or tickle, without elegance or grandeur; and also entirely free from an ever-continued and sickening seesaw of the swell, thought to be so exquisitely fine by some organists in England and America.**

The morning was inauspicious. The clouds, dark and heavy, at once shut out the cheerful light of the sun, and poured out a cold, continuous rain, which was anything but musical in its appearance and influence. We left our lodgings about an hour before the time appointed for the commencement of the performance, and as we came to the street leading directly to the Hall, we found the sidewalks filled with people of all ages and descriptions, who, notwithstanding the mud and wet, had taken their stand to look into the carriages as they passed. The row of carriages at this time extended full a quarter of a mile from the Hall. The police regulations were excellent, and officers were stationed all around to see that they were observed. As the carriages were not permitted to break the line, and moved very slowly, a fine opportunity was afforded to those on the sidewalks to get a glimpse of the beautiful ladies and their elegant dresses, — and this was about all that the common people could get of the Festival

After a few minutes' recess, the competition in comic song followed. **Eight societies had entered their names as candidates, and sang successively for the prize picture. Some of them produced roars of laughter, and every one of them was received with more or less merriment and glee. One song (they were all part-songs) was truly good and exceedingly well done, but the others were commonplace,** or even low and frivolous; so much so, as to appear quite at variance with the idea that these festivals are designed for improvement in musical taste. It was somewhat sad too, to observe that **those songs which seemed to bring down humanity the nearest to mere animal being were the most admired, so that one in which imitations of the bleating of sheep and the cries of the domestic animals were introduced, called forth the loudest laugh and the most violent clapping of hands.**

It is no unimportant lesson for a teacher or **conductor** of **music** to learn what **music** is appropriate to the occasion, what comes fairly within the capacities of the **performers** and the under- standing of the **hearers**, what is suitable for children, — what for congregations, — what for choirs without **orchestra**, and for **choirs** with **orchestra**, etc. And how shall one learn these things and a thousand others? Answer. By the study of **music** under the **direction** of those who are competent to teach. How long will it take and what will it cost? Answer. Go to the members of other professions and ask them these questions, in relation to their own preparatory studies, multiply their answer by two, and the product will not deceive you.

Approaches

Forest-Gut Random Forest Annotator:

- based on the assumption that we can train a classifier abstracting features from LE texts, using LED Training Set (excluding books in the Gold Standard), and negative examples from RED and Reuters21578, in several configurations

Music-Gut Musical Heat Annotator:

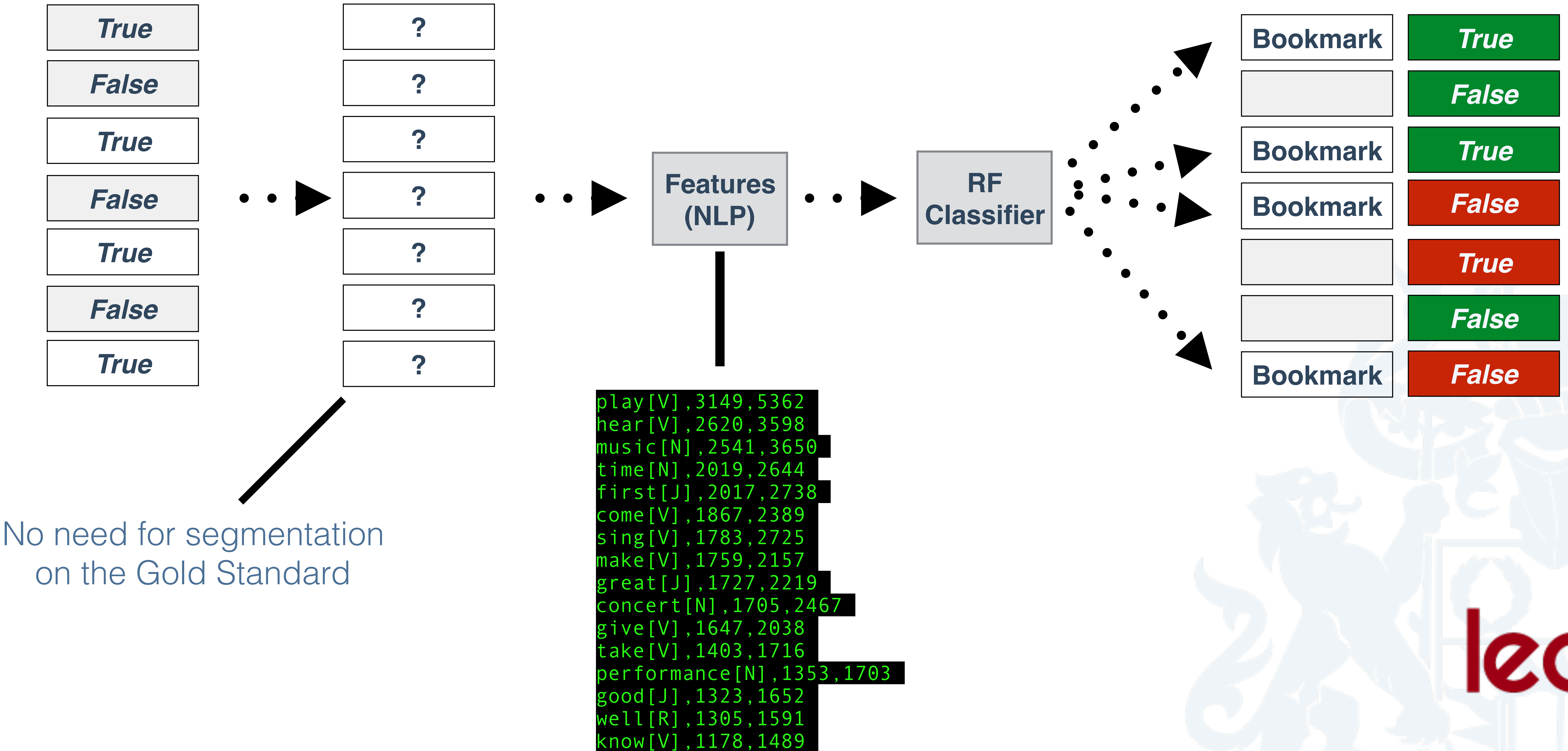
- based on the hypothesis that LEs are a subset of the texts talking about music.
- the more the text overlaps with **Gutenberg-M** the more it could be a LE

Music-Emb Musical Heat Annotator:

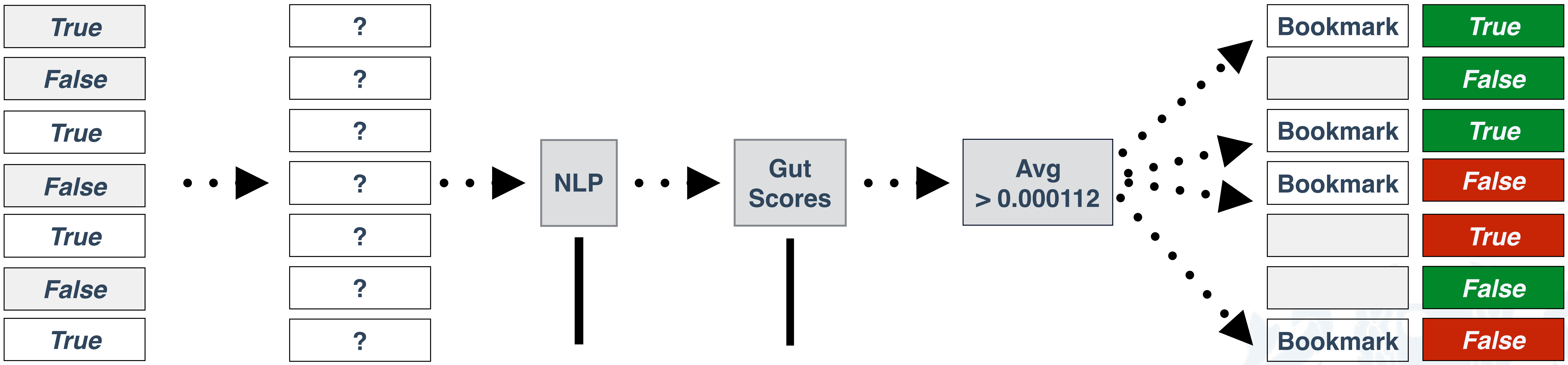
- based on the hypothesis that LEs are a subset of the texts talking about music.
- generated computing a distributed vector representations of words on the Gutenberg Corpus. Similar words are closer in the vector space. Same output as TFIDF but with a different technique (Word2Vec implementation of Spark ML library).

And others: Sentiment Analysis, combination of dictionaries, combinations of the above, …

Random Forest Annotator



Musical Heat Annotator



No need for segmentation
on the Gold Standard

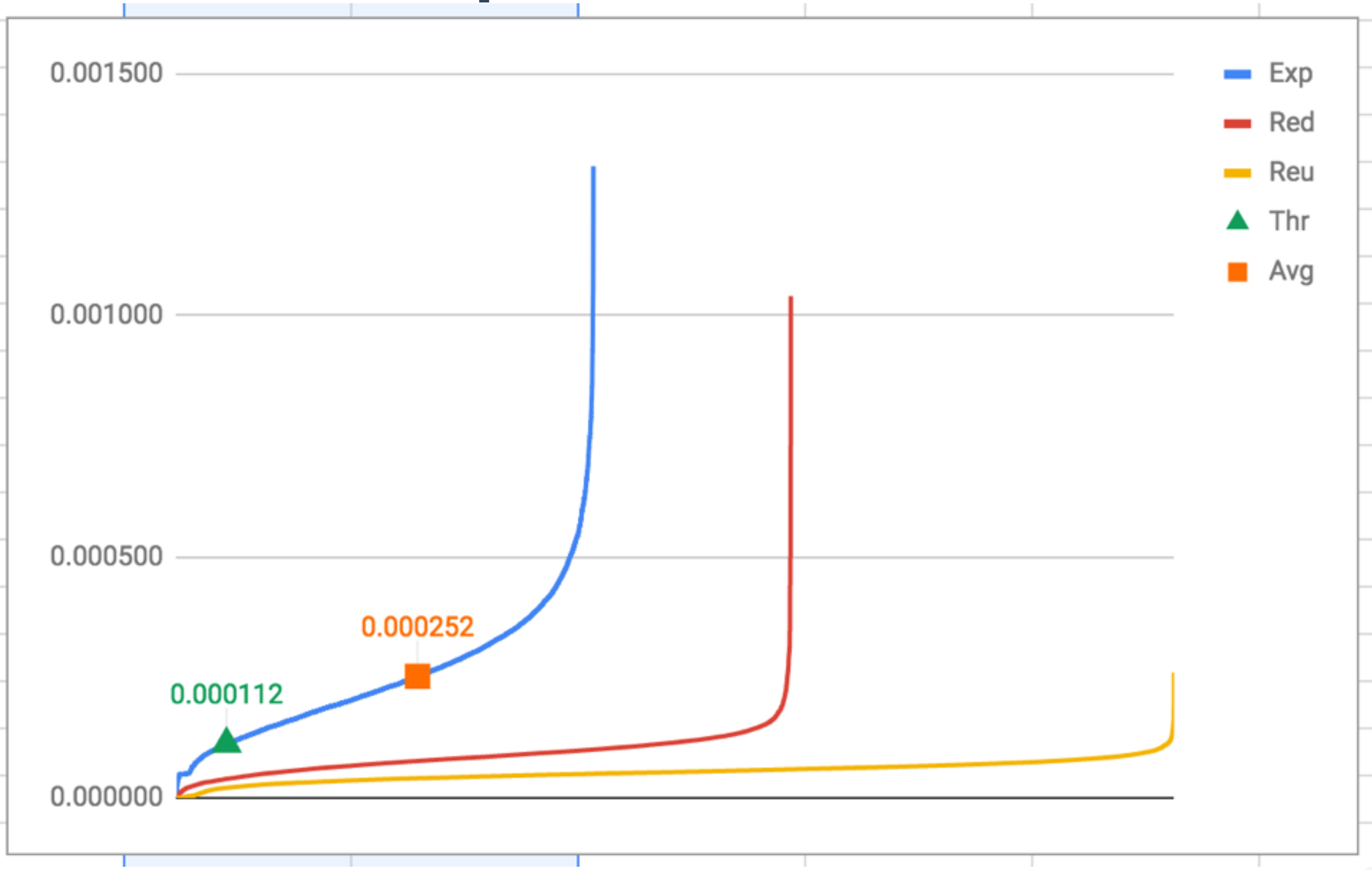
0	rontgen[N]	0.0007272755403226193
1	play[V]	0.000007967133189523013
2	Brahms[N]	0.0007256378255299395
3	symphony[N]	0.00071413210885995495
4	another[D]	0.000070088376068171141
5	musical[J]	0.00006717815731235641
6	take[V]	0.0000059785964542184945
7	always[R]	0.00005846276132915727
8	happen[V]	0.0000054224336619273
9	specially[R]	0.0005217816538462906
10	count[V]	0.0005135716029154898
11	something[N]	0.0004832913297756952
12	sort[N]	0.0004442636696952911
13	regard[V]	0.000004326928936343346
14	Fate[N]	0.000004142928952284239

Music-Gut: Development of a music dictionary

- Gutenberg corpus (english subset)
- We use NLP (StanfordNLP) to get a vector lemma+pos: music[n], play[v]*
- We calculated TF/IDF of each doc/term pair in Gutenberg
- We selected the terms in documents classified in the Music shelf and computed the AVG score

* Improved From Phase 1: V, N,… instead of NN, NNS, VBG,VBN, …

Music-Gut: Development of a music dictionary



We computed the **threshold** analysing the coverage of the vocabulary with the **LED Training Set** (excluding the books used in the Gold Standard)

Average heat value minus the standard deviation = **0.000112**

Music-Emb: Using word embeddings

- Gutenberg corpus (english subset)
- We use NLP to get a vector of tagged lemmas for each documents (StanfordNLP): music[n], play[v]
- We computed Word Vectors using the skip-ngram model, **learn vectors that are good at predicting its context in the same sentence** (Spark Word2Vec)
- We obtain a model that we can query for words that happen in the same **context**
- Querying this model we obtain a cluster of similar terms (with score)
- We generated 4 clusters: **listener[n], performer[n], event[n], music[n]**

Music-Emb: from word embeddings

music[n]
melody[n]
musical[j]
singer[n]
choral[j]
musical[n]
tune[n]
song[n]
singing[n]
flute[n]
violin[n]
improvisation[n]
orchestral[j]
cello[n]
orchestra[n]
serenade[n]
cadenza[n]
melodious[j]
accompaniment[n]
symphony[n]
lute[n]
harp[n]
playing[n]
vocal[n]
lilt[n]
orchestration[n]
love-song[n]
repertoire[n]

listen[v]
audibly[r]
tremulous[j]
echo[v]
repeat[v]
hoarse[j]
hushed[j]
awe-struck[j]
loud[r]
audible[j]
awed[j]
aloud[r]
louder[j]
softly[r]
awestruck[j]
breathlessly[r]
sonorously[r]
sing-song[j]
laughing[n]
tea-bell[n]
entranced[j]
plaintive[j]
purr[v]
burst[v]
inaudible[j]
drawling[j]
bugle-note[n]
breathe[v]

performance[n]
creation[n]
concert[n]
theatrical[j]
oratorio[n]
performer[n]
pantomimic[j]
ballet[j]
pantomime[n]
practically[r]
psychical[n]
musical[j]
entertainment[n]
theatricals[n]
dance[n]
troupe[n]
operatic[j]
orchestra[n]
vaudeville[n]
tannhauser[n]
exhibition[n]
rôles[n]
orchestral[j]
choral[j]
composer[n]
cantata[n]
repertory[n]
pianoforte[n]

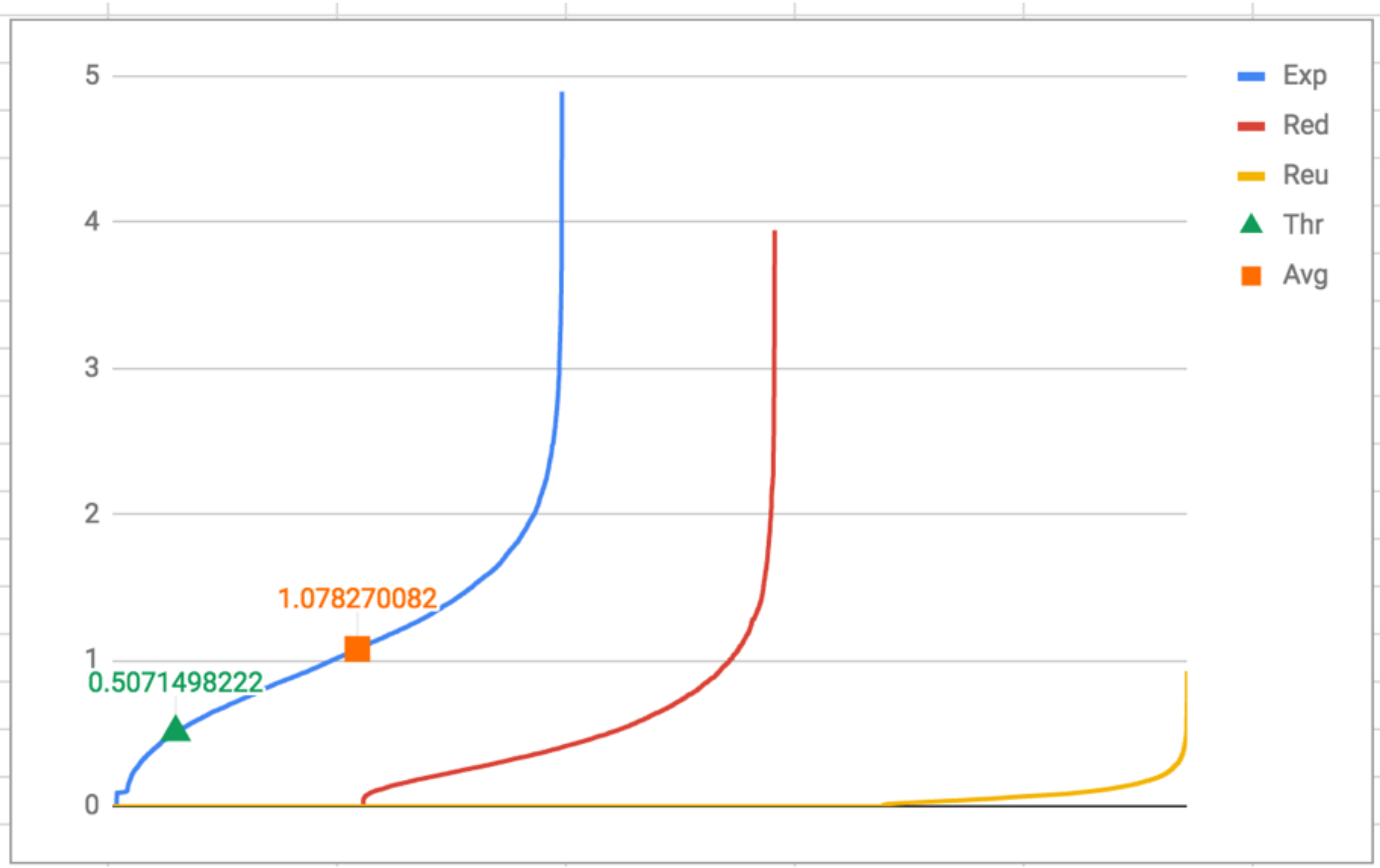
< embeddings

Gutenberg-M >

Beethoven[N]
vocal[J]
music[N]
Liszt[N]
Chopin[N]
composer[N]
Mozart[N]
musical[J]
Haydn[N]
piano[N]
aria[N]
fugue[N]
theme[N]
accent[N]
master[N]
Dickens[N]
resonance-chamber[N]
leading-tone[N]
florid[J]
sound[V]
score[N]
rondo[N]
sweet[J]
sense[N]
gesture[N]
hammer[N]
flow[N]
sorrow[N]

Music-Emb: Using word embeddings

music[n] cluster



- We computed the threshold analysing the coverage of the vocabulary with the **LED Training Set** (excluding the books used in the Gold Standard)
    - Average heat value minus the standard deviation = **0.480797**

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)

Experiments

- Music-Forest
- Music-Gut
- Music-Emb (music[n], event[n], performer[n], listener [n])
- Sentiment (extracting a dictionary from SentiWordNet)

- Composite (music[n] & event[n] & performer[n] & listener [n] & sentiment [n])
- Stacked (if method 1 =false, try method 2, …)

- *Gold Standard*: 1000 samples, 1/2 positives.

Experiment	Method	retrieved	positives	precision	recall	fmeasure	accuracy	error
rf--le-neg--gut10k-gs	FOREST	74	73	0.986486	0.146000	0.254355	0.572000	0.428000
rf--le-neg--gut5k-gs	FOREST	123	119	0.967480	0.238000	0.382022	0.615000	0.385000
rf--le-neg--le1k-gs	FOREST	202	189	0.935644	0.378000	0.538462	0.676000	0.324000
heat-sentiment-1-gs	SENTIMENT	907	457	0.503859	0.914000	0.649609	0.507000	0.493000
rf--le-reu-red--le10k-gs	FOREST	946	498	0.526427	0.996000	0.688797	0.550000	0.450000
heat-performer-1-gs	MUSIC-EMB	797	460	0.577164	0.920000	0.709329	0.623000	0.377000
heat-listener-1-gs	MUSIC-EMB	677	441	0.651403	0.882000	0.749363	0.705000	0.295000
led-components-1t-all-gs	COMPOSITE	500	411	0.851582	0.851582	0.700000	0.768386	0.789000
stacked-compo-forest-heat-4-gs	STACKED	766	486	0.634465	0.972000	0.767773	0.706000	0.294000
stacked-compo-heat-1-gs	STACKED	765	486	0.635294	0.972000	0.768379	0.707000	0.293000
stacked-forest-compo-heat-6-gs	STACKED	685	481	0.702190	0.962000	0.811814	0.777000	0.223000
stacked-compo-heat-8-gs	STACKED	684	481	0.703216	0.962000	0.812500	0.778000	0.222000
stacked-forest-heat-3-gs	STACKED	658	474	0.720365	0.948000	0.818653	0.790000	0.210000
heat-gutenberg2-000112-gs	MUSIC-GUT	657	474	0.721461	0.948000	0.819360	0.791000	0.209000
stacked-forest-music-9-gs	STACKED	555	446	0.803604	0.892000	0.845498	0.837000	0.163000
heat-music-1-gs	MUSIC-EMB	559	457	0.817531	0.914000	0.863078	0.855000	0.145000

FindLEr

HOME PROJECT DATABASE CONTACT

FindLEr

Detects traces of listening experiences in texts

Type a URL pointing to a plain text file.

https://ia801600.us.archive.org/33/items/in.ernet.dli.2015.260880/2015.2

DISCOVER!

Discussion

- The results on B3 are excellent: 86% F-Measure / 85% Accuracy
  - Probably comparable to human annotators …
- FindLEr:
  - Applies the Music-Emb method *as-is* as it is the best performing <u>in our experiments with B3</u>. B3 has is own biases…
  - The tool works differently with different books
  - Streaming the text…
- However, what actually <u>**is**</u> a listening experience?

An Ominous Fall. — I remember Count d'Orsay telling me that on the day previous to the appearance of the celebrated ordonnances, or decrees of July 27, 1830, which caused the Revolution and drove Charles X. from the throne, his sister, the Duchesse de G--- , niece by marriage to Prince- Polignac, and a violent Royalist, was seated at the piano, playing and singing with triumphant vigour, "La victoire est a nous," when suddenly the music-stool gave way, and the beautiful Duchess lay sprawling on the floor. D'Orsay, who was a Liberal, assured her, laughingly, that this fall in the midst of her Legitimist song was de tres mauvais augure, and a bad prognostic for the success of the party to which she belonged. He did not at the time believe his own prophecy, so firmly did the Bourbons appear to be established; but before the end of the month Charles X. had left France, and was followed by the fair Duchess and her husband, the most faithful friends and adherents of the fallen monarch, and as true to him in adversity as when he shone forth as one of the most powerful sovereigns of Europe.

<p>We were also present during public worship in churches at other places, as Leyden, and The Hague, but a description of one is a description of all. At Amsterdam and Rotterdam the same general style of church-singing prevails.</p>

<p>We did not learn much in relation to church music this day, either in the Moravian or Baptist Chapel.</p>

The only souvenirs that I have of that first summer are the little comedies brought out by Mr. English, and his melodrama of Rosina Meadows, which first saw light in that little theatre, and had a great run for several years. The two Chapman brothers were very clever, funny people, and were inimitable in the celebrated Mrs. Caudle's Curtain Lectures, written by Douglas Jerrold for Punch (to which he was a principal contributor), and prepared for the stage by some witty playwright. We had also the well-known Mrs. Drake and her daughter, — ^the latter a beautiful girl and a wonderful soubrette, who afterward married Harry Chapman. "Gentleman Fenno" was the leading man.

How to improve the tool

- The tool applies the method as-is
    - Parts of texts that are not content can be picked (e.g. ToC)
    - The threshold is fixed. Maybe a **slider** should help on customising the threshold on the actual book
    - Implement different methods and combinations?
- Support:
    - file uploads
    - sharing of annotation

Future work

- Try to compose the different approaches developing fine grained heuristics (methods to return a score instead of True/False)
- The methods implemented so far do not have any notion of Context
- Test with LE that do not talk about music
- Test with musical texts that are not LEs
- How to capture better the semantics of a LE: what ***is*** a Listening Experience?