PITCHFORK MUSIC REVIEWS
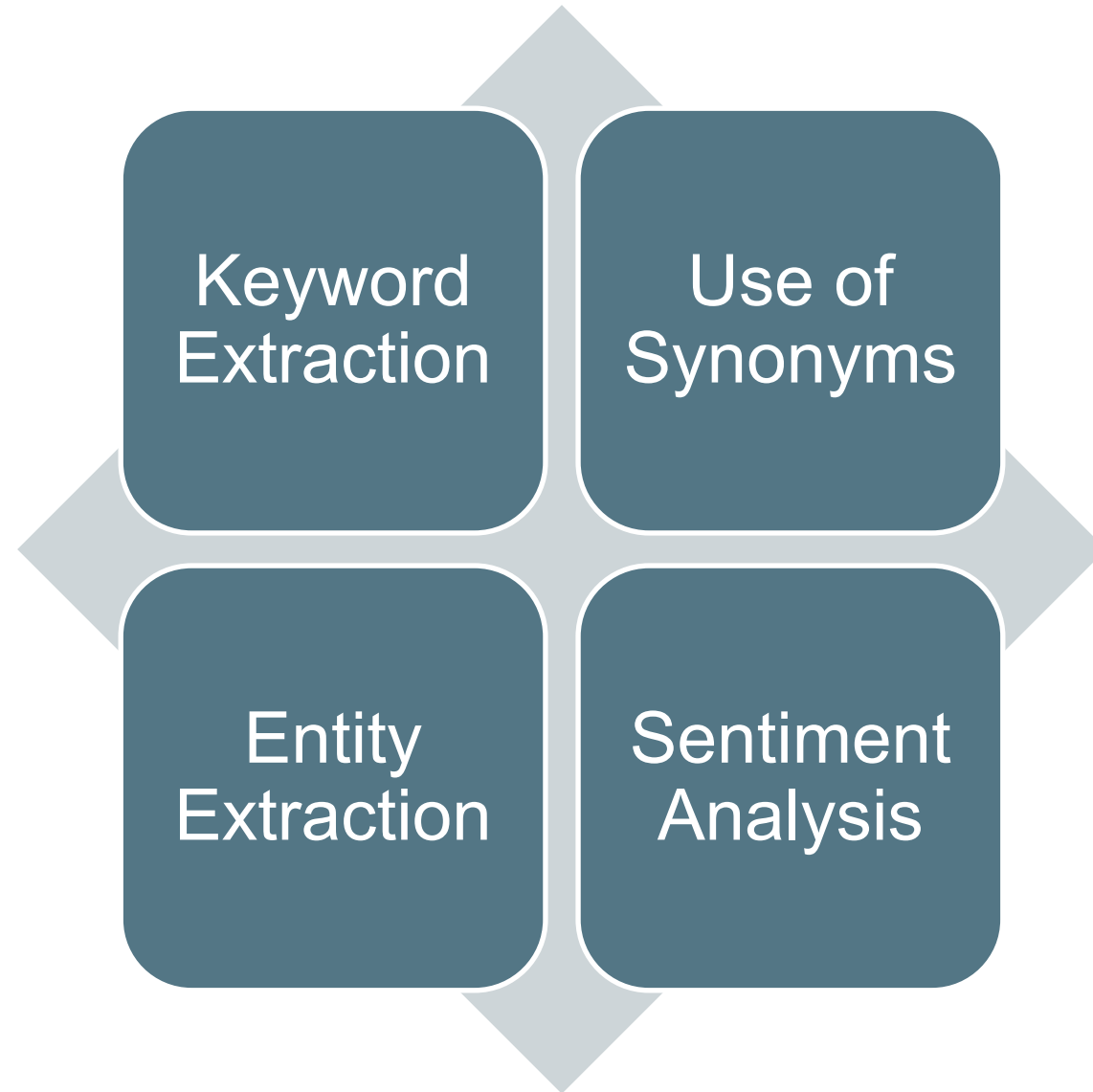## MILESTONE #3

INFORMATION PROCESSING AND RETRIEVAL

FABIO HUANG - UP201806829,

LISA SONCK - UP202202272

Used Systems - Schemas

- Basic Schema
- Extended Schema
- Synonym Schema

Used Systems - Schemas

- Basic Schema
- Extended Schema
- Synonym Schema

Type	Field	Indexed	MultiValued
standard_text	title	True	False
	artist	True	False
	author	True	False
	song release year	True	False
	genre	True	False
	record_label	True	False
single_words	keywords	True	True
	organizations	True	True
	places	True	True
	persons	True	True
	facilities	True	True
	law	True	True
	work_of_art	True	True
	events	True	True
	products	True	True
	norp	True	True
free_text	review_content	True	False
float	score	True	False
int	number of entities	True	False
string	reviewid	False	False

Used Systems - Schemas
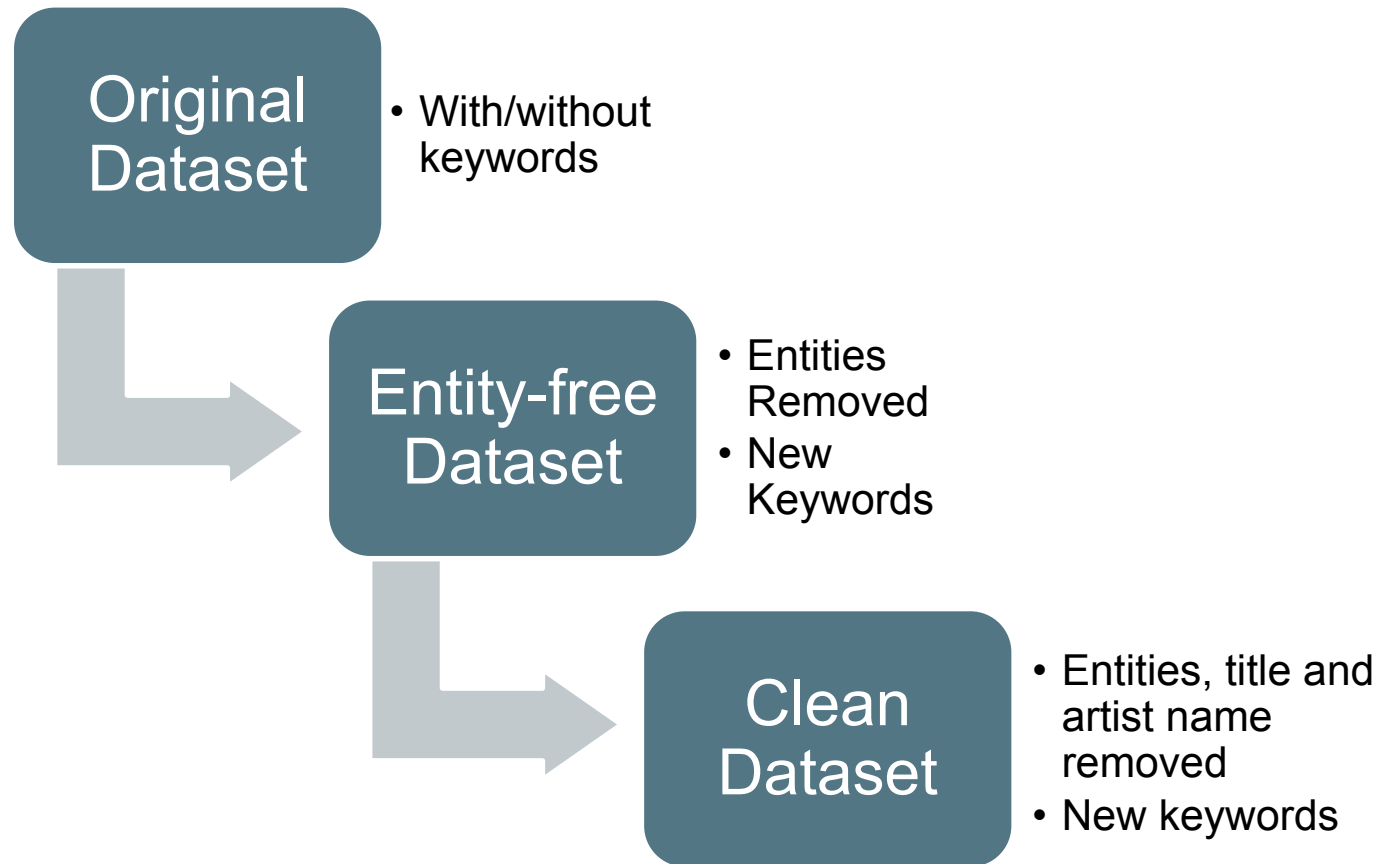
- Basic Schema
- Extended Schema
- Synonym Schema

standard_text	Tokenizer
	solr.ClassicTokenizerFactory
	Filters
	solr.ASCIIFoldingFilterFactory
	solr.LowerCaseFilterFactory
	solr.TrimFilterFactory
single_words	Tokenizer
	solr.KeywordTokenizerFactory
	Filters
	solr.ASCIIFoldingFilterFactory
	solr.LowerCaseFilterFactory
	solr.EnglishMinimalStemFilterFactory
	solr.EnglishPossessiveFilterFactory
	solr.HyphenatedWordsFilterFactory
	solr.KStemFilterFactory
	solr.TrimFilterFactory
free_text	Tokenizer
	solr.ClassicTokenizerFactory
	Filters
	solr.ASCIIFoldingFilterFactory
	solr.LowerCaseFilterFactory
	solr.EnglishMinimalStemFilterFactory
	solr.EnglishPossessiveFilterFactory
	solr.KStemFilterFactory
	solr.TrimFilterFactory
	solr.StopFilterFactory

Used Systems - Schemas

- Basic Schema
- Extended Schema
- Synonym Schema

standard_text	Tokenizer	
	solr.ClassicTokenizerFactory	
	Filters	
	solr.ASCIIFoldingFilterFactory	
	...	
single_words	Tokenizer	
	solr.KeywordTokenizerFactory	
	Filters	
	solr.ASCIIFoldingFilterFactory	
	...	
keywords_type	Tokenizer	
	solr.KeywordTokenizerFactory	
	Filters	
	solr.SynonymGraphFilterFactory	
	solr.ASCIIFoldingFilterFactory	
	solr.LowerCaseFilterFactory	
	solr.EnglishMinimalStemFilterFactory	
	solr.EnglishPossessiveFilterFactory	
	solr.HyphenatedWordsFilterFactory	
	solr.KStemFilterFactory	
	solr.TrimFilterFactory	
free_text	Tokenizer	
	solr.ClassicTokenizerFactory	
	Filters	
	solr.ASCIIFoldingFilterFactory	
	...	

Used Datasets

Original Dataset
- With/without keywords

Entity-free Dataset
- Entities Removed
- New Keywords

Clean Dataset
- Entities, title and artist name removed
- New keywords

Used Systems

	Basic Schema	Extended Schema	Synonyms Schema
Original Dataset without Keywords	system_1	system_2	
Original Dataset with Keywords		system_3	
Entity-free Dataset		system_4	
Cleaned Dataset		system_5	system_6

Information Needs

- Find albums that are perceived as boring.
- Find songs and albums that are ideal for studying.
- Find albums that contain a drum solo
- Find reviews that talk about Bob Dylan, but he is not the artist of the reviewed album.
- Find reviews of albums that include jazz and piano.
- Find reviews where the author praises the artist's voice.

QUERY 1:
DRUM SOLO

q:

   drum solo

q.op:

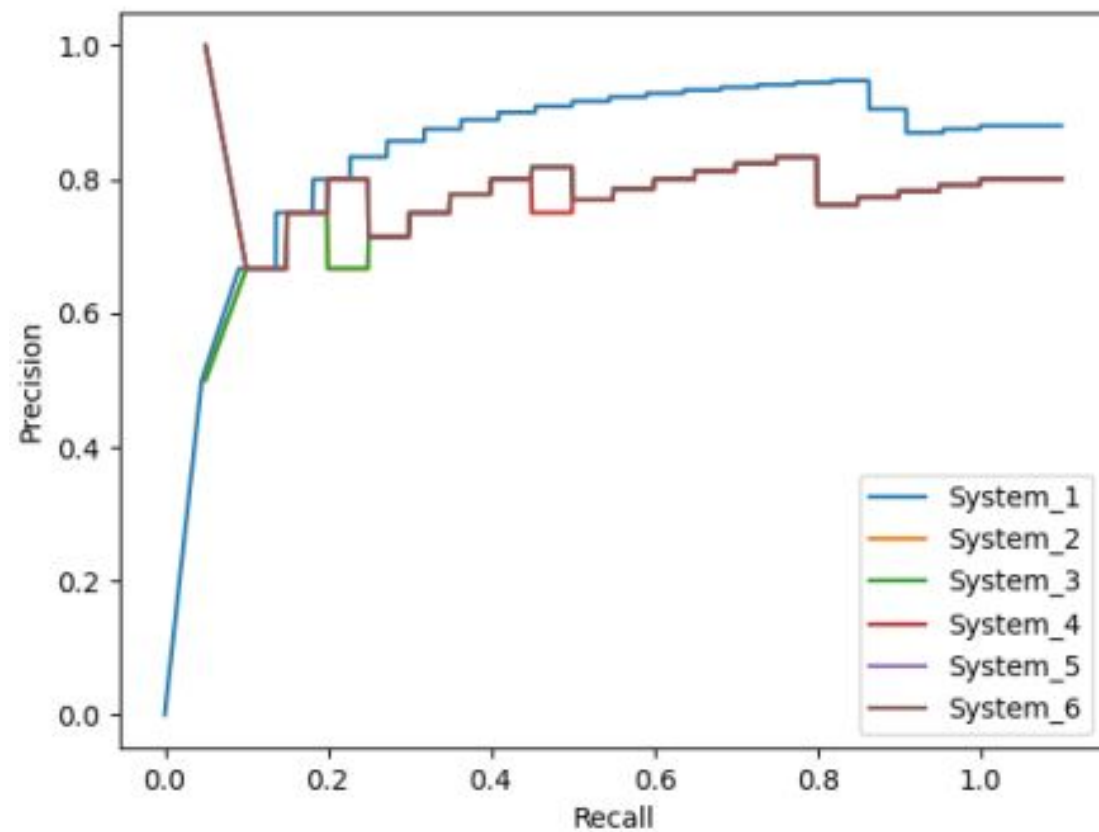   AND

qf:

   review_content^20
   keywords^50

pf:

   review_content^10

ps:

   5

Results

	Average Precision	Recall
system_1	0.831695	1.0
system_2	0.777176	1.0
system_3	0.777176	1.0
system_4	0.801482	1.0
system_5	0.804954	1.0
system_6	0.804954	1.0

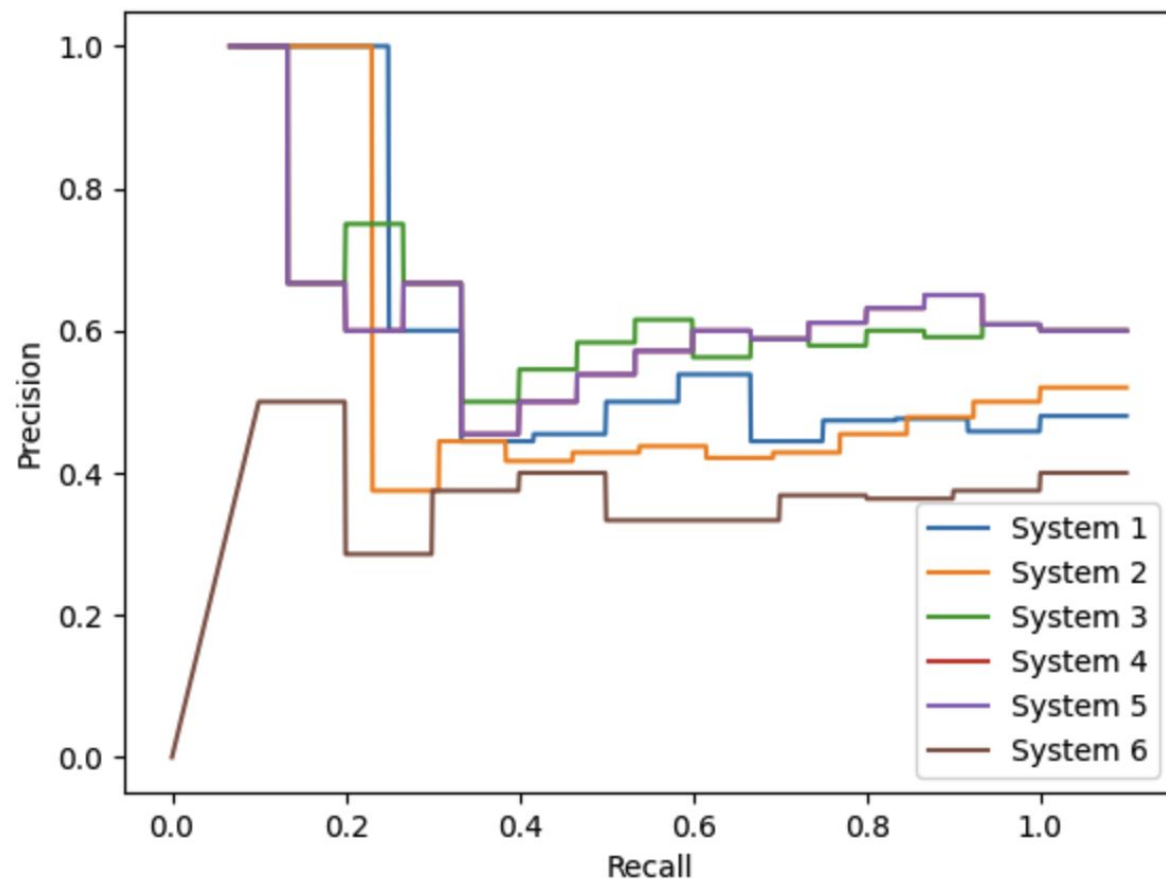|	P@05	P@10	P@15	P@20	P@25
system_1	0.8	0.9	0.933333	0.95	0.88
system_2	0.8	0.8	0.8	0.8	0.8
system_3	0.8	0.8	0.8	0.8	0.8
system_4	0.8	0.8	0.8	0.8	0.8
system_5	0.8	0.8	0.8	0.8	0.8
system_6	0.8	0.8	0.8	0.8	0.8

Results

QUERY 2: Praise to Voice

q:

review_content: ("I love voice" 5)^5
review_content: ("unique voice" 5)^5
review_content: "voice"^2
review_content: "vocal"^2
review_content: "I love"
review_content: "great"
review_content: "beautiful"
!review_content: "terrible"
!review_content: "bad"
!keywords: "terrible"
!keywords: "bad"
keywords: voice^4
keywords: unique^2
keywords: perfect^2

q.op:

OR

	P@05	P@10	P@15	P@20	P@25
system_1	0.6	0.5	0.533	0.5	0.48
system_2	0.6	0.5	0.467	0.45	0.52
system_3	0.8	0.5	0.6	0.6	0.6
system_4	0.6	0.5	0.6	0.65	0.6
system_5	0.6	0.5	0.6	0.65	0.6
system_6	0.4	0.4	0.33	0.4	0.4

	Average Precision	Recall
system_1	0.580492	1.0
system_2	0.537957	1.0
system_3	0.652069	1.0
system_4	0.641053	1.0
system_5	0.641053	1.0
system_6	0.383002	1.0

voice

Analyse Fieldname / FieldType: keywords ▾ ⑦ Schema Browser   ☑ Verbose Output   ▼ Analyse Values

KT	text	voice
	raw_bytes	[76 6f 69 63 65]
	start	0
	end	5
	positionLength	1
	type	word
	termFrequency	1
	position	1

SGF	text	articulation	juncture	junction	joint	join	articulatio
	raw_bytes	[61 72 74 69 63 75 6c 61 74 69 6f 6e]	[6a 75 6e 63 74 75 72 65]	[6a 75 6e 63 74 69 6f 6e]	[6a 6f 69 6e 74]	[6a 6f 69 6e]	[61 72 74 69 63 75 6c 61 74 69 69 6f]
	start	0	0	0	0	0	0
	end	5	5	5	5	5	5
	positionLength	15	15	15	15	15	15
	type	SYNONYM	SYNONYM	SYNONYM	SYNONYM	SYNONYM	SYNONYM
	termFrequency	1	1	1	1	1	1
	position	1	1	1	1	1	1

Results

	Mean Average Precision
system_1	0.693956
system_2	0.696333
system_3	0.703448
system_4	0.778395
system_5	0.779289
system_6	0.688257