Improving case-based search of clinical trials
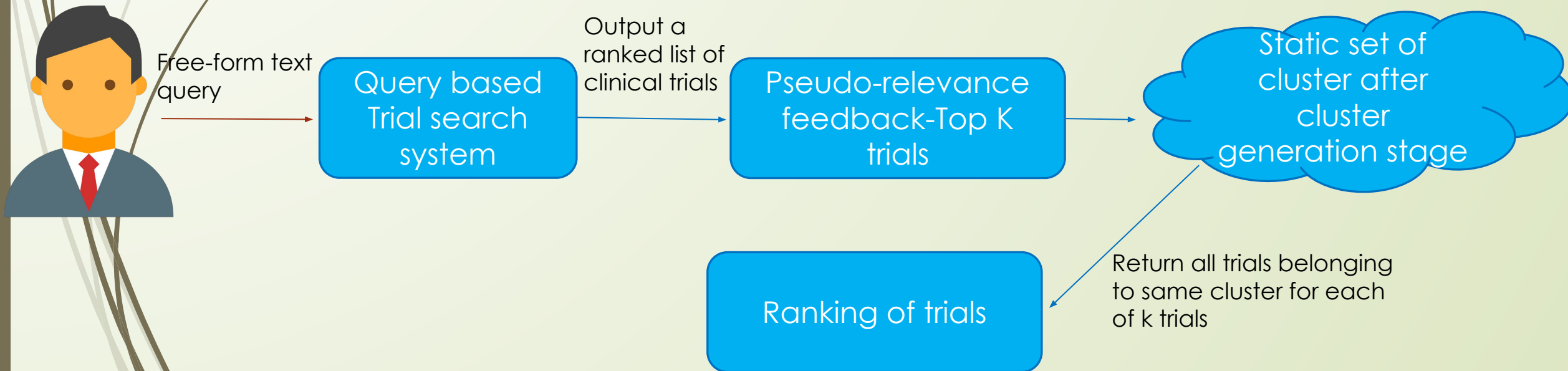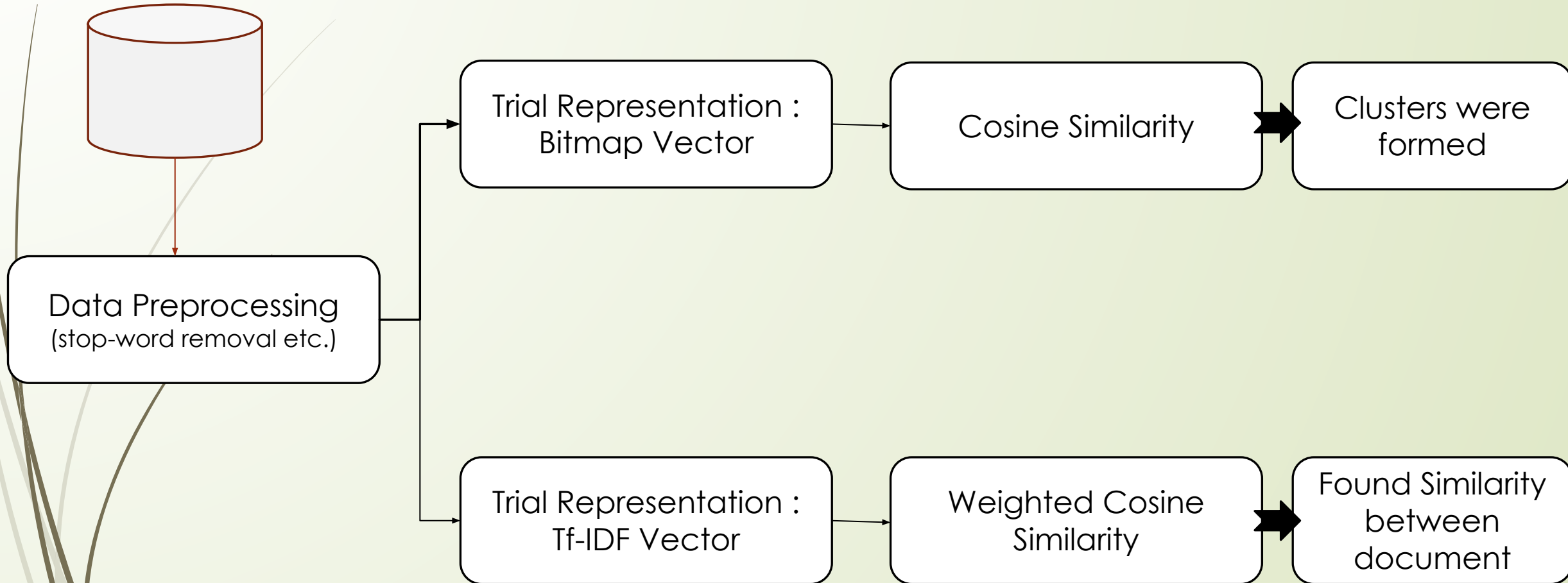
PROBLEM STATEMENT

CASE-BASED TRIAL SEARCH SYSTEM

HOW CLUSTERING FIT IN CASES-BASED SEARCH

1. DEVELOP OPTIMAL TRIAL REPRESENTATION WHICH BEST IMPROVES CLUSTERING QUALITY.

2. BETTER THE CLUSTERING QUALITY BETTER IS OUR CASE BASED SEARCH RETRIEVAL RESULTS

3. SINCE GENERAL USER HAS PROBLEM REFORMULATING EFFECTIVE QUERY , WE USE PSUEDO RELEVANCE OF TOP K TRIALS AS USER INPUT.

Free-form text query

Output a ranked list of clinical trials

Query based Trial search system

Pseudo-relevance feedback-Top K trials

Static set of cluster after cluster generation stage

Ranking of trials

Return all trials belonging to same cluster for each of k trials

Previous Work

Data Preprocessing
(stop-word removal etc.)

Trial Representation : Bitmap Vector

Cosine Similarity

Clusters were formed

Trial Representation : Tf-IDF Vector

Weighted Cosine Similarity

Found Similarity between document

SimScore(i, j) =

α * CosSim(incVec$_i$, incVec$_j$) +β * C

RESULTS- kNN CLUSTERING

SIMILARITY	TRIAL 1	TRIAL 2
0.9	NCT01977131	NCT01239862
0.8	NCT02066558	NCT01750723
0.9	NCT01971255	NCT01646138
0.7	NCT01813084	NCT01400906
0.9	NCT01711593	NCT01612936
0.7	NCT01651377	NCT01062945
0.8	NCT01489306	NCT01355016
0.9	NCT01405651	NCT01297582

ONLY A FEW CLUSTERS WERE FORMED AND ONLY THE DOCUMENTS WITH COSINE SIMILARITY SCORE CLOSE TO 1 WERE SIMILAR.

ISSUES FACED IN PREVIOUS METHODS
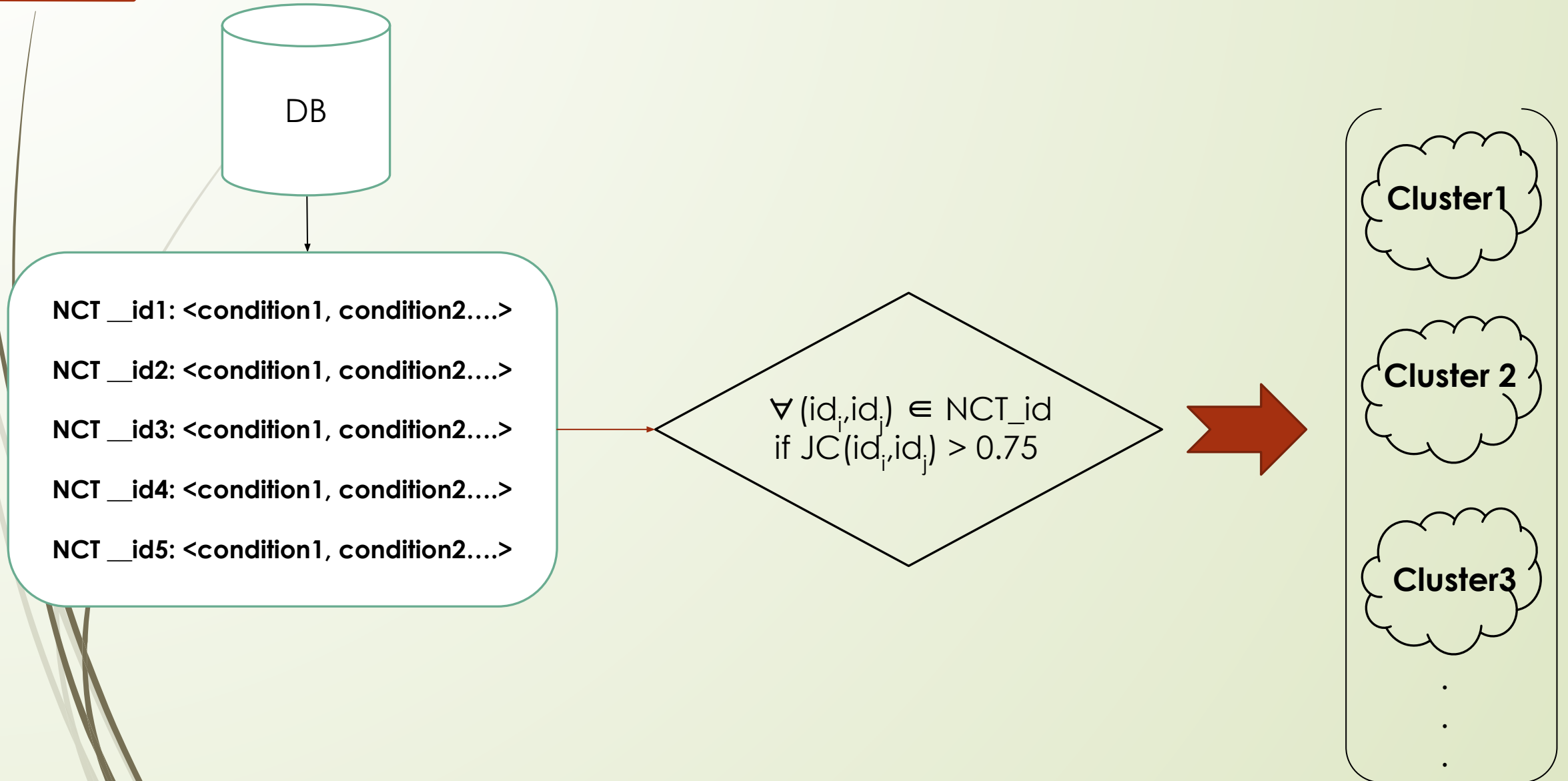
1. MEMORY OVERFLOW
2. SMALL CLUSTER

POST MIDSEM

METHODS USED

- JC SIMILARITY ON CONDITIONS LIST- kNN CLUSTERING

- JC SIMILARITY ON MESH THESAURUS

- COSINE SIMILARITY ON UMLS MODIFIED TRIAL

- DBSCAN CLUSTERING

- ABLATION ANALYSIS : K-MEANS CLUSTERING

JC SIMILARITY ON CONDITIONS LIST

DB

NCT __id1: <condition1, condition2....>

NCT __id2: <condition1, condition2....>

NCT __id3: <condition1, condition2....>

NCT __id4: <condition1, condition2....>

NCT __id5: <condition1, condition2....>

$$\forall\, (id_i, id_j) \in NCT\_id$$
$$\text{if } JC(id_i, id_j) > 0.75$$

Cluster1

Cluster 2

Cluster3

.
.
.
.

RESULTS

Safety, Pharmacokinetic and Preliminary Efficacy Study of AC0010MA in Advanced Non Small Cell Lung Cancer

⚠ The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Know the risks and potential benefits of clinical studies and talk to your health care provider before participating. Read our disclaimer for details.

ClinicalTrials.gov Identifier: NCT02448251

Recruitment Status ❶: Recruiting
First Posted ❶: May 19, 2015
Last Update Posted ❶: June 19, 2018

See **Contacts and Locations**

Risk Factors of Medistinal Metastasis in Endoscopic Staging of Lung Cancer

⚠ The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Know the risks and potential benefits of clinical studies and talk to your health care provider before participating. Read our disclaimer for details.

ClinicalTrials.gov Identifier: NCT02991924

Recruitment Status ❶: Recruiting
First Posted ❶: December 14, 2016
Last Update Posted ❶: March 30, 2018

See **Contacts and Locations**

Sponsor:

JC SIMILARITY ON CONDITIONS LIST

- **<u>RESULTS</u>**

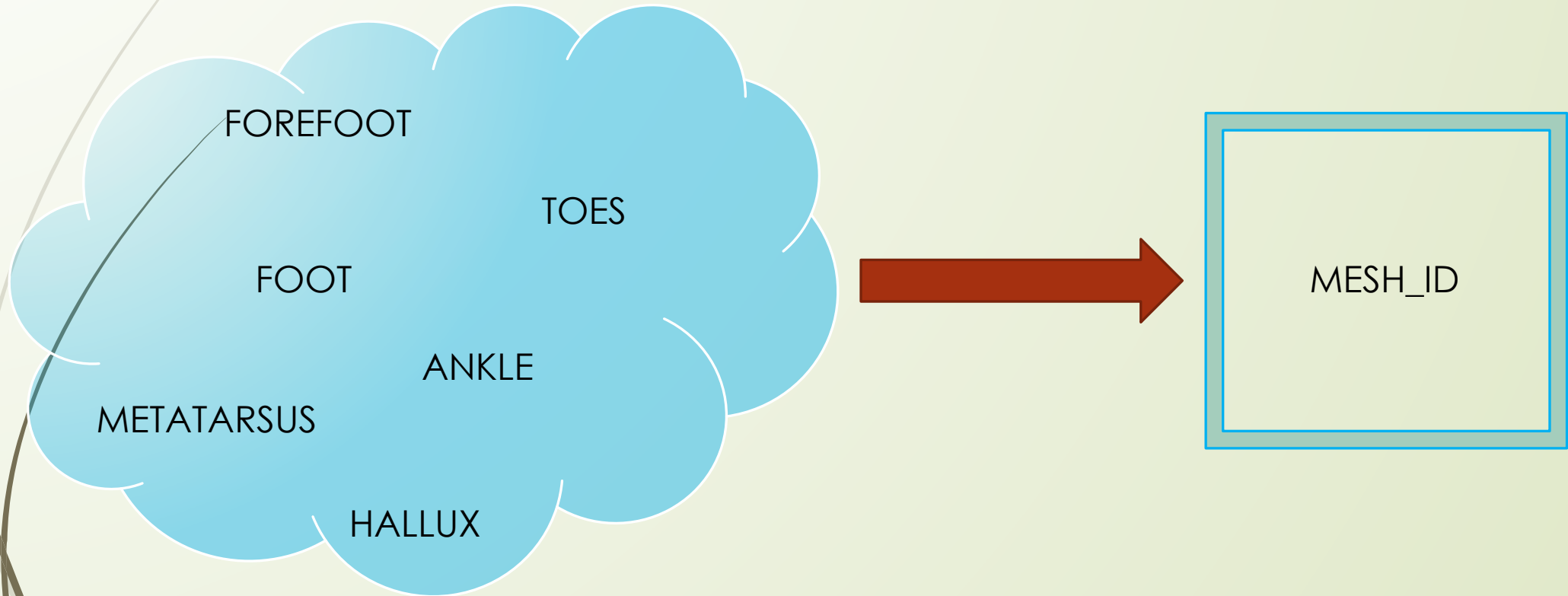 EVEN THE TRIALS WITH JC=1 WERE NOT SIMILAR IN SOME OF THE CASES.

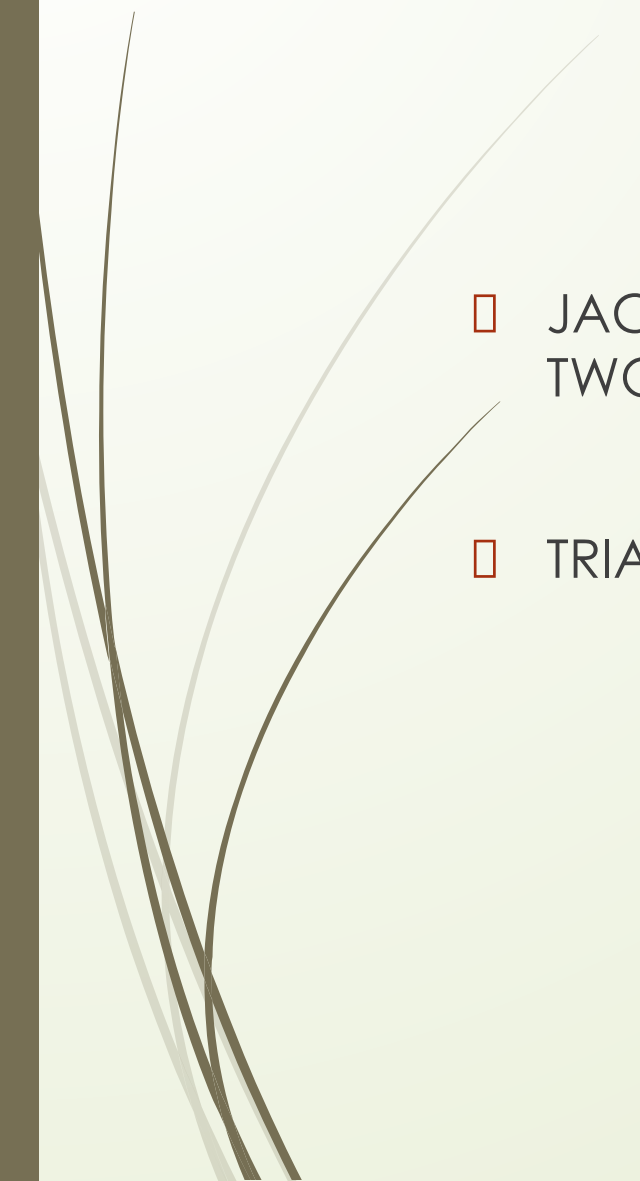 REASON : EACH TRIAL WAS JUST REPRESENTED BY 2-3 CONDITIONS

  (#features too low)

INTUTION FOR NEXT APPROACH

- DIFFERENT TERMS MAY CONVEY THE SAME MEANING IN MEDICAL FIELD

- HENCE WORDS IN TRIALS SHOULD BE COMPARED WITH ITS SYNONYMS.

- USED MeSH THESAURUS TO FIND SYNONYMS.

- MeSH (Medical Subject Headings) IS THE NLM CONTROLLED VOCABULARY THESAURUS USED FOR INDEXING ARTICLES FOR PubMed.

ftp://nlmpubs.nlm.nih.gov/online/mesh/

JC SIMILARITY ON MESH THESAURUS

FOREFOOT

TOES

FOOT

ANKLE

METATARSUS

HALLUX

MESH_ID

- JACCARD COEFFICIENT WAS USED TO MEASURE SIMILARITY BETWEEN TWO TRIALS.

- TRIALS WITH JC > 0.9 WERE SIMILAR.

Trial record **1 of 1** for: NCT03156231

Previous Study | Return to List | Next Study

Effect Of Acetazolamide On Altitude Related Illness In Patients With Respiratory Disease

⚠ The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Know the risks and potential benefits of clinical studies and talk to your health care provider before participating. Read our disclaimer for details.

ClinicalTrials.gov Identifier: NCT03156231

Recruitment Status ❶: Recruiting
First Posted ❶: May 17, 2017
Last Update Posted ❶: June 7, 2017

See Contacts and Locations

Sponsor:
University of Zurich

Collaborator:
National Center of Cardiology and Internal Medicine, Kyrgyzstan

Information provided by (Responsible Party):
University of Zurich

Trial record **1 of 1** for: NCT03173508

Previous Study | Return to List | Next Study

Effect of Acetazolamide on Right Ventricular Function at Rest in Patients With Respiratory Disease at Altitude

⚠ The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Know the risks and potential benefits of clinical studies and talk to your health care provider before participating. Read our disclaimer for details.

ClinicalTrials.gov Identifier: NCT03173508

Recruitment Status ❶: Recruiting
First Posted ❶: June 1, 2017
Last Update Posted ❶: June 7, 2017

See Contacts and Locations

Sponsor:
University of Zurich

Collaborator:
National Center of Cardiology and Internal Medicine, Kyrgyzstan

Information provided by (Responsible Party):
Silvia Ulrich Somaini, University of Zurich

Trial record **1 of 1** for: NCT03165890

Previous Study | Return to List | Next Study

Effect of Acetazolamide on Cognition in Patients With Respiratory Disease at Altitude

⚠ The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Know the risks and potential benefits of clinical studies and talk to your health care provider before participating. Read our disclaimer for details.

ClinicalTrials.gov Identifier: NCT03165

Recruitment Status ❶: Recruiting
First Posted ❶: May 24, 2017
Last Update Posted ❶: June 7, 2017

See Contacts and Locations

Sponsor:
University of Zurich

Collaborator:
National Center of Cardiology and Internal Medicine, Kyrgyzstan

Information provided by (Responsible Party):
University of Zurich

Trial record **1 of 1** for: NCT03167034

Previous Study | Return to List | Next Study

Effect of Acetazolamide on Right Ventricular Function During Exercise in Patients With Respiratory Disease at Altitude

⚠ The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Know the risks and potential benefits of clinical studies and talk to your health care provider before participating. Read our disclaimer for details.

ClinicalTrials.gov Identifier: NCT03167034

Recruitment Status ❶: Recruiting
First Posted ❶: May 25, 2017
Last Update Posted ❶: June 7, 2017

See Contacts and Locations

Sponsor:
University of Zurich

Collaborator:
National Center of Cardiology and Internal Medicine, Kyrgyzstan

Information provided by (Responsible Party):
Silvia Ulrich Somaini, University of Zurich
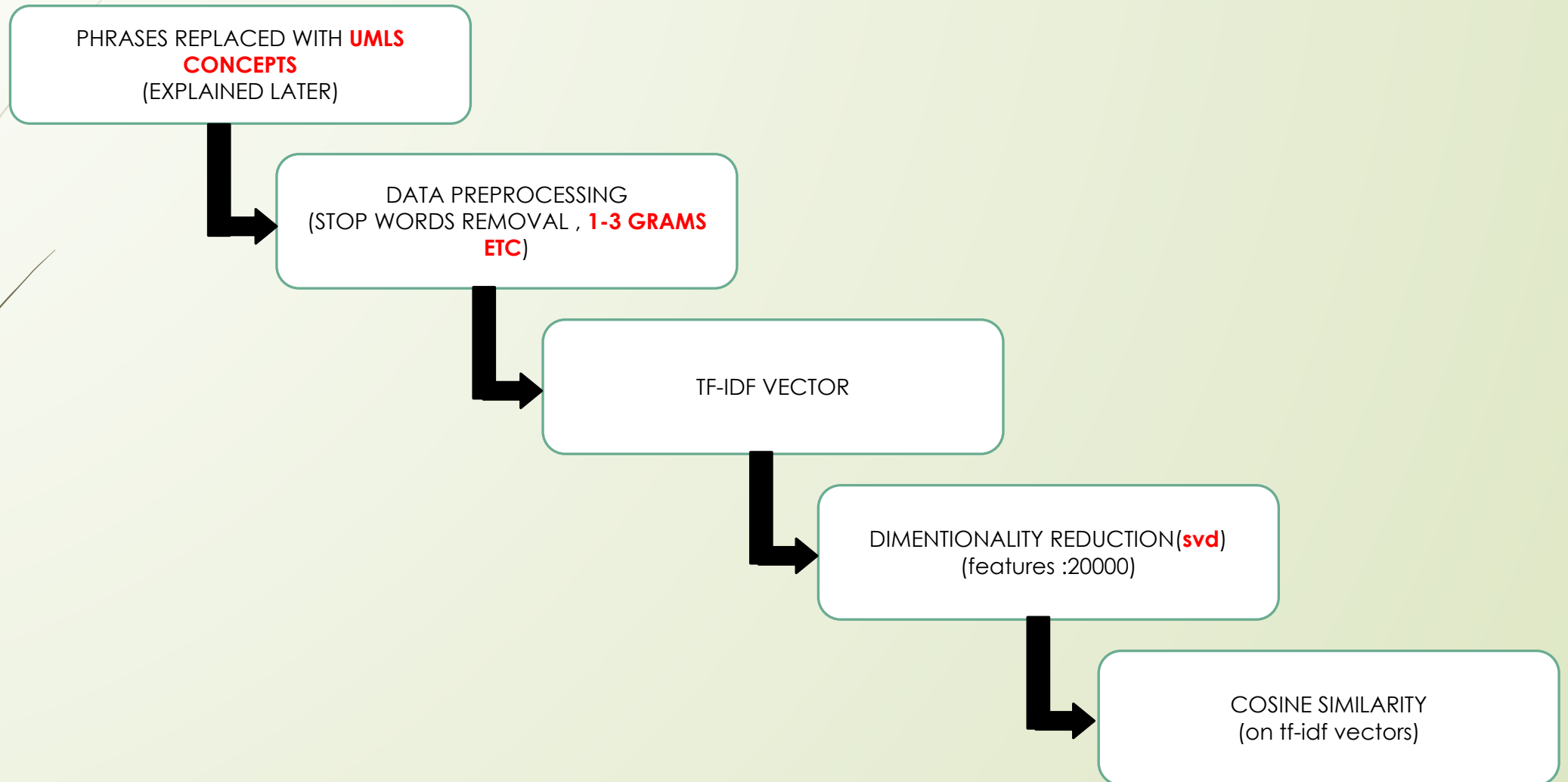
◻ RESULTS

Avg. Cluster Size :　9.72

No of outliers　　:　2016

No of clusters　　:　3894

Total trials　　　:　25536

◻ INTUITION :

FROM THE PREVIOUS 2 METHODS WE FOUND:

ONLY MEDICAL TERMS ARE NOT ENOUGH TO REPRESENT TRIAL.

HENCE IN NEXT APPROACH WE CONSIDERED REMAINING
.　TERMS AS WELL.

COSINE SIMILARITY

PHRASES REPLACED WITH **UMLS CONCEPTS**
(EXPLAINED LATER)

DATA PREPROCESSING
(STOP WORDS REMOVAL , **1-3 GRAMS ETC**)

TF-IDF VECTOR

DIMENTIONALITY REDUCTION(**svd**)
(features :20000)

COSINE SIMILARITY
(on tf-idf vectors)

RESULTS

```
Similarity values via TF_IDF Matrix on Inclusion Criteria using UMLS concepts

No of Trials Similar With Trial  NCT02301663  =  0
No of Trials Similar With Trial  NCT01356966  =  0
No of Trials Similar With Trial  NCT00300872  =  0
Trial:  NCT02220231  and Trial:  NCT02300649  with similarity  0.8127167481271855
Trial:  NCT02220231  and Trial:  NCT02185430  with similarity  0.6029134183778617
Trial:  NCT02220231  and Trial:  NCT02288780  with similarity  0.6329534159001815
Trial:  NCT02220231  and Trial:  NCT02237664  with similarity  0.5710271755055794
Trial:  NCT02220231  and Trial:  NCT02185378  with similarity  0.6742721277698331
Trial:  NCT02220231  and Trial:  NCT02365311  with similarity  0.9053575664597288
No of Trials Similar With Trial  NCT02220231  =  6
Trial:  NCT02695264  and Trial:  NCT02284217  with similarity  0.7251673036015771
Trial:  NCT02695264  and Trial:  NCT02772471  with similarity  0.773307957570136
```

Previous Method (Mesh)		Current Method	
Avg. Cluster Size	: 9.72	Avg. Cluster Size	: 14.28
No of outliers	: 2016	No of outliers	: 1378
No of clusters	: 3894	No of clusters	: 2214
Total trials	: 25536	Total trials	: 25536

Trial: NCT02220231 and Trial: NCT02365311 with similarity 0.9053575664597288

NIH U.S. National Library of Medicine
ClinicalTrials.gov

Find Studies ▾   About Studies ▾   Submit Studies ▾   Resources ▾   About Site ▾

Home >   Search Results >   Study Record Detail   ☐ Save this study

Trial record **1 of 1** for: NCT02220231

Previous Study | Return to List | Next Study

Echocardiographic Evaluation of the Change on Pulmonary Blood Flow and Cardiac Function Induced by Capnothorax During One Lung Ventilation

NIH U.S. National Library of Medicine
ClinicalTrials.gov

Find Studies ▾   About Studies ▾   Submit Studies ▾   Resources ▾   About Site ▾

Home >   Search Results >   Study Record Detail   ☐ Save this study

Trial record **1 of 1** for: NCT02365311

Previous Study | Return to List | Next Study

Echocardiographic Evaluation of the Change on Pulmonary Blood Flow and Cardiac Function During One-lung Ventilation

Reason:

Inclusion Criteria for NCT02220231

Inclusion Criteria:

1. Above 20 years of age.

2. American Society of Anesthesiologists (ASA) Physical Status I, II, III.

3. Thoracic surgical procedure (video-assisted thoracoscopic extended thymectomy)
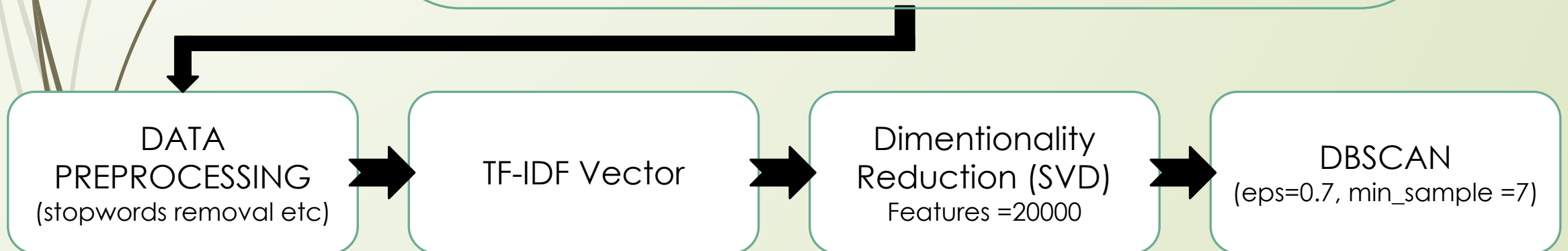
Inclusion Criteria for NCT02365311

Inclusion Criteria:

- Above 20 years of age.

- American Society of Anesthesiologists (ASA) Physical Status I and II

- Thoracic surgical procedure (video-assisted)

DBSCAN CLUSTERING

"Intrabronchial Route of Drug Administration"   :   "C1512914"

"Intracartilaginous Route of Drug Administration"   :   "C1512917"

"Intracisternal Route of Drug Administration"   :   "C1512924"

"Intracorneal Route of Drug Administration"   :   "C1512927"

"LDL-C results documented or reviewed - LDL-C 100-129 mg/dl"   :   "C2114378"

"LDL-C results documented or reviewed - LDL-C < 100 mg/dl"   :   "C2114376"

"primary repair of lateral collateral ligament of knee"   :   "C2091899"

UMLS CONCEPTS

| DATA PREPROCESSING (stopwords removal etc) | → | TF-IDF Vector | → | Dimentionality Reduction (SVD) Features =20000 | → | DBSCAN (eps=0.7, min_sample =7) |

TRIALS FROM SAME CLUSTER WERE NOT FOUND TO BE SAME

NIH U.S. National Library of Medicine

ClinicalTrials.gov

Find Studies ▾   About Studies ▾   Submit Studies ▾   Resources ▾   About Site ▾

Home >   Search Results >   Study Record Detail   ☐ Save this study

Trial record **1 of 1** for: NCT00156793

Previous Study | Return to List | Next Study

Phase 3 Imaging and Safety Study of AI-700 in Patients With Suspected Coronary Artery Disease Undergoing SPECT Imaging

NIH U.S. National Library of Medicine

ClinicalTrials.gov

Find Studies ▾   About Studies ▾   Submit Studies ▾   Resources ▾   About Site ▾

Home >   Search Results >   Study Record Detail   ☐ Save this study

Trial record **1 of 1** for: NCT00939237

Previous Study | Return to List | Next Study

Trial of Lycopene/Ateronon for Secondary Prevention of Coronary Heart Disease

Results

- **(ON 25536 TRIALS):**

  NO OF CLUSTERS FORMED = 20

  AVERAGE SIZE OF CLUSTER = 28

  MOST OF THE TRIALS WERE FOUND TO BE IN SAME CLUSTER

  NO OF OUTLIERS = 23447

Conclusion:

IT SUPPORTS THE ESTABLISHED IDEA THAT DBSCAN DOES NOT WORK WELL WITH HIGH DIMENSIONAL DATA AND REDUCING THE NUMBER OF DIMENSIONS CAN'T REPRESENT THE TRIAL PROPERLY.

ABLATION ANALYSIS -
K MEANS CLUSTERING

- MEDICAL TERMS WERE REPLACED WITH UMLS CONCEPTS.

- TF-IDF VECTOR WAS THEN CREATED TO REPRESENT THE TRIAL.

- K MEANS CLUSTERING WAS APPLIED WITH NO OF CLUSTERS = 100

- VECTOR SIZE = 20000

- RESULTS:

  - For k = 100:
    - No of clusters   : 100
    - Avg Cluster Size   : 189.2
    - No of Trials   :   25536
    - No outliers

Ablation analysis

Silhouette coefficient vs #Features
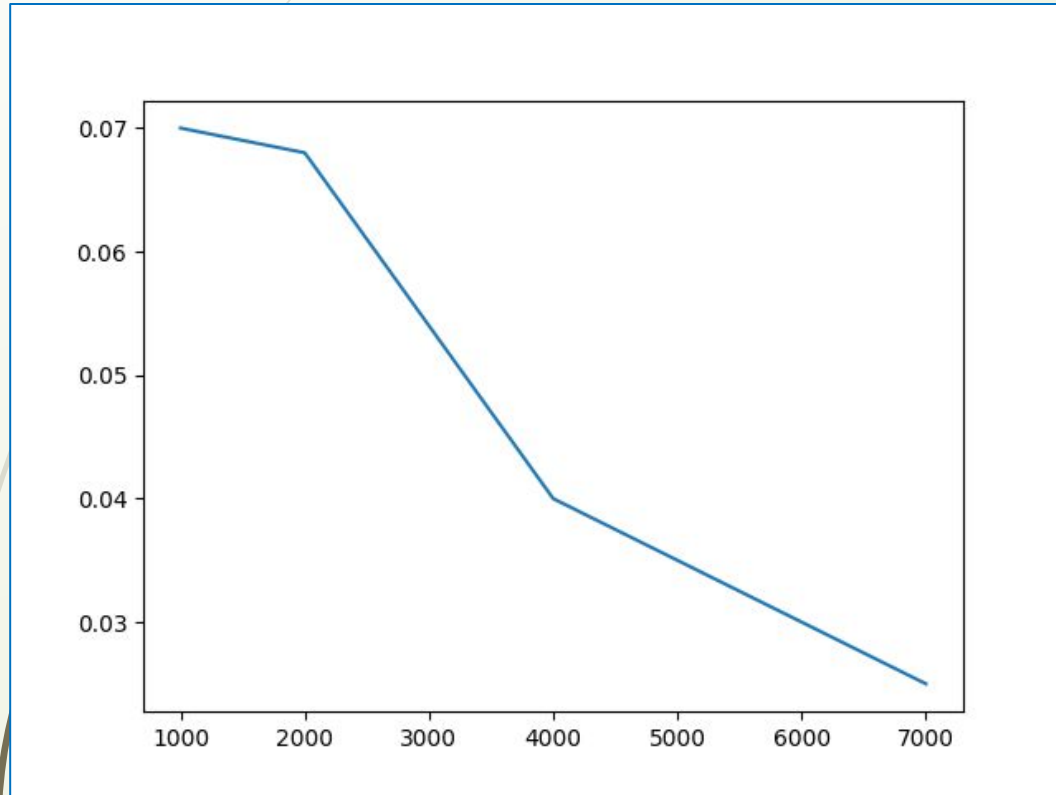
More the no of features, poorer the results.

Silhouette coefficient vs #Clusters
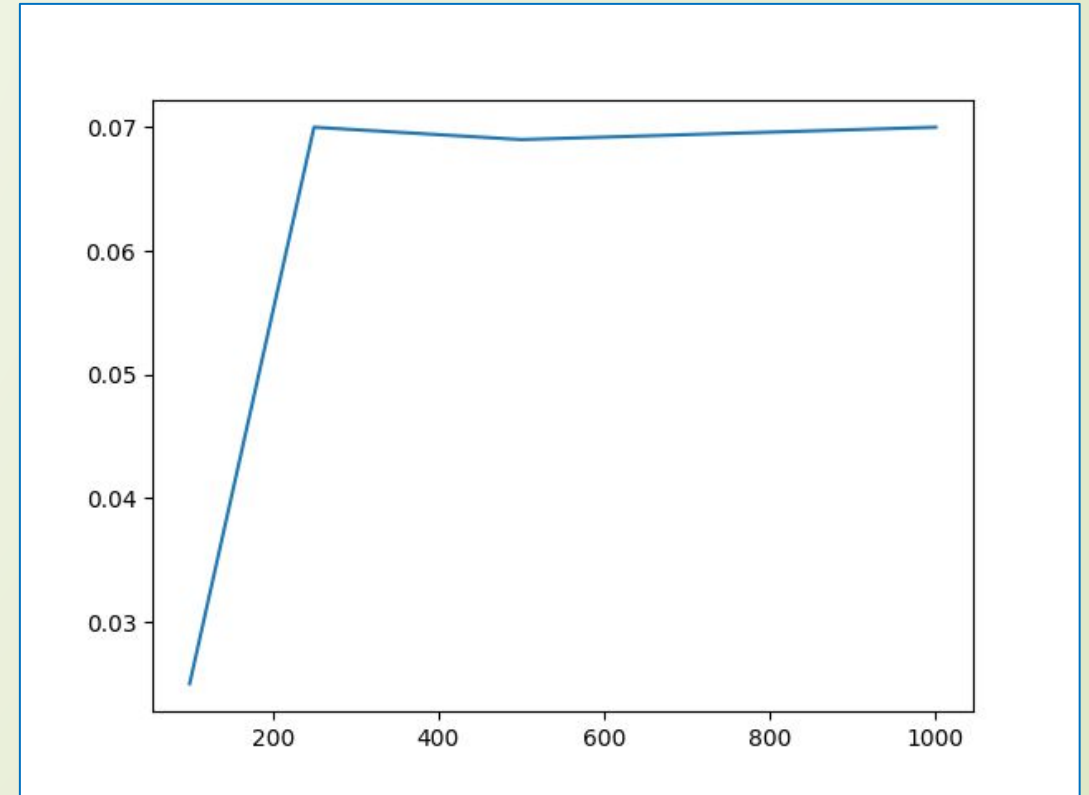
More the no of clusters, better the results.

Panel 1 (Top Left)

☐ Save this s

Trial record **1 of 1** for: NCT02573506

Previous Study | Return to List | Next Study

Study of Split-course Chemoradiotherapy For Bulky Local Advanced None-small Cell Lung Cancer

ClinicalTrials.gov Identifier: NCT02573506

The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Read our disclaimer for details.

Recruitment Status ❶: Active, not recruiting
First Posted ❶: October 9, 2015
Last Update Posted ❶: September 18, 2018

or:
 Yat-sen University

ation provided by (Responsible Party):
 Liu, Sun Yat-sen University

Details | Tabular View | No Results Posted | Disclaimer   ▣ How to Read a Study Record

Panel 2 (Top Right)

Trial record **1 of 1** for: NCT03006575

Previous Study | Return to List | Next Study

Study of Split-course Chemoradiotherapy For Postoperative Locoregional Recurrence Of Non-small Cell Lung Cancer

ClinicalTrials.gov Identifier: NCT03006575

The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Know the risks and potential benefits of clinical studies and talk to your health care provider before participating. Read our disclaimer for details.

Recruitment Status ❶: Recruiting
First Posted ❶: December 30, 2016
Last Update Posted ❶: September 26, 2018

See Contacts and Locations

or:
 Yat-sen University

ation provided by (Responsible Party):
 Liu, Sun Yat-sen University

Panel 3 (Bottom Left)

☐ Save this study

Trial record **1 of 1** for: NCT01977209

Previous Study | Return to List | Next Study

Gossypol Combined With Docetaxel and Cisplatin Scheme in Advanced Non Small-cell Lung Cancers With APE1 High Expression (GTCA)

ClinicalTrials.gov Identifier: NCT01977209

The safety and scientific validity of this study is the responsibility of the study sponsor and investigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Read our disclaimer for details.

Recruitment Status ❶: Unknown
Verified October 2013 by Dong Wang, Third Military Medical University.
Recruitment status was: Recruiting
First Posted ❶: November 6, 2013
Last Update Posted ❶: November 6, 2013

 Military Medical University

ation provided by (Responsible Party):
 Wang, Third Military Medical University

Panel 4 (Bottom Right)

☐ Sav

Trial record **1 of 1** for: NCT03135977

Previous Study | Return to List | Next Study

Apatinib and Etoposide as the Maintenance Therapy in Extensive-stage Small Cell Lung Cancer After First-line Chemotherapy

ClinicalTrials.gov Identifier: NCT03135977

he safety and scientific validity of this study is the responsibility of the study sponsor and nvestigators. Listing a study does not mean it has been evaluated by the U.S. Federal Government. Know the risks and potential benefits of clinical studies and talk to your health are provider before participating. Read our disclaimer for details.

Recruitment Status ❶: Recruiting
First Posted ❶: May 2, 2017
Last Update Posted ❶: December 7, 2017

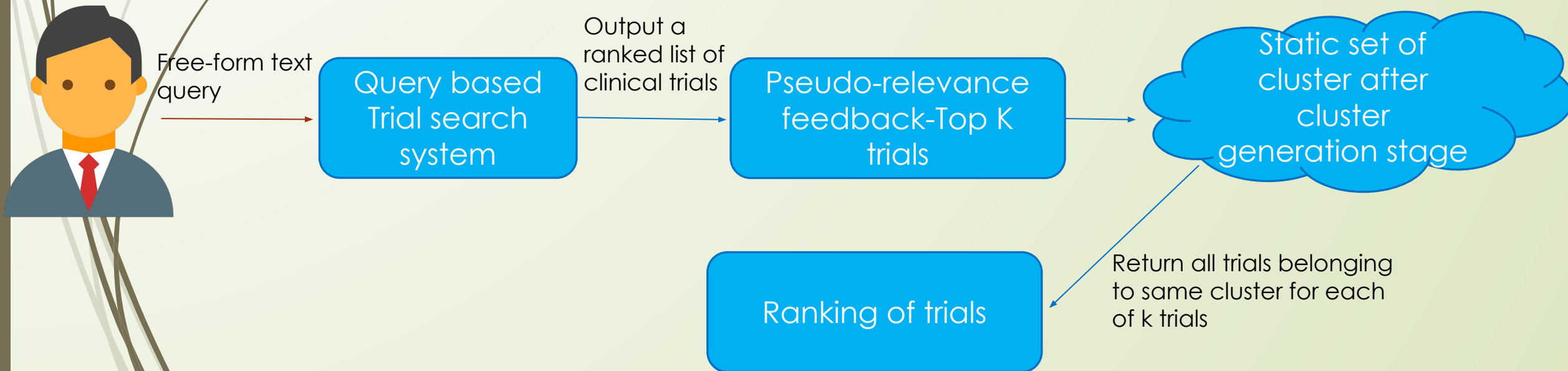See Contacts and Locations

n Cancer Hospital and Research Institute

on provided by (Responsible Party):
 ong cai, Sichuan Cancer Hospital and Research Institute

CASE-BASED TRIAL SEARCH SYSTEM

HOW CLUSTERING FIT IN CASES-BASED SEARCH

1. DEVELOP OPTIMAL TRIAL REPRESENTATION WHICH BEST IMPROVES CLUSTERING QUALITY.

2. BETTER THE CLUSTERING QUALITY BETTER IS OUR CASE BASED SEARCH RETRIEVAL RESULTS

3. SINCE GENERAL USER HAS PROBLEM REFORMULATING EFFECTIVE QUERY , WE USE PSUEDO RELEVANCE OF TOP K TRIALS AS USER INPUT.

Free-form text query

Output a ranked list of clinical trials

Query based Trial search system

Pseudo-relevance feedback-Top K trials

Static set of cluster after cluster generation stage

Return all trials belonging to same cluster for each of k trials

Ranking of trials

THANK YOU