INF FAKULTÄT FÜR INFORMATIK

How do patients talk about Tinnitus

Date: 28.03.2017

Supervised by:
Prof. Myra Spiliopoulou

Paras Multani
Rajeev Motwani
Kantharaju C Narayanappa
Prakruthi Shivanna

Contents

Tinnitus

•Tinnitus is a problem of hearing sounds when no external sound is present.

•Cause – It may be caused due to ear infections, age-related hearing loss, exposure to loud noise, earwax blockage or ear bones change.

• Symptoms – hearing loss.

•Diagnosis – Usually based on person‘s description, audiogram and neurological exam, imaging test, psychological assessment of patients with depression, anxiety, stress.

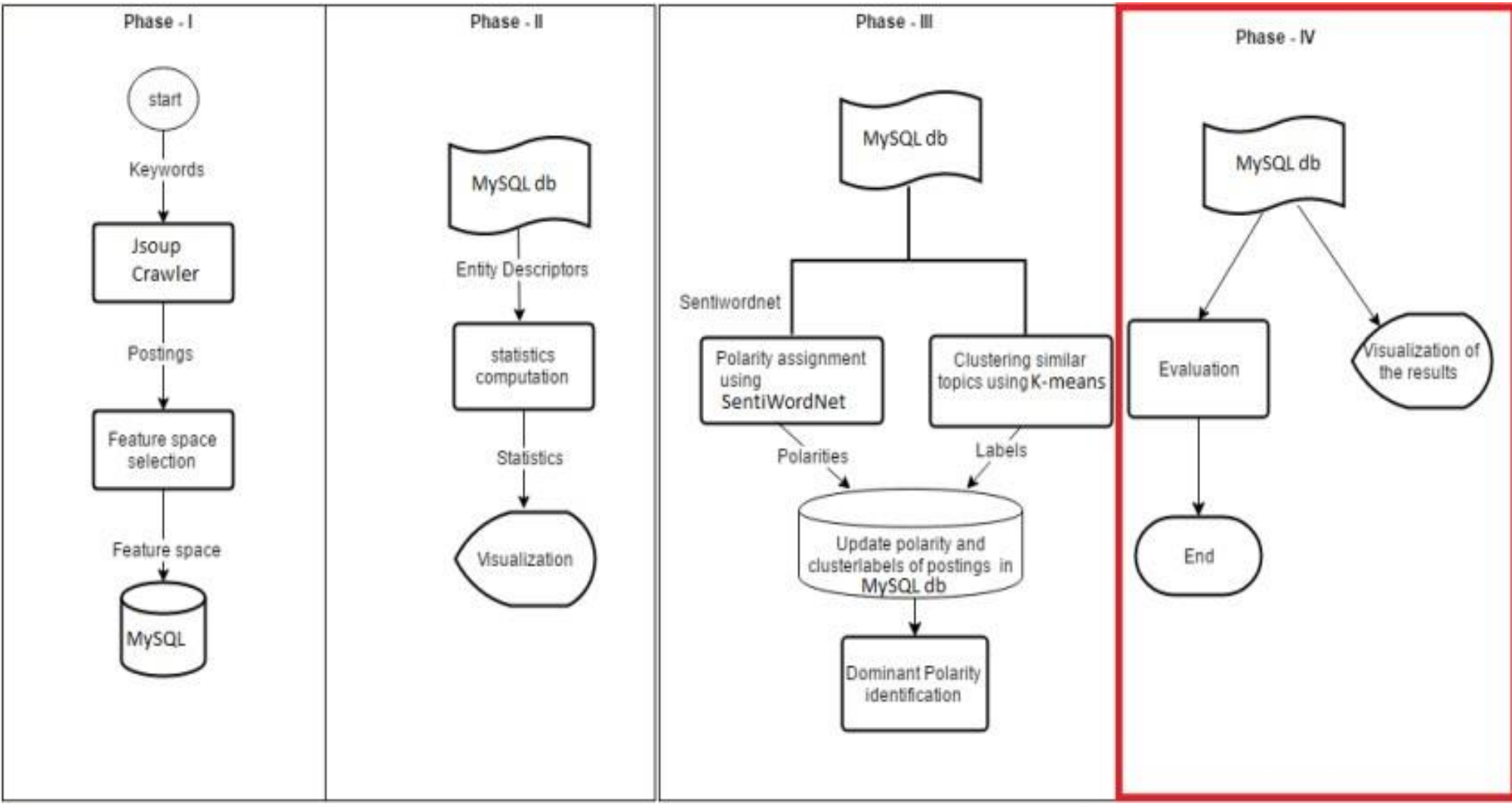
• As of 2014,there were no medications effective for tinnitus.

Motivation

•Patients post their problems about tinnitus on tinnitustalk.com. The posts include their problems, symptoms and their knowledge about the treatment of tinnitus.

•Often, providing a treatment for a particular symptom seems difficult. But,if we identify the topics in these posts, classify them and build a model, then it would help medical experts in making better decisions regarding the treatment of tinnitus.

Architecture



Web Crawling

- Implementation of a web crawler using Jsoup.
- Introduce Yourself and Treatments as Seed urls.
- Extraction of tags - User, Title, Publish Date and Description.
- Storing the data in MySQL.
- Note : Crawling has been performed for initial posts by patients and the replies have not been considered while performing any activity.

Data Schema

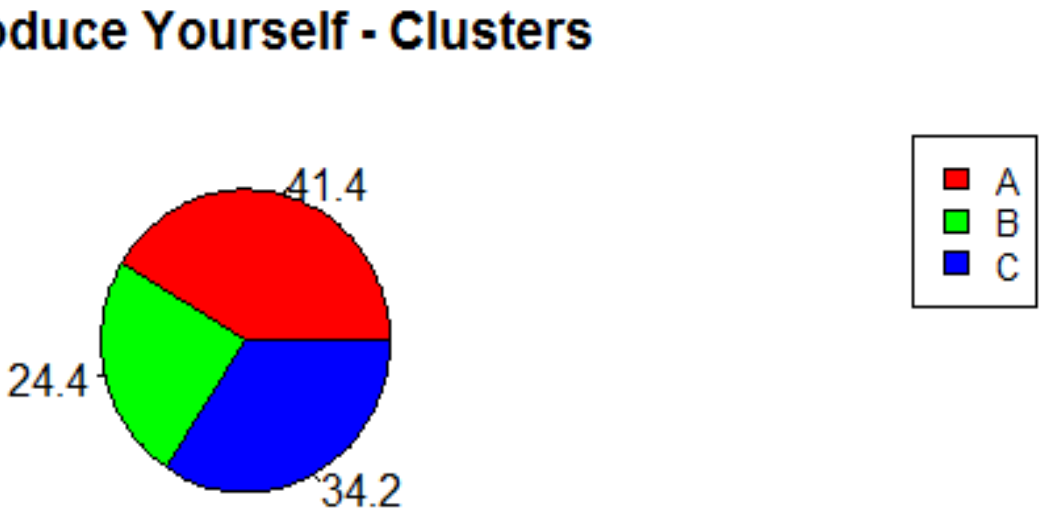
userid	title	publishd	description
.bill	Hello	Feb 1...	Had a case of flu Then ears felt plugged up Then the ringing started sto...
1MW	SSNHL Tinnitus Hyperacusis	Nov ...	Hello Some day after afternoon sleep when i wake up noticed my ear rin...
1Regcabguy$$	Newbie Here	Jan 3...	Hello New tinnitus sufferer here Injured my hearing recently with firear...
2131e	New to This How Loud Is Your Tinni...	Sep 1...	Hi everyone Im so grateful to have found all of you guys here on this fo...
2Fight	Aneurysm Survivors Story — Onset...	Jun 2...	I just had my 2nd Anniversary of Aneurysm rupture in the right PICA of ...
2lo	37 Years Old Just Got Diagnosed wi...	Aug ...	Hi all New to this situation Not a happy camper I guess you could say 37...
2LoudnHear	Its 2LoudnHear	Aug ...	Im treading water in a sea of sound
2Nine	A New Tinnitus Kid in Town	Sep 1...	Hello there my name is Pavel and I'm a musician I play the guitar not in a...
400runner	AntiInflammatory Induced	Nov ...	Hi Everyone I had a back injury at work for which I was prescribed the a...

Posts Clustering

- Data cleaning using NLP tasks.
- Term document matrix creation using tf-idf vector weight scheme.
- Normalizing euclidean distance and performing Kmeans clustering on the matrix.
- Developed using text mining package “tm” in R.

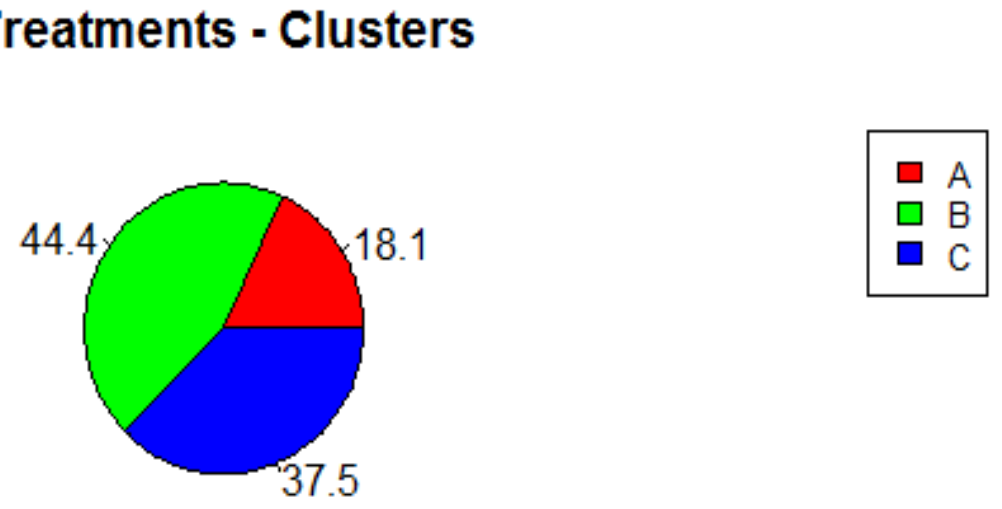
Posts Clustering

Total Posts=2485 → A=1026, B=606, C=853



Posts Clustering

Total Posts=72 → A=13, B=32, C=27



Text Classification

•Implementation of a simple classifier in Java using SentiWordNet.

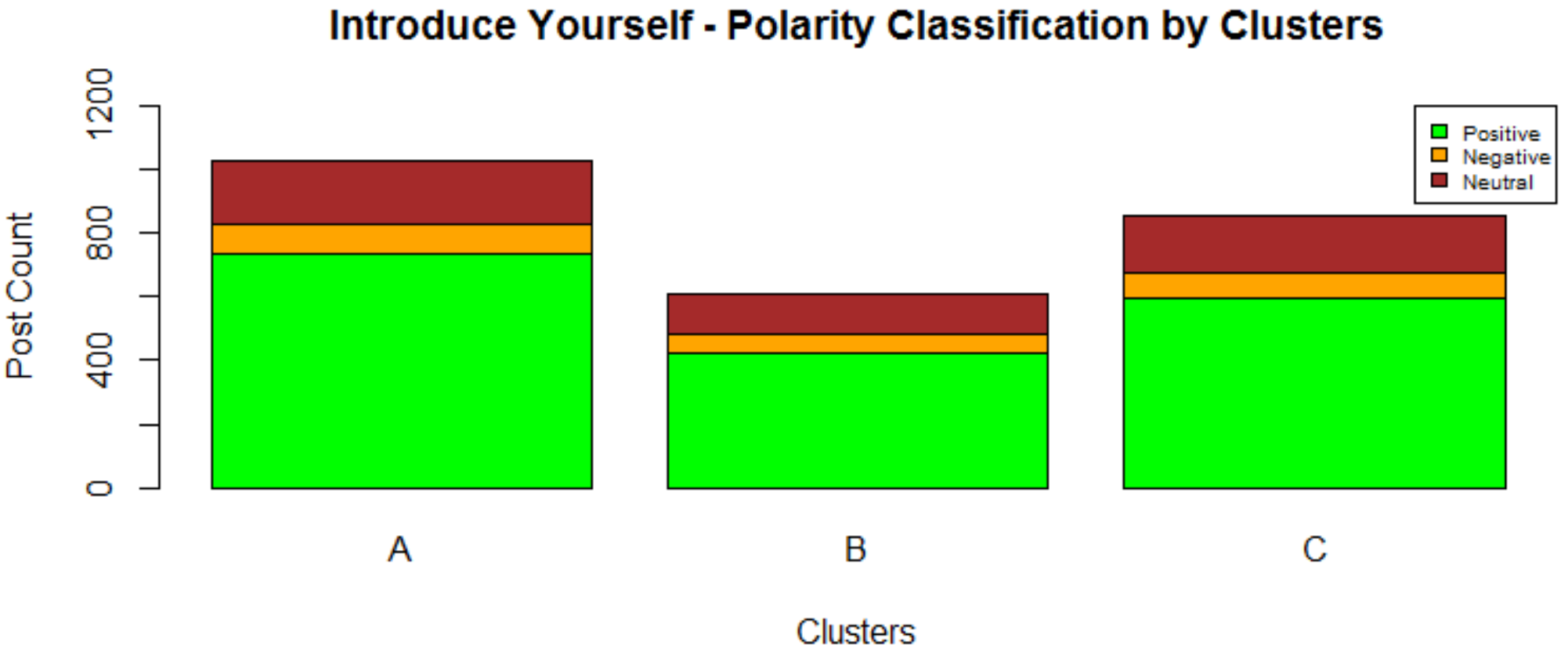
• Assigning a post to either positive, negative or neutral class.

• Polarity scale1*: -1 <= p <=1 => Neutral Post
p > 1 => Positive Post
p < -1 => Negative Post

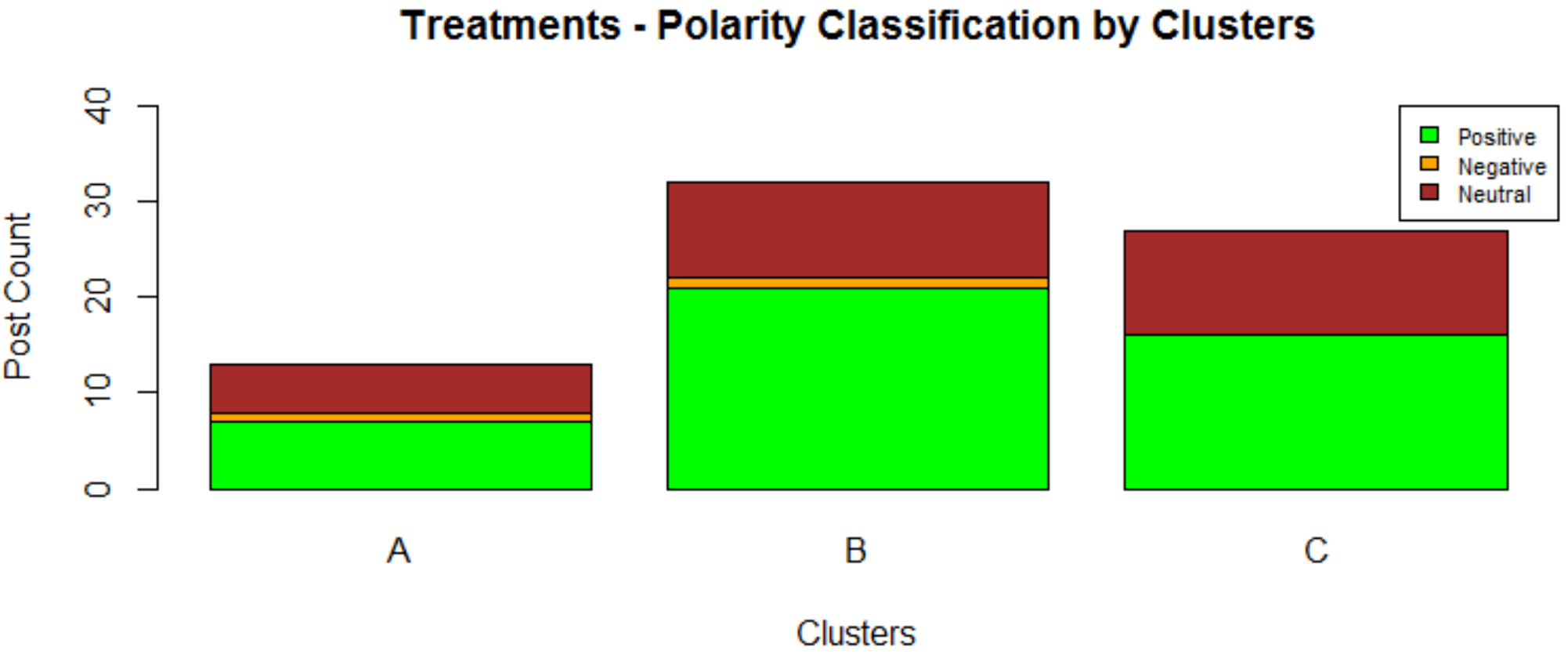
•Identification of dominant polarity in introduce yourself and treatments.

1* Agrawal, Shaishav. "Using syntactic and contextual information for sentiment polarity analysis." Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human. ACM, 2009.

Text Classification



Text Classification



Topic Modelling

- Posts are a mixture of topics, a topic is a probability distribution over words.

- LDA technique is used to obtain topics learnt using Gibbs sampling.

- Fusion to abstractly describe each topic within a cluster.

- Labelling every cluster.

Topic Modelling

Introduce_Clusters	Topic 1	Topic 2	Topic 3
A	vertigo	loud	plugs
	hissing	neck	vertigo
	annoying	hyperacusis	neck
B	perforation	whistling	sudafed
	beep	pregnancy	lexapro
	mucus	fever	earache
C	humming	dose	siren
	pulsing	whistling	hiss
	whistling	intensity	block

Topic Modelling

Cluster A : Abstract Topic 1 Name {Disease,Sound}
Abstract Topic 2 Name {Disease,Sound}
Abstract Topic 3 Name {Disease}
Cluster B : Abstract Topic 1 Name {Infection}
Abstract Topic 2 Name {Infection,Sound}
Abstract Topic 3 Name {Decongestants,Infection}
Cluster C : Abstract Topic 1 Name {Sound}
Abstract Topic 2 Name {Quantity,Sound}
Abstract Topic 3 Name {Sound}

Cluster A: Disease, Cluster B: Infection, Cluster C: Sound

Topic Modelling

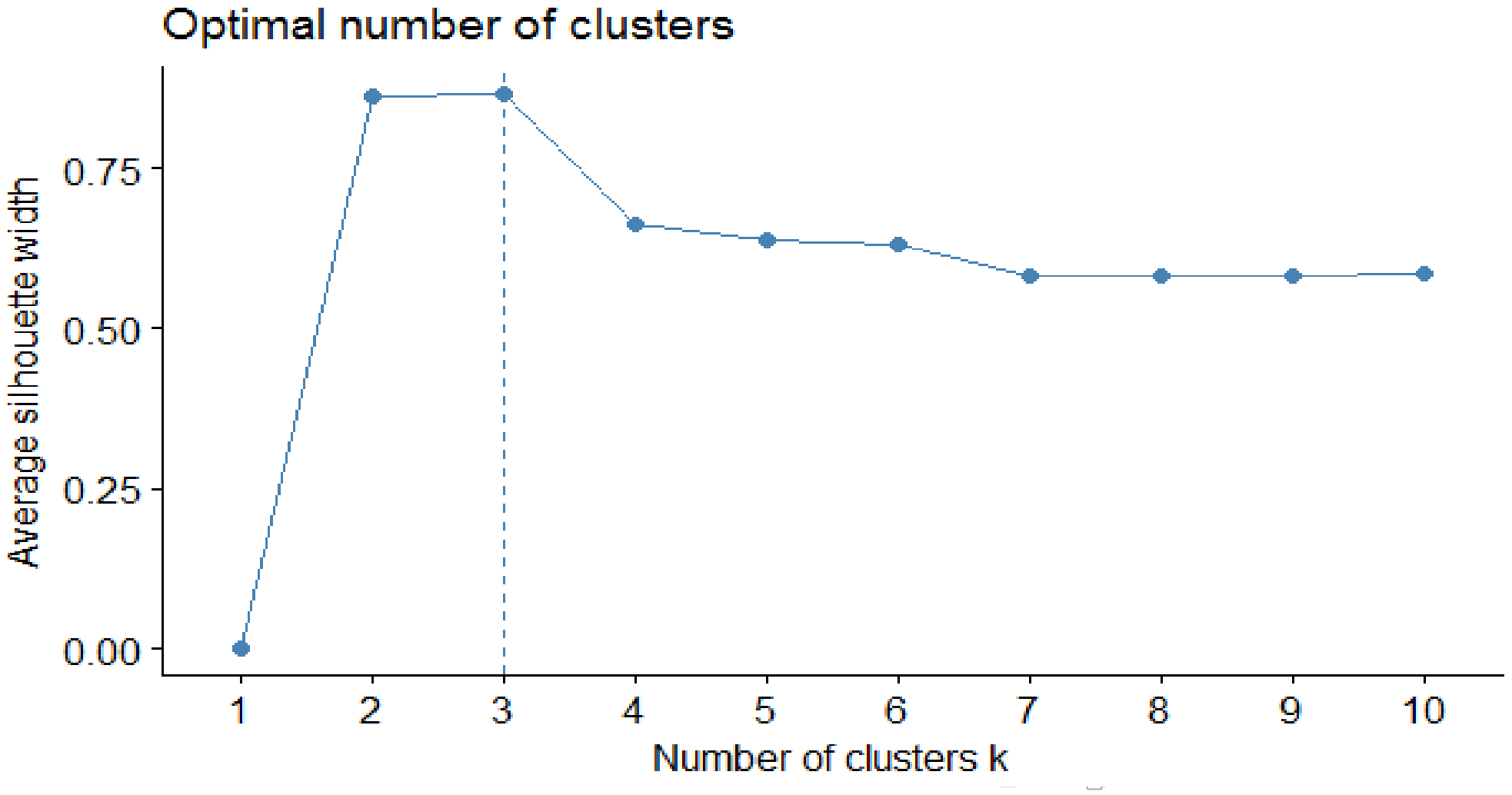
Tretment_Clusters	Topic 1	Topic 2	Topic 3
A	cbt	drug	retigabine
	biloba	antidepressants	drug
	bppv	elacin	neurofeedback
B	gaba	otic	estimme
	vascular	ganglion	neurostimulator
	neuromonics	otoharmonics	zinc
C	stem	epilepsy	carbamazepine
	ringing	loudness	ringing
	melatonin	magnesium	noises

Topic Modelling

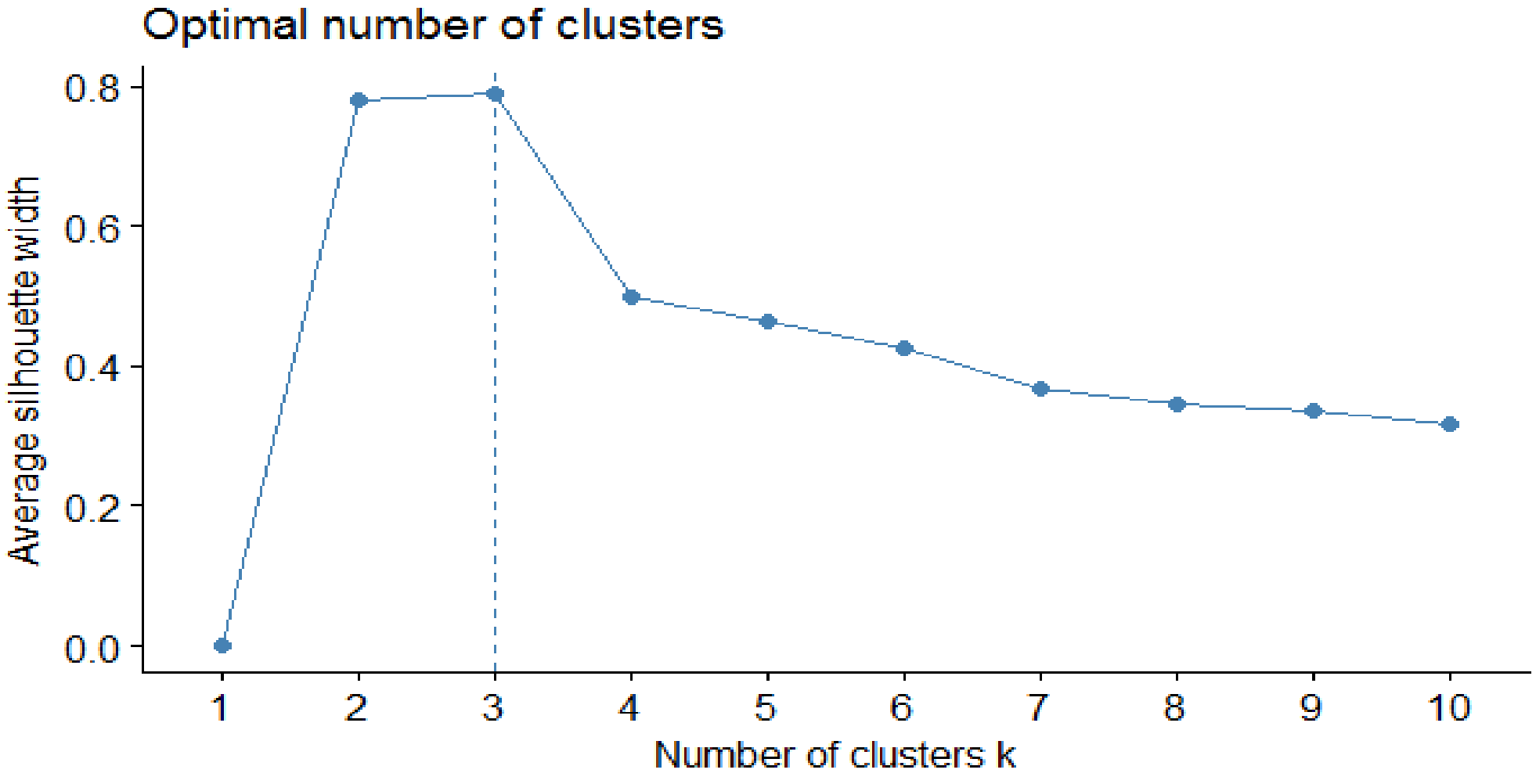
Cluster A : Abstract Topic 1 Name {Drugs,Therapy}
Abstract Topic 2 Name {Drugs}
Abstract Topic 3 Name {Anticonvulsant,Drugs}
Cluster B : Abstract Topic 1 Name {Devices,Supplements}
Abstract Topic 2 Name {Disorder,Supplements}
Abstract Topic 3 Name {Treatment,Supplements}
Cluster C : Abstract Topic 1 Name {Sound,Treatment}
Abstract Topic 2 Name {Health Condition,Sound}
Abstract Topic 3 Name {Sound,Drugs}

Cluster A: Drugs, Cluster B: Supplement, Cluster C: Sound

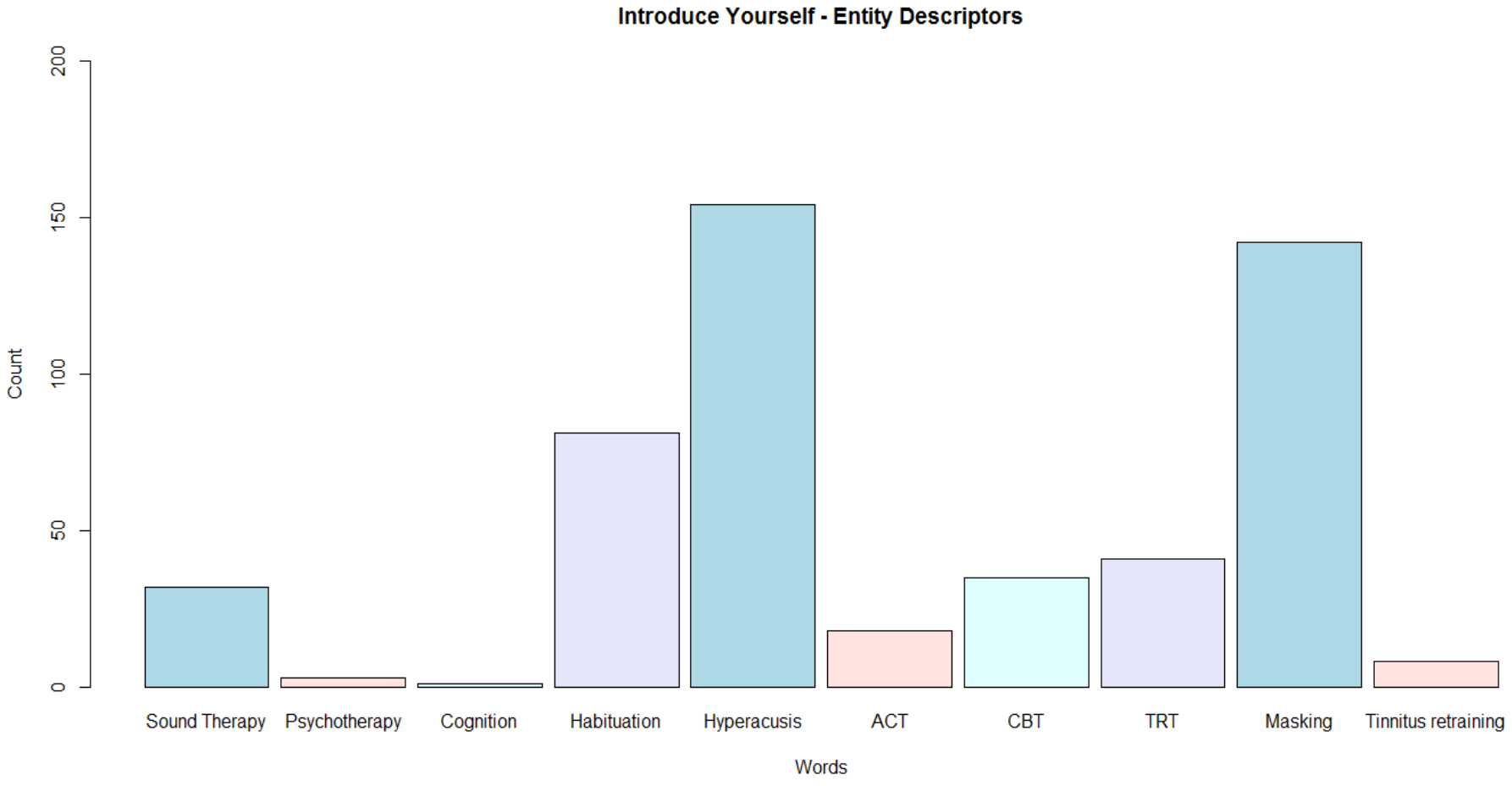
Silhouette Evaluation – Introduce Yourself



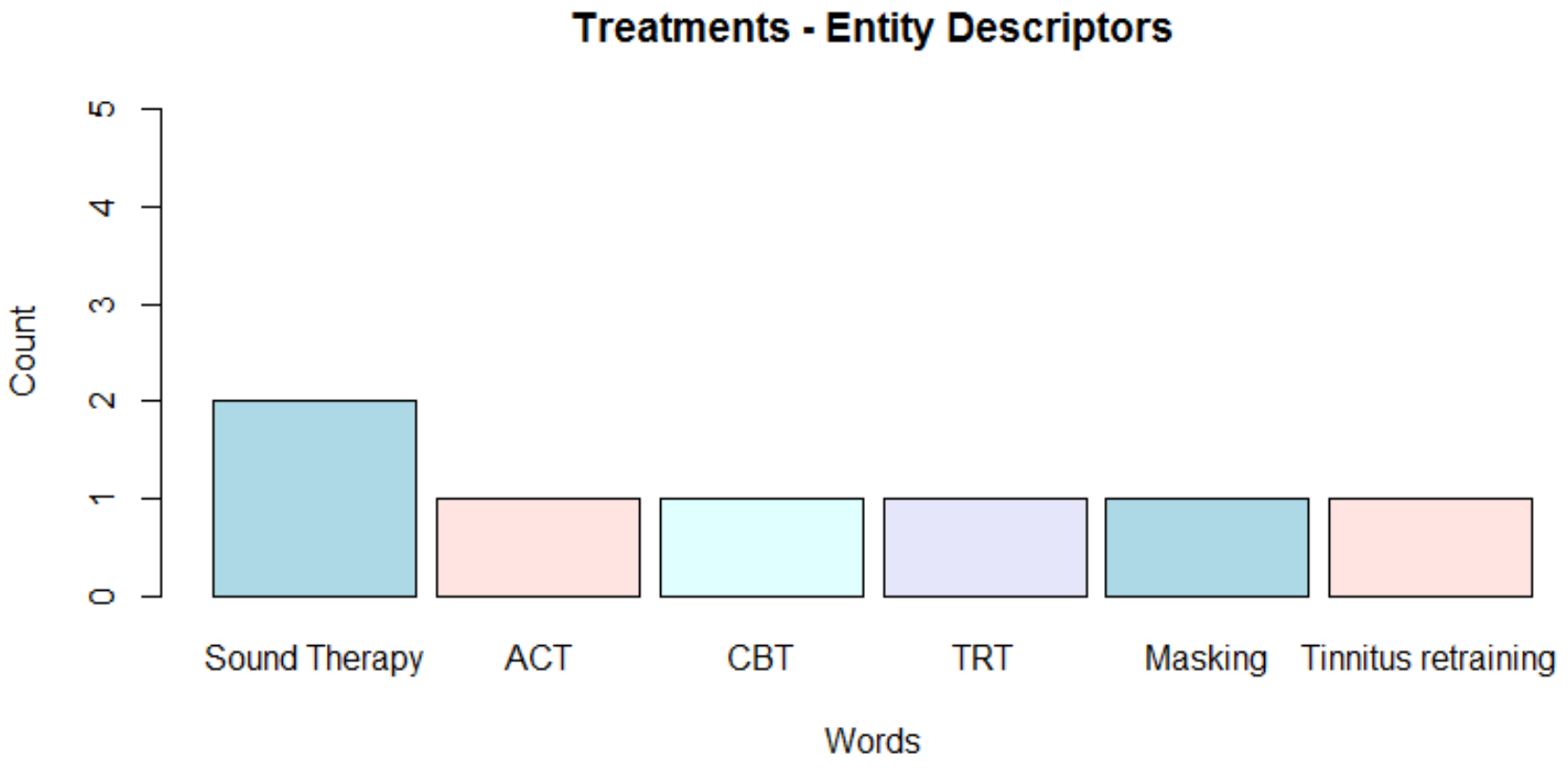
Silhouette Evaluation – Treatments



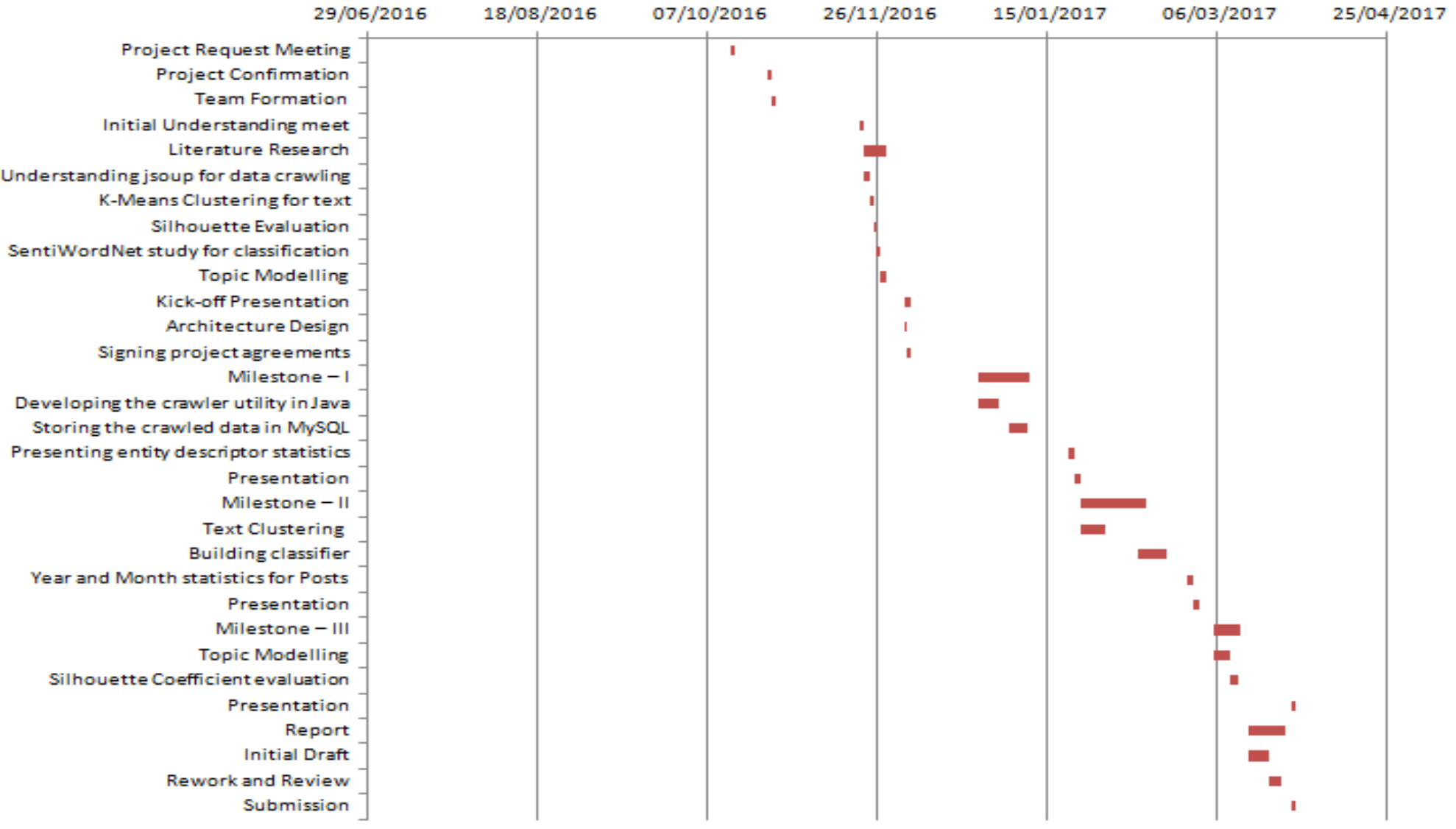
Statistics


Introduce Yourself - Entity Descriptors
Count
0
50
100
150
200
Sound Therapy
Psychotherapy
Cognition
Habituation
Hyperacusis
ACT
CBT
TRT
Masking
Tinnitus retraining
Words

Statistics


Treatments - Entity Descriptors
Count
0
1
2
3
4
5
Sound Therapy
ACT
CBT
TRT
Masking
Tinnitus retraining
Words

Timeline


29/06/2016
18/08/2016
07/10/2016
26/11/2016
15/01/2017
06/03/2017
25/04/2017
Project Request Meeting
Project Confirmation
Team Formation
Initial Understanding meet
Literature Research
Understanding jsoup for data crawling
K-Means Clustering for text
Silhouette Evaluation
SentiWordNet study for classification
Topic Modelling
Kick-off Presentation
Architecture Design
Signing project agreements
Milestone – I
Developing the crawler utility in Java
Storing the crawled data in MySQL
Presenting entity descriptor statistics
Presentation
Milestone – II
Text Clustering
Building classifier
Year and Month statistics for Posts
Presentation
Milestone – III
Topic Modelling
Silhouette Coefficient evaluation
Presentation
Report
Initial Draft
Rework and Review
Submission

Technical Specification

- Java programming.
- Python programming.
- R programming.
- MySQL database.
- Ms Excel.

References:

•Agrawal, Shaishav. "Using syntactic and contextual information for sentiment polarity analysis." Proceedings of the 2nd International Conference on Interaction Sciences: Information Technology, Culture and Human. ACM, 2009.

Thank You!