TOPIC MODEL VISUALISATIONS: A QUALITATIVE USER EVALUATION STUDY

MSc Data Science Dissertation

August 2024

Varbanov, Alex
av2049@hw.ac.uk
H00456607
Master of Science in Data Science (F2D1-DSC)

supervised by Le Bras, Pierre (PhD, FHEA)
pierre.le_bras@hw.ac.uk

Contents

List of Figures

List of Tables

Abstract

This research project explores the preferences of non-expert users of interactive topic modelling visualisations by developing a usability testing experiment and conducting a qualitative evaluation of the user experience. More specifically, we are interested in two topic modelling visualisation techniques – projection and cluster-based mapping and their efficacy in conveying the thematic structure of a large corpus of text documents.

First, we explore and identify state-of-the-art topic modelling visualisation strategies. Focusing on visualising output produced by probabilistic generative models such as Latent Dirichlet Allocation (LDA) (Blei, et al., 2003), our investigation notes the extensive amount of effort put towards evaluating the computational performance and algorithmic simplicity of topic models, parallel to the restricted amount of effort put towards usability evaluation of corresponding visualisations.

We found LDAvis (Sievert & Shirley, 2014) to be the most prominent visualisation tool used by researchers (Maskat, et al., 2023) and as such it served as the baseline interface for our usability testing. It implements dimensionality reduction to project topics onto a 2D plane aiming to maintain distance as a similarity measure.

However, such projections can become cluttered and cognitively strenuous to use. In our review, we also noted the importance of displaying a large number of topics in an interpretable manner, especially when presenting to non-expert users (Padilla, et al., 2014). Research into human perception and memory recall suggests that a cluster-based approach (one that visually groups elements on the screen) would have a lesser cognitive load on the user.

To test this, we devised an experimental study in a controlled environment, producing a second interface utilising a cluster-based approach, alongside a training interface with no mapping of the topic space. We then asked non-expert users to perform thematic analysis on a corpus of discussion forum messages. We used a modified Post-Study System Usability Questionnaire to conduct an interview evaluating their experience. Finally, we processed their input and analysed their answers.

Chapter 1

Introduction

How does one extract information from our shared global digital knowledge? To this day the most common way to find and navigate through documents online is by using keyword searching and linked documents (Blei, 2012). This technique is good for finding an item that fits a user-defined topic but is not fit for capturing the intrinsic structure of the topic itself or its broader thematic relationships.

On the other hand, the ever-increasing amount of data being generated, digitalised, and stored makes conventional searching even harder. However, the simultaneous increase in computational power enables us to explore this problem through machine learning, natural language processing (*NLP*) and statistical modelling (Egger & Yu, 2022).

In recent years several topic modelling methods have been developed to discover latent topics in a collection of documents with minimal human involvement. These models produce high-dimensional data structures that are not easily interpretable. Through various techniques, we can visualise the thematic relationships discovered by a certain model to understand them better (Sievert & Shirley, 2014).

Topic models developed over the past two decades have been evaluated not only by their performance but also by the visual representation of the model output (Egger & Yu, 2022; Maskat, et al., 2023). The latter is important for analysis as a list of topics alone conveys little knowledge about the corpus and its domain.

1.1 Background

Visualisation of topic modelling output still poses a challenge to researchers alongside more performance-based aspects such as *interpretability, computational complexity* and *output stability* (Maskat, et al., 2023).

Digestibility (how quickly a user can extract knowledge and the scale of the visualisation), *expressiveness* (how much knowledge a user can extract), *autonomy* (how much human involvement is required for production) and *interactivity* have been highlighted as desired goals when developing visualisation methods (Padilla, et al., 2014; Sievert & Shirley, 2014). To achieve these, developers have employed a plethora of visualisation strategies. While term-based visualisations such as *lists* and *word clouds* are most common, other representations such as *maps, graphs, charts, matrices* and more have also been developed and used (Maskat, et al., 2023).

The type of visualisation produced is very much dependent on the corpus domain, problem and the research questions being investigated. More complex issues, for example, require the integration of a set of different visualisations (Helldin, et al., 2018; Choi, et al., 2018).

It has been suggested that more attention needs to be paid to the development of user interfaces (Blei, 2012; Sievert & Shirley, 2014). Researchers develop visualisations to meet certain functional criteria and evaluate their work based on these quantitative metrics. They often cite usability as potential for future work and acknowledge the need to involve domain-expert users in the development process (Helldin, et al., 2018; Blei, 2012).

On the other hand, the need to present topic-modelling output to experts outside of the field of computer and data science has also been acknowledged (Padilla, et al., 2014). The lack of technical background can lead to bigger mental strain and therefore lessen the ability of an interface to convey the thematic structure of the corpus.

There seems to be little research on how well different methods of topic visualisations convey themes – both to expert and to non-expert users. This study aims to provide more insights into the matter.

1.2. Aim and Objectives

The overall goal of this research project is to evaluate the usability of different forms of topic visualisations. While ways to improve interpretability are often built into models on a mathematical or statistical level, qualitative evaluation of the produced visualisations is also needed to explore the best visual strategy to achieve higher interpretability of the corpus as a whole. We will focus on users who have little understanding of the inner workings of a topic model, with the hope of providing a general solution to the problem.

To achieve the main goal, we need to address the following questions:

Q1. What different methods for visualising output produced by topic models are there and what is the current state of the art?

Q2. What layout mapping strategies are these methods employing to portray the underlying thematic structure of the corpus?

Q3. How efficient are these methods in relaying this structure to the user?

In practice, we can investigate these questions by conforming the study to the following objectives:

O1. To provide a comprehensive review of the literature about topic modelling visualisations.

O2. To identify the functional properties of state-of-the-art topic modelling visualisations, particularly in their ability to convey the thematic structure of the corpus.

O3. To devise a user-based experiment evaluating the quality of visualisations in terms of usability and user preference.

The findings of this study will benefit topic modelling researchers by providing a more informed way to assess their model performance and improve how they present their findings to colleagues and non-expert users alike.

Chapter 2

Literature Review

In this chapter, we will address our first two objectives, i.e. to conduct a comprehensive literature review and identify the functional properties of state-of-the-art topic modelling visualisations.

We will first consider how one example of a machine learning model extracts topics from large text corpora and what is the raw output it produces. We will then see what visualisation strategies are employed to translate this output onto the screen. Finally, we will see how basic and complex topic representations can work alongside user interaction to produce more insightful visualisations.

2.1. Topic Modelling: Latent Dirichlet Allocation

We will start by examining *Latent Dirichlet Allocation (LDA)* (Blei, et al., 2003) as it is the topic model behind the most prominent topic visualisation tool used by researchers - *LDAvis* (Maskat, et al., 2023). It is important to understand how it works and what raw output it produces, as this is the heart of the visualisation problem.

LDA is a three-level hierarchical Bayesian model for collections of discrete data such as text corpora. Each document in the collection is modelled as a finite mixture over an underlying set of topics. Each topic is modelled as an infinite mixture over an explicit representation of the documents (i.e. the words they consist of).

With generative probabilistic models, we assume that the process that generated the data includes both observed and hidden variables (Blei, 2012). This defines a joint probability distribution over them, which we can use to compute the conditional distribution of the hidden variables, also called the posterior distribution.

2.1.1. LDA: Formal Definition

In the context of topic modelling, our observed variables are the documents themselves and the hidden variables - the topics and thematic structure of the text corpus. We can formally describe the generative process using Blei's notation (Blei, 2012). We define:

- $\mathbf{K}$ to be the finite set of topics,
- $\mathbf{D}$ to be the corpus, i.e. the collection of documents,
- $\mathbf{n}$ to be the index of a word in a document,
- $\boldsymbol{\beta}_{\mathrm{k}}$ to be a topic, i.e. a distribution over the vocabulary of the corpus,
- $\boldsymbol{\theta}_{\mathrm{d}}$ to be the mixture of topics per document d,
- $\mathbf{w}_{\mathrm{d}}$ to be the observed words in a document d,
- $\mathbf{z}_{\mathrm{d}}$ to be the topic assignment for each word in document d.

The generative process for LDA therefore can be described with the following equation:

$$p(\beta_{1:\mathrm{K}}, \theta_{1:\mathrm{D}}, \mathrm{Z}_{1:\mathrm{D}}, \mathrm{W}_{1:\mathrm{D}},) = \prod_{i=1}^{K} p(\beta_i) \prod_{d=1}^{D} p(\theta_d) \left(\prod_{n=1}^{N} p(\mathrm{Z}_{d,n} \mid \theta_d)\, p(\mathrm{W}_{d,n} \mid \beta_{1:\mathrm{K}}, \mathrm{z}_{d,n}) \right)$$

Equation 1: Blei, 2012

A thing to note here is the presence of conditional probabilities, e.g. that of the per-word topic assignments and the observed words per document. These and other dependencies define LDA and are how we discover the hidden structure of topics through the only observed variable w.

We will examine LDA's probabilistic graphical model to understand how these dependencies are connected. This is a visual way to describe families of probability distributions often used by researchers (Blei, et al., 2003; Buenaño-Fernandez, et al., 2020; Kumari, et al., 2021; Cheng, et al., 2020). Each node in the model is a random variable in the generative process. LDA's probability graphical model is shown below (*Fig.1*; Kumari, et al., 2021).

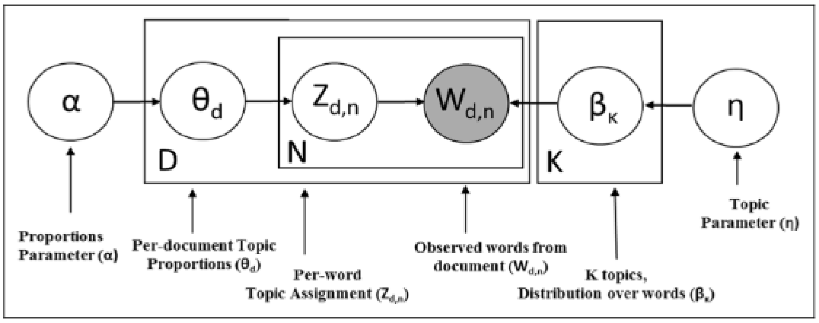

α
θd
Zd,n
Wd,n
βK
η
D
N
K
Proportions Parameter (α)
Per-document Topic Proportions (θd)
Per-word Topic Assignment (Zd,n)
Observed words from document (Wd,n)
K topics, Distribution over words (βk)
Topic Parameter (η)

FIGURE 1: LDA'S GRAPHICAL MODEL (KUMARI, ET AL., 2021)

2.1.2. LDA: A Practical Example

A brief look at a more practical example from Blei in their article on *Probabilistic Topic Models* (*Fig.* 2; Blei, 2012): on the left-hand side is depicted the set of topics **K**, where every topic $\boldsymbol{\beta}_k$ is defined by the probabilities of words from the vocabulary (e.g. topic 1 consists of the terms: "*gene*", "*dna*", "*genetic*" with the respective probabilities of occurrence 0.04, 0.02 and 0.01).

The probability of occurrence of a certain term from the vocabulary within a particular topic is usually estimated using *Variational Bayes methods* or *Collapsed Gibbs Sampling* (Griffiths & Steyvers, 2004), and the probability of occurrence of the term from the empirical distribution of the corpus (Sievert & Shirley, 2014). These sampling methods typically output two key structures: the distribution of topics in documents and the distribution of top words (for efficiency) in topics.

In the middle, we can see how a document expresses three out of the four topics listed in the example set. On the right-hand side, depicted are the topics distributed over the whole collection of documents (shown via histogram). Keywords in the document are highlighted with the colour corresponding to their topic.

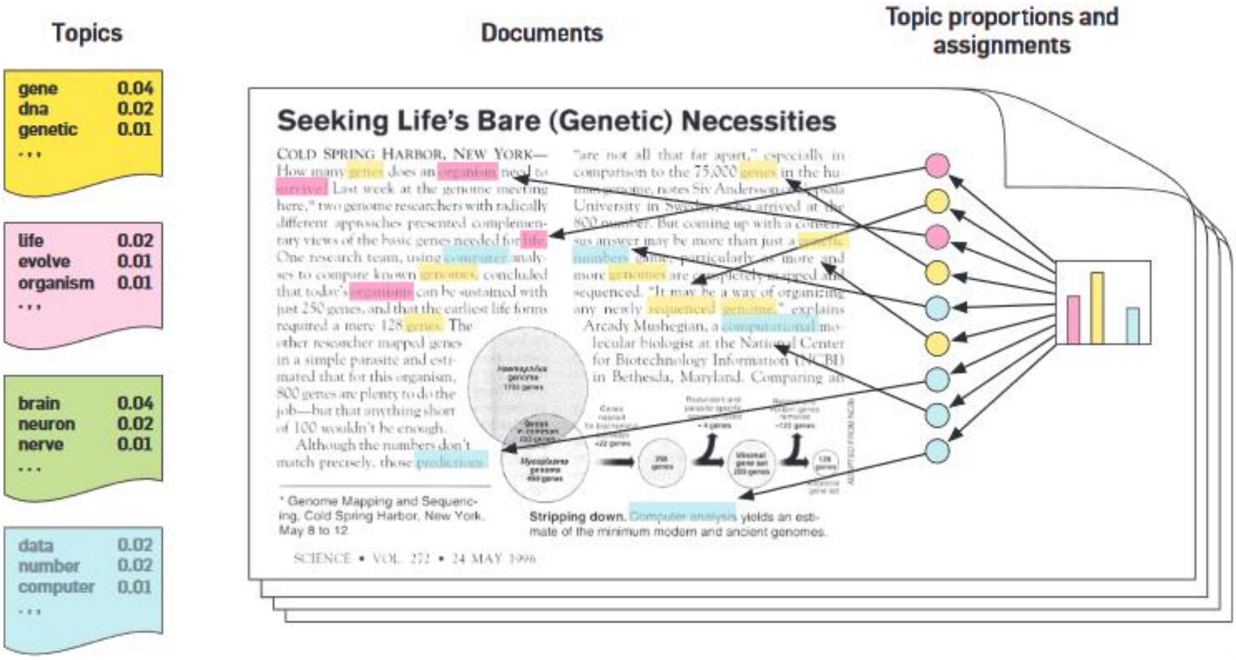

Topics
Documents
Topic proportions and assignments
gene 0.04
dna 0.02
genetic 0.01
...
life 0.02
evolve 0.01
organism 0.01
...
brain 0.04
neuron 0.02
nerve 0.01
...
data 0.02
number 0.02
computer 0.01
...
Seeking Life's Bare (Genetic) Necessities
Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.
SCIENCE • VOL. 272 • 24 MAY 1996

FIGURE 2: LDA EXAMPLE (BLEI, 2012)

2.1.3. LDA: Summary

LDA takes a pre-processed corpus (Egger & Yu, 2022) of documents and produces **K** topics. These topics are described in two ways: either by a probability distribution over the vocabulary (every word in every document) or a distribution over every document. In practice, these distributions are represented by matrices, where one dimension is **K** and the other is the vocabulary or the corpus.

When we take a particular topic **k** expressed by the vocabulary and order its terms by the probability of occurrence we can try and interpret what this topic could be about. Moreover, we can use this topic vector **k** as raw data input into a different system that visualises these topics on the screen to ease and enhance interpretability.

The next section of the literature review breaks down topic visualisation techniques, starting with *LDAvis* - the most prominent topic visualisation tool in recent years (Maskat, et al., 2023).

2.2. Topic Visualisations

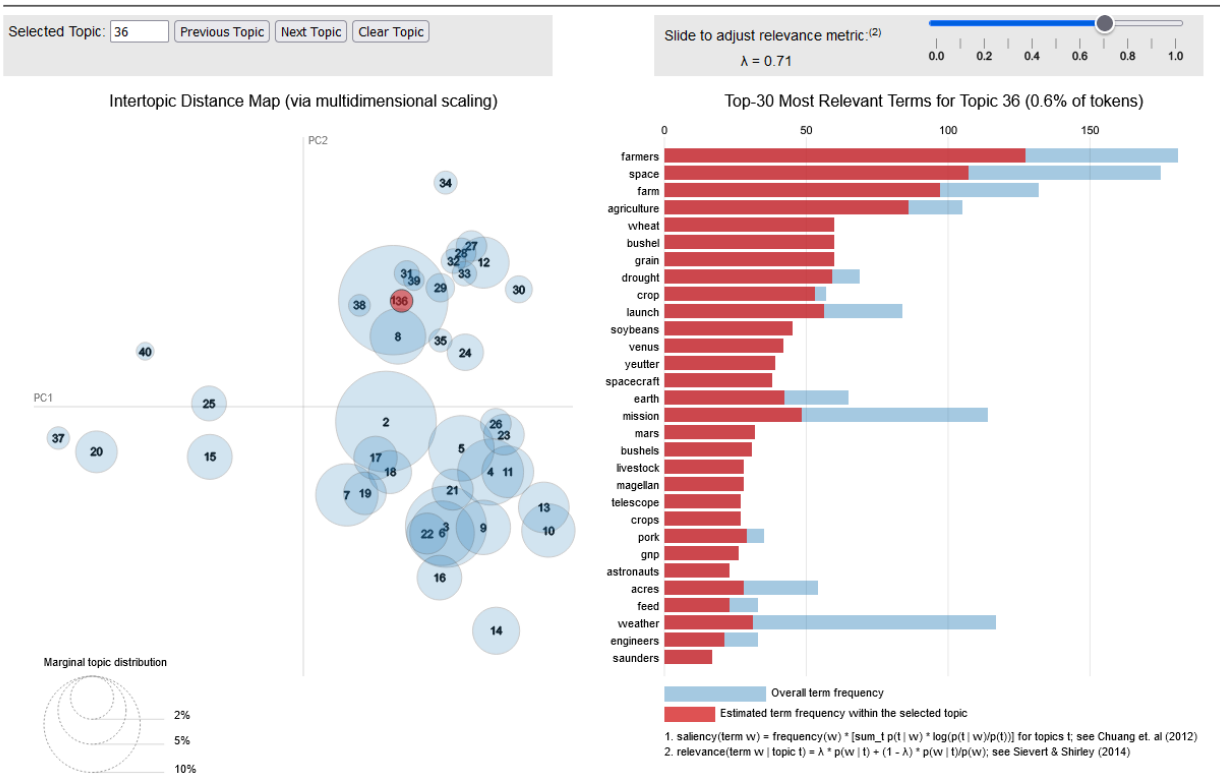
The output produced by the topic models can then be further processed to gain more insights about the topics. How popular is a certain topic within the text corpus, what are the top terms that define it and how similar is it to other topics are just a few of the questions that researchers can investigate.

As is typical for data analysis, these research objectives are strongly aided by visualisations, with interactivity being named as the most promising among basic techniques. The remainder of this literature review will investigate different types of visualisations and visualisation-producing methods. Some related figures have been placed in Appendix A: *Topic Modelling Visualisation Figures* to maintain text formatting and image clarity.

2.2.1. LDAvis

LDAvis, as mentioned previously, has been discovered to be the most popular tool for visualising topic research. It is a web-based interactive system, developed using R and D3.js (Sievert & Shirley, 2014). *LDAvis* takes the data structures produced by a Latent Dirichlet Allocation model – topic-vocabulary distribution and topic-document distribution, alongside other corpus data to generate a projection-mapping of these topics in the 2D space, using *Jensen-Shannon divergence* to compute inter-topic distance and *Principal Component Analysis* as a scaling method. An implementation called *pyLDAvis* has also been available as a free licence Python library (Mabey, 2015), which has contributed to the method's popularity amongst researchers.

The interactive dashboard *(Fig. 3)* offers users two viewports: on the left-hand side – a global view of the topic distribution space and on the right – a bar chart describing a particular topic by its terms. A settings pane sits on top of the viewports which allows, amongst other things, for adjusting the parameter **λ**, which controls term *relevance* within a topic.



1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)
2. relevance(term w | topic t) = λ * p(w | t) + (1 - λ) * p(w | t)/p(w); see Sievert & Shirley (2014)

FIGURE 3: LDAVIS SHOWING 2246 ASSOCIATED PRESS DOCUMENTS (SIEVERT & SHIRLEY, 2015)

LDAvis: Term relevance

Sievert and Shirley define relevance as the balance between the probability of a term occurring within a topic and its *lift*, i.e. the ratio of said probability to the term's marginal probability over the corpus. The parameter **λ** is the weight coefficient of a logarithmic scale of those two factors. Values of **λ** closer to 0 mark topic-specific terms as more relevant, while values closer to 1 prioritise more generic terms.

LDAvis: Topic Insights and Interactivity

The visualisation uncovers more or further builds upon insights from the model's output. For instance, the model partitions the corpus into the preset number of topics. In the LDAvis global overview, a user can quickly deduce the relationship between two topics: proximity depicting similarity, size corresponding to the topic's probability of occurrence within the collection, and both combined can describe a level of abstraction (e.g. "*global politics*" vs "*national politics*"). The authors also propose k-means clustering to facilitate semantic zooming.

The bar chart view displays the most popular 30 terms in a selected topic or the whole corpus of no topic is selected. The red and blue bars depict terms' relevance and *saliency* (the overall occurrence of a term within the corpus). This is where interactivity starts playing a role in better understanding the topic search space.

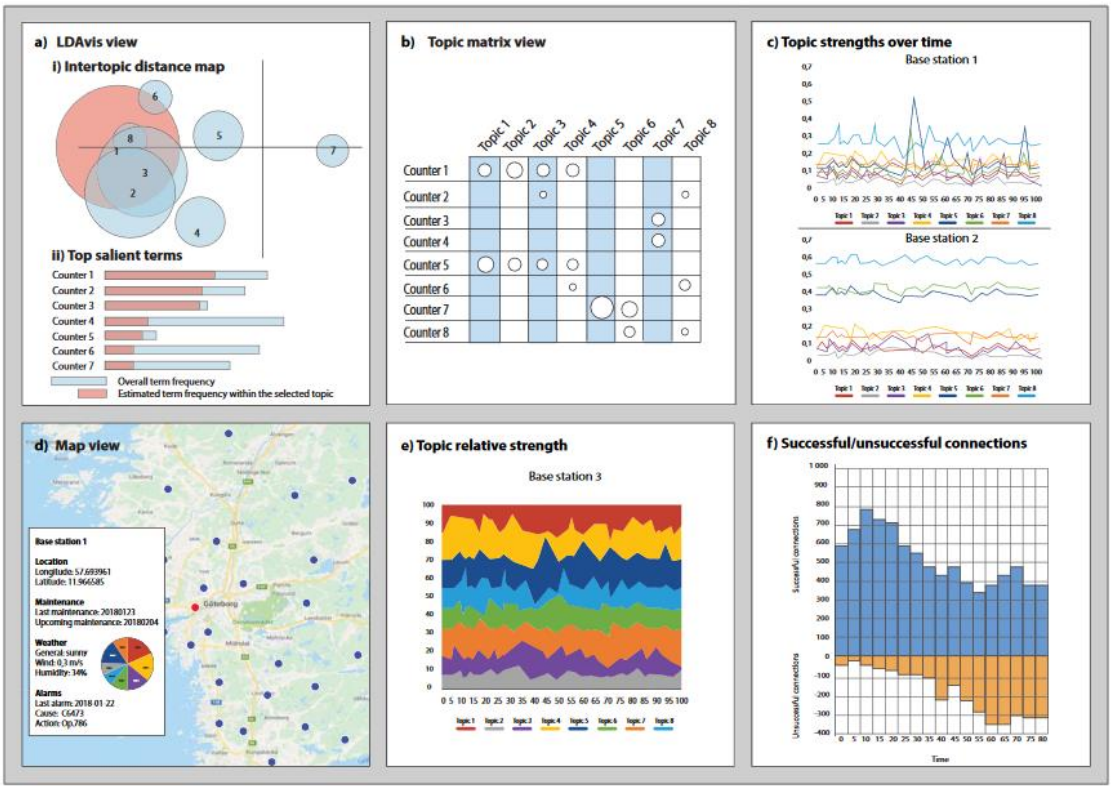
As noted before LDA has been found to produce topics, more difficult to interpret than other models. The weight coefficient **λ** can be adjusted to update the term relevance in real-time and to improve topic interpretability. The authors suggest a value of 0.6 as a starting point but acknowledge that this might differ from corpus to corpus. Hovering over a certain term would then update the topic overview panel to depict how this term is distributed among different topics.

LDAvis: Limitations

LDAvis has been proven popular with NLP researchers as its interactive elements allow users to dig deeper into topic-term relationships (Maskat, et al., 2023). However, certain problems warrant more complex visualisations or entirely different visualisation formats, to depict temporal changes, geographical data, and specific term distribution, among others.

Helldin, et al., (2018) use *LDAvis (Fig. 4.a)* in conjunction with a *streamgraph (Fig. 4.e)*, *term-topic probability matrix (Fig. 4.b)* and a *line chart (Fig. 4.c)*, to better describe the situation awareness in telecommunication networks. They provide more context by showing a *map visualisation (Fig. 4.d)* of control towers and a *stacked bar chart* of connections success *(Fig. 4.f)* in their *visualisation grid*.

Han, et al., (2020) have designed their interactive dashboard *(Fig. 5)* to visualise a sentiment index over time via a *line chart*, a *topic-document map* to visualise topic distributions within the corpus and a *word cloud* to display the term distribution across the corpus. These are just two examples of how different visualisation elements complementing each other can help the observer gain knowledge. We will investigate further in the next section.



Base station 1

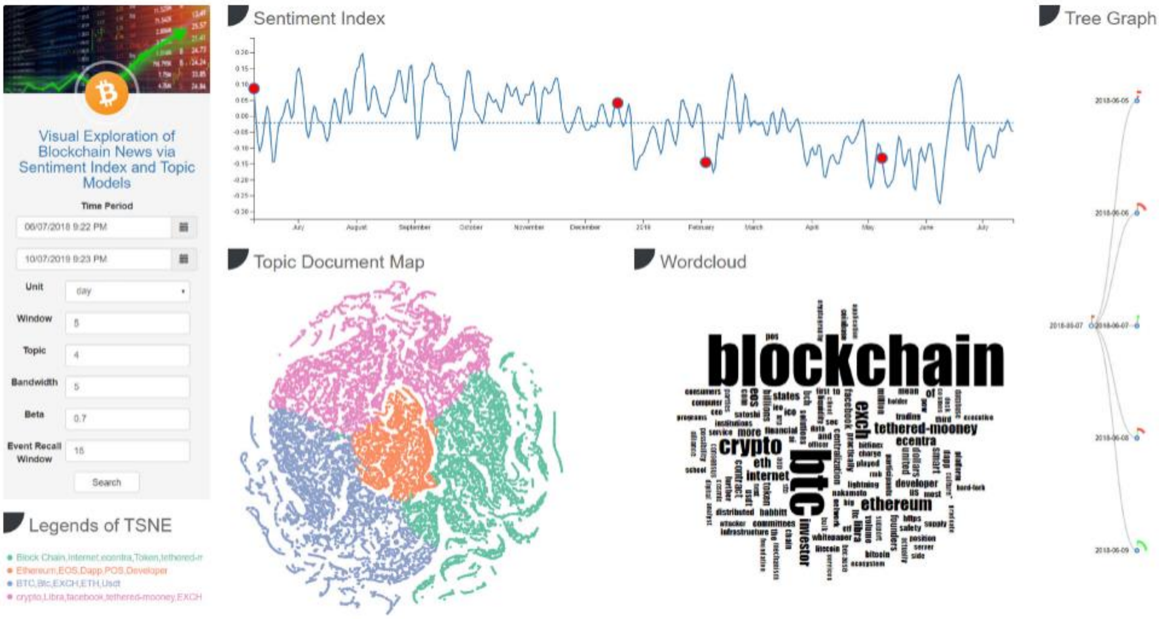
Location
Longitude: 57.693961
Latitude: 11.966585

Maintenance
Last maintenance: 20180123
Upcoming maintenance: 20180204

Weather
General: sunny
Wind: 0,3 m/s
Humidity: 34%

Alarms
Last alarm: 2018 01 22
Cause: C6473
Action: Op.786

FIGURE 4: VISUALISATION GRID SHOWING TELECOMMUNICATION DATA (HELLDIN, ET AL., 2018)



Visual Exploration of Blockchain News via Sentiment Index and Topic Models

Time Period
06/07/2018 9:22 PM
10/07/2019 9:23 PM
Unit: day
Window: 5
Topic: 4
Bandwidth: 5
Beta: 0.7
Event Recall Window: 15

- Block Chain,Internet,econtra,Token,tethered-n
- Ethereum,EOS,Dapp,POS,Developer
- BTC,Btc,EXCH,ETH,Usdt
- crypto,Libra,facebook,tethered-mooney,EXCH

FIGURE 5: VISUALISATION DASHBOARD SHOWING FINANCIAL NEWS DATA (HAN, ET AL., 2020)

2.2.2. Basic Topic Visualisations

Word Clouds

We start by looking at the different building blocks of topic model visualisations. These include *lists*, *graphs*, *grids* and *line charts* among others. To depict a body of text on screen, arranging words in a *list is* the most basic approach alongside the *word cloud* (Maskat, et al., 2023).

A *word cloud* is a cluster of keywords extracted from a body of text, usually coloured for clarity, where a term's relevance is encoded in its depiction in font size. This technique could be used to describe the whole corpus (Lee & Ostwald, 2024) or a particular topic (Goswami, et al., 2019). In terms of layout, the words can be positioned randomly around the cluster (*Fig. 5 – bottom right*; Lee & Ostwald, 2024; Koylu, 2019), could be concentrated with the biggest terms appearing towards the centre (Appendix A: *Fig. A*, Kumari, et al., 2021; Le Bras, et al., 2020) or could also follow an outward spiral path of ranking (Appendix A: *Fig. C;* Choi, et al., 2018).

Lists

A few years before *LDAvis* was first introduced, Chaney & Blei (2012) produced a list-based navigational tool for Wikipedia articles and a few other collections of documents. It provides 2 types of *list visualisations* – document and topic-related over 3 different visualisation layouts (Appendix A: *Fig. B*):

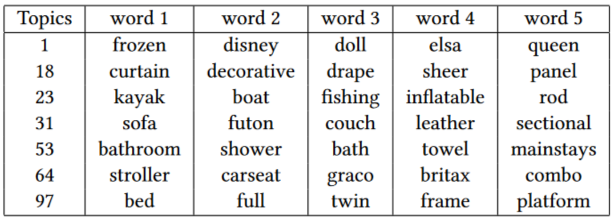
1. An overview of the corpus structure described by the topics (overview page).
2. A topic-specific view (topic page).
3. A document-specific view (document page).

The ordering of the list can depict either the probability of occurrence or similarity between items. In this implementation, the lists are laid out vertically, which is usually, but not necessarily the case (Cheng, et al., 2020).

Grids and Tables

Pieces of text could also be laid out in two dimensions using grids and tables. Goswami, et al. (2019) use the 5 top terms to define their topics and have separated those into a table as they aim to subtract the overlapping words from two separate word clouds representing supply and demand in e-commerce (*Fig. 6*). The vertical orientation of the list of topics is following the previously mentioned example, however, the addition of columns for order of terms allows for quick navigation and word search across the table. Topics are also given an index for identification.

Helldin, et al. (2018) use a similar visualisation method as part of their visualisation grid. The topic matrix view (*Fig. 4.b*) displays the probability of occurrence of terms within topics using a matrix where the horizontal axis represents the topics and the vertical - the terms (in this case counters).



Topics	word 1	word 2	word 3	word 4	word 5
1	frozen	disney	doll	elsa	queen
18	curtain	decorative	drape	sheer	panel
23	kayak	boat	fishing	inflatable	rod
31	sofa	futon	couch	leather	sectional
53	bathroom	shower	bath	towel	mainstays
64	stroller	carseat	graco	britax	combo
97	bed	full	twin	frame	platform

FIGURE 6: TABLE OF TOPIC DEFINITIONS (GOSWAMI, ET AL. 2019)

The probability is depicted using different-sized circles (or bubbles) within a cell of a topic-term pair. This further builds on using two-dimensional space to show topic-term relationships by introducing size as an analogue for the probability of occurrence (in this case of a word). We will now further examine how distance and position in two-dimensional space can be used to encode insights produced by the topic model.

2.2.3. Complex 2D Representations

Maskat, et al. (2023) surveyed forms of visualisation in 42 published articles about topic modelling. They have come up with 5 different categories of visualisation:

1. **maps**, where geographical information is present in the data.
2. **networks**, where the connections or associations between items are described in the data.
3. **evolution-based**, aiming to show progression over a period.
4. **charts**, i.e. any other topic visualisation where none of the above are depicted. These are the common figures used in data visualisation – bar charts, scatter plots, pie charts, etc.
5. **others** – custom-made tools for a specific problem in a given domain.

These methods utilise 2D space differently alongside other visualisation marks encodings and representation channels. *Charts* are usually used alongside the first three categories described above and the *others* category covers niche solutions. So, our focus will remain on the *maps*, *evolution-based* and *networks* categories.

We can further extend the relationship-focused category in terms of the layout of the topic or term marks on screen. These three sub-categories are as follows:

6. **projections**, where position and distance in the 2D space convey similarity.
7. **clustering**, where similar items are grouped to convey hierarchy.
8. **networks**, where the connections or associations between items are visualised and may convey information about their relationship.

Geo-based Layouts

Geographical data can come into play in topic analysis as simple as a visual representation of the probability of a topic occurring at a particular geo-location and its geographical spread (Asghari, et al., 2020) to visualising several topics occurring in the same region (Appendix A: *Fig. C.ii*, Choi, et al., 2018; Martin & Schuurman, 2017).

Glyph maps contain additional charts mapped to the geographical coordinates associated with the data. Term-specific geo-information can also be depicted. Geographical representations are often concerned with detecting outliers among their immediate neighbours as seen in Appendix A: *Fig. C.i* where they are highlighted with a red box.

Temporal-based Layouts

Some research is concerned with how trends in topics change over time. The most common way to portray this in two-dimensional space is to use the horizontal axis as an analogue for time and the vertical axis as the attribute change, we want to measure.

Traditionally, line charts have been used as a simple way to implement this. Helldin, et al. (2018) use line charts as part of their visualisation grid to depict topic strength over time (*Fig. 4.c*) while Han, et al. (2020) tracked the sentiment score developed by them to detect real-world events that might have triggered the conception of new trends within a text corpus of financial news (*Fig. 5*).

Le Bras, et al. (2020) and Kumari, et al. (2021) use line charts to visualise the volume of documents increasing over time. Choi, et al. (2018) allow the user to select a point in time through a slider in their interactive visualisation *TopicOnTiles* (Appendix A: *Fig. C*), however, we will examine interactivity in depth in the following section.

Let us briefly consider two more complex types of temporal data representations – the *streamgraph* and the *evolutionary path map*. Haidar, et al., (2016) use streamgraphs to show temporal term strength in a corpus of news articles about World War 1 (Appendix A: *Fig. D,* Haidar, et al., 2016).

They are basing their work on *ThemeRiver* (Havre, et al., 2002) which is used to visualise temporal term strength in several different text corpora while employing Byron & Wattenberg's (2008) pioneering streamgraph visualisation model which emphasises *legibility* and has proved popular with online users.

We can apply the same strategy to topics, while also visualising how they combine or split up over time to form new abstract or specific topics. Liu, et al., (2020) have produced several evolution path mappings to visualise the research field of artificial intelligence and machine learning throughout the 1990s and mid-2000s (Appendix A: *Fig. E*).

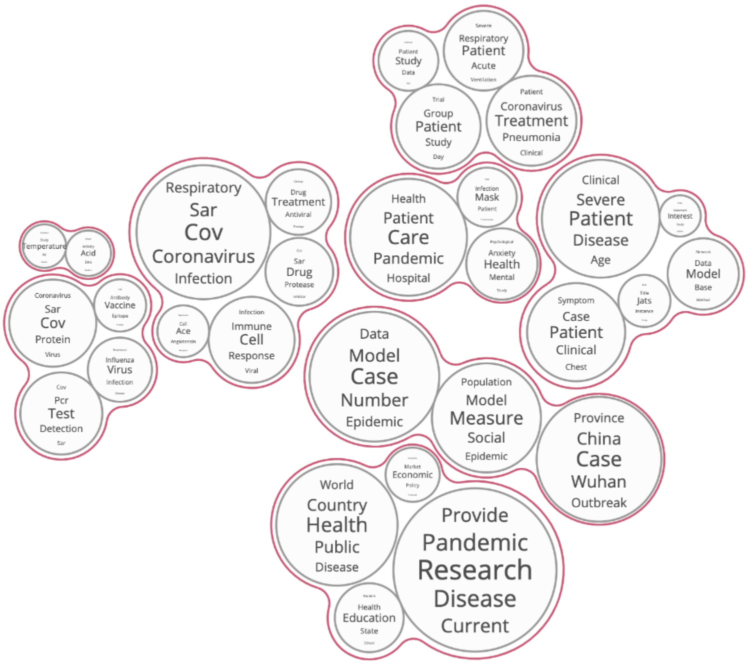
If we track the change of the beige stream of *Database Systems/ XML Data* we can see how it merges with *Natural Language System/ Machine Translation* to form the new topic of *Data Mining/ Association Rules,* when research was at its peak around the middle of the decade, indicated by the number of streams and their relative thickness.

Relationship-based Layouts

We saw how 2D space is utilised to map geographical metadata about topics. We also observed how this space can be used to depict change over time. We can otherwise use the topic mark's position on the screen to either show its relative similarity to other topics or to cluster together similar objects to depict hierarchy. Padilla, et al. (2014) use position to convey similarity by dispersing topics in a hexagonal grid.

LDAvis achieves both by using a bubble chart in its topic space representation of the corpus. This approach can be applied to other, non-probabilistic topic models as well (Egger & Yu, 2022). Bubbles are close or far away from each other based on topic similarity. One topic can appear to encircle many others, which can be interpreted as a form of hierarchy, e.g. in *Fig. 3* we can see that topic 1 encapsulates topics 8, 29, 31, 36, 38 and 39.

Le Bras, et al. (2020) depict hierarchy by performing *agglomerative clustering* of topics, based on the cosine distance between the topics-document vectors and drawing an outline around clusters (*Fig. 7*) to visualise a theme grouping of several document-level topics. A bubble and a five-term word cloud depict a topic. Sievert & Shirley (2014) have shown the possibility of applying further k-means clustering in the *LDAvis* projection mapping based on the position of the topic marks to partition the topic space in **k** different areas. Han, et al. (2020) use the same strategy to depict the corpus' structure by clustering documents into topics *(Fig.5)* and visualising the corpus in a *topic document map.*



Patient Study Data
Severe Respiratory Patient Acute Ventilation
Trial Group Patient Study Day
Patient Coronavirus Treatment Pneumonia Clinical
Study Temperature Air
Activity Acid Dna
Respiratory Sar Cov Coronavirus Infection
Clinical Drug Treatment Antiviral Therapy
Cov Sar Drug Protease Inhibitor
Coronavirus Sar Cov Protein Virus
Antibody Vaccine Epitope
Respiratory Influenza Virus Infection Disease
Cov Pcr Test Detection Sar
Cell Ace Angiotensin
Infection Immune Cell Response Viral
Health Patient Care Pandemic Hospital
Infection Mask Patient
Psychological Anxiety Health Mental Study
Clinical Severe Patient Disease Age
Statement Interest Study
Network Data Model Base Method
Title Jats Instance
Symptom Case Patient Clinical Chest
Data Model Case Number Epidemic
Population Model Measure Social Epidemic
Province China Case Wuhan Outbreak
Market Economic Policy
World Country Health Public Disease
Provide Pandemic Research Disease Current
Student Health Education State School

FIGURE 7: CLUSTER-BASED MAPPING (LE BRAS, ET AL., 2020)

We can also use two dimensions to spread topics, terms or documents and draw connections between items to encode semantic relationships. Buenaño-Fernandez, et al. (2020) use a bipartite *network diagram* to depict the documents directly and link them to their assigned topics (Appendix A: *Fig. F.i*).

The reader can instantly get a rough estimate of the number of topics exhibited in a document by looking at the clustering of documents and the lines drawn to the respective topic circles. They can observe topic associations through the lines linking documents.

Connections are drawn between lists of terms, topics and pieces of text in the computer-assisted argument extraction tool Topics2Themes (Appendix A: Fig. G, Skeppstedt, et al., 2018). Kumari, et al. (2021) use a network diagram to depict the connectedness of topics by colour-coding clusters and depict relationship strength through the connecting line's thickness (Appendix A: *Fig. F.ii*).

Won, et al. (2021) use *network diagrams* to describe the topics themselves, rather than to show inter-topic relationships. In *Fig. 8* we can see how they link the topic's top terms, in this case about infection control, and how the term's probability is encoded into the size of the circle representing it.

All terms depicted could be traced to the topic's centre node. Keywords could also be connected, thus conveying a deeper understanding of the topic structure and allowing the researcher to define it with a single-phrase label.

These methods use the available space to evenly disperse the mark representations on screen, to improve legibility. The horizontal axis can be used similarly to the temporal designs to depict hierarchy. Cheng, et al. (2020) use an interactive tree diagram to visualise this and directly show which documents exhibit which topics (Appendix A: *Fig. H*).

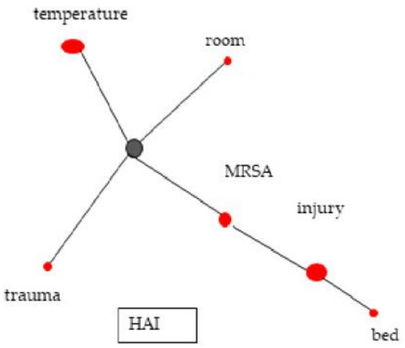


FIGURE 8: TERM-TOPIC NETWORK DIAGRAM
HAI: healthcare-associated infection

(WON, ET AL. 2021)

2.2.4. Summary

Topic modelling output is generally visualised with various charts using different strategies for encoding insights into the 2D space. *Table 1.* describes the visualisation elements mentioned so far and the insights they convey to the observer. In the reviewed articles, only visualisations depicting the output of topic models were considered. Some other basic charts, frequently used in data analysis are also included as they provide meaningful insights about the topic space.

These images are usually shown when reporting in scientific papers but can also be part of interactive visualisation dashboards. We will investigate a few examples of the latter in the following section. Temporal and spatial visualisations are usually reserved for specific use-case scenarios, so we will focus on how the 2D screen space is utilised to convey insights about the thematic structure of the corpus.

Table 1. Visualisation Elements Summary Table

Basic Charts	
word cloud	
insights:	relevance of a word in a body of text; the body of text defined by its top terms;
used in:	*Padilla, et al., 2014; Goswami, et al., 2019; Kumari, et al., 2021; Lee & Ostwald, 2024; Choi, et al., 2018; Le Bras, et al., 2020; Egger & Yu, 2022;*
list of items	
insights:	corpus, topic or document structure described by topic, document or word definitions; ranking of items according to relevance;
used in:	*Buenaño-Fernandez, et al., 2020; Skeppstedt, et al., 2018; Odlum, et al., 2020; Kumari, et al., 2021; Sievert & Shirley, 2014; Chaney & Blei, 2012; Cheng, et al., 2020; Le Bras, et al., 2020; Martin & Schuurman, 2017; Han, et al. 2020 (via LDAvis);*
line chart	
insights:	topic strength over time; topic strength within the corpus; top terms strength over time; number of documents per topic over time;
used in:	*Helldin, et al., 2018; Koylu, 2019; Kumari, et al., 2021; Le Bras, et al., 2020; Zhang, et al., 2020;*
stacked bar chart	
insights:	term frequency for a specific interval; term frequency for both topic and corpus;
used in:	*Havre, et al., 2002; Sievert & Shirley, 2014; Han, et al., 2020 (via LDAvis);*
pie chart/doughnut chart	
insights:	topic exhibition per document; topic exhibition per geo-location;
used in:	*Chaney & Blei, 2012; Cheng, et al., 2020; Choi, et al., 2018;*
bar chart	
insights:	total number of documents in the corpus at a point in time over a period; term frequency within a topic, document or corpus; total number of topics assigned to documents in a corpus; total number of documents assigned to a topic; topic strength for a specific time interval; topics strength within the corpus;
used in:	*Buenaño-Fernandez, et al., 2020; Koylu, 2019; Chaney & Blei, 2012; Cheng, et al., 2020; Le Bras, et al., 2020;*

Table 1. Visualisation Elements Summary Table

2D Space Utilisation	
dendrogram/ tree diagram	
insights:	topic hierarchy within the corpus space;
used in:	*Buenaño-Fernandez, et al., 2020; Cheng, et al., 2020; Egger & Yu, 2022;*
grid/table/matrix	

insights:	show the relationship between two or more entities (topic-document, topic-word, etc.);
	the most relevant terms;
	comparison of items;
	summary of items;
	ranking;

used in:	*Goswami, et al., 2019; Helldin, et al., 2018; Cheng, et al., 2020; Buenaño-Fernandez, et al., 2020; Won, et al., 2021; Zhang, et al., 2020;*
geo-map	
insights:	show how strongly a topic is depicted at a geographical location at a point in time or period; show if a topic is depicted in a specific geographical region; show how a topic is spread across a geographical space according to its terms;
used in:	*Asghari, et al., 2020; Koylu, 2019; Liu, et al., 2020; Choi, et al., 2018; Martin & Schuurman, 2017;*
streamgraph	
insights:	term strength over time in a body of text ; topic strength over time in a corpus;
used in:	*Havre, et al., 2002; Helldin, et al., 2018; Haidar, et al., 2016;*
evolution path map	
insights:	semantic evolution of the topic over time; topic strength over time; topic's relationship to other topics;
used in:	*Liu, et al., 2020; Koylu, 2019;*
projection-based layout	
insights:	inter-topic similarity; topic term relevance[1]
used in:	*Helldin, et al., 2018; Padilla, et al., 2014; Sievert & Shirley, 2014; Choi, et al., 2018; Han, et al., 2020 (via LDAvis); Egger & Yu, 2022 (via LDAvis and BERTopic);*

[1] Spiral layout in a word cloud (Choi 2018)

Table 1. Visualisation Elements Summary Table	
2D Space Utilisation	
cluster-based layout	
insights:	partition a topic space into general areas; partition a document space into topic areas; depict the relationship between a high-level topic and its sub-topics;
used in:	*Han, et al., 2020; Le Bras, et al., 2020; Sievert & Shirley, 2014;*[2]
network-based layout	
insights:	association between words, topics, themes (high-level topics) and documents; the strength of the connection;
used in:	*Skeppstedt, et al., 2018; Buenaño-Fernandez, et al., 2020; Kumari, et al., 2021; Odlum, et al., 2020; Won, et al., 2021; Zhang, et al., 2020;*

TABLE 1: VISUALISATION ELEMENTS SUMMARY TABLE

[2] optional K-means clustering on topic projections, not cluster-based.

2.3. Interactive Dashboards

Interactive visualisations can have different levels of complexity depending on the specific problem, target audience and the problem's domain. Interactions with the streamgraphs, for example, are simple and only limited to *mouse hovering*.

Chaney & Blei's *List-based Navigational Tool* (Appendix A: *Fig. B*) and *LDAvis* are visualisation tools that offer topic modellers several additional means of interaction and are more insightful but to various degrees.

Choi, et al.'s *TopicOnTiles* (2018) and Le Bras, et al.'s *Covid-19 Bubble Maps* (2020) are all dashboards that surpass the earlier-mentioned methods in terms of complexity and information gain. This shows us that interactivity allows visualisations to be simultaneously compact and thorough, something Sievert & Shirley (2014) set out to do when they developed *LDAvis*. In the following paragraphs, we will examine how different visualisation elements can work together with interactivity to produce more insightful ways of conveying knowledge.

2.3.1. Simple Interactive Dashboards

Although a layout comprised of a few list views is relatively simple, two strategies can greatly improve the interpretability of the corpus - ordering the lists and allowing the user to navigate quickly across related topics and documents.

A small preliminary study was conducted with positive reviews of the *List-based Navigational Tool*. One user noted they had discovered insights they would otherwise not have used traditional Wikipedia browsing. There is a version available[3] online which has a few additions but is still limited to lists and tables.

List-based Navigational Tools

In the original published design, the user is initially presented with an overview page consisting of the list of topics defined by their top three terms laid over a horizontal bar chart depicting the probability of occurrence of topics in the corpus. By clicking on a selected topic, the user navigates to that topic's page where he can see more of the

[3] https://www.cs.mcgill.ca/~isavov/arxiv_demo/browse/topic-presence.html

words defining the topic, a list of related documents for that topic, ordered by their topic exhibition, and a list of related topics, ordered by similarity (Appendix A: *Fig. B.ii*).

The user can then select a particular document and navigate to that document's page. There they can see the document in its original form, a pie chart showing the proportions of the topics exhibited in it, a list of related topics and a list of related documents, ordered by similarity.

Topics2Themes (Appendix A: Fig. G) is a more recent tool with a similar general layout. In addition to exploring different topics, this tool allows the user to create "themes" and link them to the relevant text and by extension - topics and terms.

The tool consists of three list views - a list of terms, a list of topics and a list containing snippets of labelled text. Connections between items on those lists are visualised and hovering over an item highlights those connections in the other lists. Clicking on an item immediately brings up all connected elements to the top of the list. The panels are also equipped with a search box and sorting options.

While both visualisations use lists as their key method of organising data they differ in some major ways. First, the *corpus and documents are described in the List-based Navigational Tool* by the probabilities of the topics exhibited and *Topics2Themes* does not offer such insights.

Second, *Topics2Themes* offers slightly more interactivity by allowing lists to be reordered based on user input. When comparing the two, we must be mindful that *Topics2Themes* is a tool developed to assist the user in argument extraction not with the prime goal to visualise the topic space.

We can, however, say that they both operate within the realm of topic analysis. So does *LDAvis*. However, its projection-based depiction of the corpus, contrary to the *List-based Navigational Tool*'s simple annotated bar chart, offers insights into inter-topic relationships and a deeper understanding of the corpus' thematic structure.

2.3.2. Projection-based Layouts

LDAvis Interactivity

LDAvis's dashboard design (Fig. 3) was discussed earlier when introducing the concept of topic visualisations (see *Section 3.1*). Also briefly mentioned was its *interactivity*, i.e. reordering topic terms according to the balance of their respective within-topic and within-corpus relevance. We can compare how ranking items can offer differing insights in different scenarios.

Topics2Themes ranks items by their association with the user's selection aiding in discovering *thematic relationships between topics*. *LDAvis* re-orders terms to allow the user to *interpret a topic itself* better. To aid topic interpretability *Topics2Themes* automatically groups related words.

LDAvis offers additional interactivity via a navigation panel allowing the user to traverse through the topics in order and to clear the selection.

The biggest advantage of *LDAvis* over the other dashboards examined so far is its *projection-based layout* of the topic space. The user can at a glance understand the corpus' *thematic structure* in terms of similarity between topics, find any big clusters indicating more abstract themes, the topic sizes, etc. This mapping also serves as the main selection method for topics. In the case of topics nearby, the user can easily switch and quickly compare one and the other.

The *List-based Navigational Tool* only offers insight into the corpus structure by showing the probability distribution of topics via a horizontal bar chart of 40 topics. The user is forced to scroll down the page to investigate topics with a lower probability of occurrence.

The topic probability distribution is also portrayed in *LDAvis* via the topic's bubble size. The whole corpus is fitted in a single viewport that is always present, which greatly eases interpretability. *Topics2Themes* offers no insights into the corpus structure.

Lastly, there are some functionalities of the list-based dashboards that *LDAvis* does not implement. These are the ability to inspect raw bits of the text data and to search the corpus or output by a specific keyword.

2.3.3. Cluster-based Layouts

Hierarchical Clustering

Cheng, et al. (2020) use LDA, Hierarchical LDA (HLDA) and Phrase Mining (PH) to extract topics from bioengineering text corpora. They visualise their results using several simple tools packed under the name *TopExplorer*.

List views are used to visualise the LDA and PH modelling results, while HLDA output is represented through an interactive tree diagram (Appendix A: *Fig. H*).

The topic lists allow the user to select a topic to display its related documents and a document to display its related topics. The tree diagram allows the user to click to expand or collapse a branch at a topic node. To get to a detailed view, containing the text and its contextual topic breakdown, the user can select a document in both views.

In this example, the global topic space is observable in its two-tier top-down hierarchical tree representation. *Topic 1* is the corpus-level topic, and *Topics 2, 7* and *11* are the thematic groupings of the document-level topics. Following a tree path can broaden the contextual interpretation of a topic.

TopExplorer similarly to Chaney & Blei's visualisation functions more like a website hosting multiple webpages than a dashboard (where all interactive elements are always present in a single window). It is reasonable to suspect an increase in the observer's cognitive load, rendering the tool less effective.

To conclude our review of visualisations we will last investigate how hierarchical clustering can be visualised in conjunction with bubble size encoding (also utilised by *LDAvis*) in a singular dashboard window via the use of Bubble Treemaps (Gortler, et al., 2018).

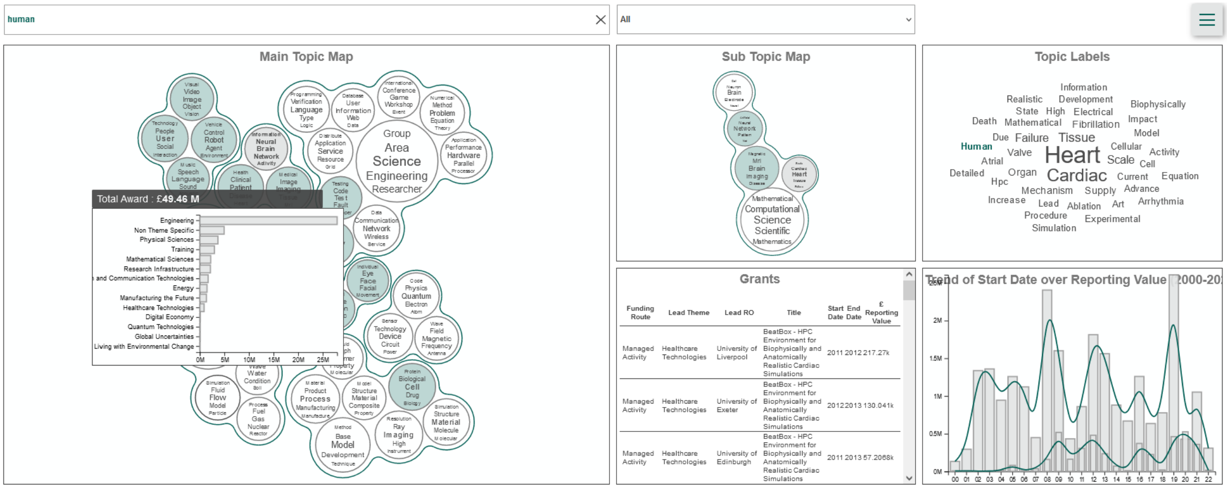
Bubble Maps

Le Bras, et al. (2020) use a simplified Bubble Treemap to depict hierarchy more insightfully. They have modelled two different COVID-19 research datasets of considerably different sizes (about 80:1GB ratio, cluster mapping of the smaller set shown in *Fig. 7*). The TopExplorer demo available online[4] and the figures depicted in the research article show only the results of modelling a corpus of 20 documents over 10 topics (Cheng, et al., 2020).

However, when the corpus size is increased to 600 documents, the topics in the second level become 7 and they increase to 60 in the last. Scaling up requires more visual elements on the screen, i.e. nodes and their connections. This will likely hinder the user's cognitive performance.

Le Bras, et al. (2020) use LDA to model 30 and 50 different main topics respectively for their chosen datasets and 200 and 400 sub-topics (based on word vector similarity as mentioned in *Section* 3.3). They handle hierarchy by introducing a separate view showing the user selection's sub-topic breakdown.

The latest implementation of this dashboard design (Gharavi, et al., 2022) uses a corpus of UKRI research grants, exhibiting 40 main, otherwise called *super-topics,* and 200 sub-topics. We will continue referring to this dashboard henceforth as it has been extended with some additional functionality (*Fig. 9; Appendix A: Fig. I*).



Grants

Funding Route	Lead Theme	Lead RO	Title	Start Date	End Date	£ Reporting Value
Managed Activity	Healthcare Technologies	University of Liverpool	BeatBox - HPC Environment for Biophysically and Anatomically Realistic Cardiac Simulations	2011	2012	217.27k
Managed Activity	Healthcare Technologies	University of Exeter	BeatBox - HPC Environment for Biophysically and Anatomically Realistic Cardiac Simulations	2012	2013	130.041k
Managed Activity	Healthcare Technologies	University of Edinburgh	BeatBox - HPC Environment for Biophysically and Anatomically Realistic Cardiac Simulations	2011	2013	57.2068k

FIGURE 9: UKRI RESEARCH GRANTS BUBBLEMAP DASHBOARD (GHARAVI, ET AL., 2022)

[4] https://sites.google.com/unomaha.edu/topexplorer

The dashboard consists of a global topic space view, showing these 30 – 50 super-topics, clustered together according to their similarity using Bubble Treemaps (Gortler, et al., 2018). Similar topics are placed closer together, so proximity is a similarity measure, but exact distance is not.

In addition to the sub-topic view, the detail view panels on the right-hand side reveal more information based on the user selection. The topic's keyword structure is depicted in a word cloud with more important terms closer to the centre. Links to the documents ordered by relevance are also available. Finally, temporal data about the publication date of research articles has been processed with the weight of super or sub-topics over time to produce trends visualised through a stacked bar chart and a line chart overlay.

The final bit of functionality we will examine is case-specific to some extent but demonstrates how pre-existing labels can be incorporated with generated topics to uncover valuable knowledge.

The documents in the raw dataset are labelled with one of 14 different categories, depending on the subject field of the research grant. Selecting a particular category using a drop-down menu will change the transparency of the bubble depending on the distribution of documents per topic per category. As these grants have an award value associated with each one, which is also of key interest to the target audience, a bar chart appears when the user hovers over a bubble, showing the total award value for all documents under a certain topic and broken down by category in an ordered bar chart.

2.3.4. Summary

Adding interaction to existing visualisations can reveal more insights about the corpus's domain. We have seen how simple visualisations become more insightful by introducing interactivity, e.g. in the cases of the navigational list views and streamgraphs. Sophisticated dashboards with more interactivity, such as the *UKRI Research Grant BubbleMap,* paint a more detailed picture of the research space.

Ranking items also plays an important role in displaying relevant terms, topics, documents and categories. We have also examined how hierarchy can be visualised via the classic tree diagram or the more functional Bubble Treemap.

We have considered some factors that might impact the user experience. For example, dashboards reduce the cognitive load by eliminating the need for navigation between viewports thus freeing the user from committing observations to memory.

It is best to keep the topic space representation condensed in a single viewport that is always present. It is also advantageous to utilise the 2D space by encoding the thematic structure in the topic space representation. This viewport can also act as the main selection tool for the dashboard and can be used to highlight the distribution of terms according to a user selection, such as in LDAvis and UKRI Research Grant examples.

Finally, we saw how additional charts and document metadata can provide context to the user's selection such as temporal bar charts or distribution pie charts.

2.4. Literature Review Summary

In this chapter, we reviewed literature about topic modelling visualisations. We set out to identify the state-of-the-art methods and the various strategies for portraying the thematic structure of the corpus on screen.

We identified *LDAvis* as the most widely used tool for visualising topic modelling output. We briefly investigated *Latent Dirichlet Allocation* to understand the basis of the visualisation problem. LDA is the topic modelling algorithm that produces the raw data structures visualised on screen by the interactive dashboard. *LDAvis* works on this output data using dimensionality reduction in a projection-based strategy to depict topics in the 2D space (*Fig. 3)*.

Other visualisation methods, however, can range from basic – showing ranked lists and grids of items alongside basic charts – to more complex dashboards, including multiple and more insightful visualisation elements, thus providing an opportunity for in-depth analysis. The complexity level of visualisations depends on the underlying domain and research problem; however, the high dimensionality of the fitted model adds to the challenge of visualising its output (Sievert & Shirley, 2014).

No matter the level of complexity, when designing their visualisations researchers were often concerned with two main ideas – the *ranking of items* and *layout of marks* on screen, i.e. the utilisation of the 2D space.

Ranking is inherently important. For example, a topic is intuitively defined by its most used terms, and we can define a theme by looking at the biggest topics it contains.

Layout, while significant on its own in terms of perception, can also be used to convey thematic relationships and corpus/topic structure. *Proximity* and *distance* can depict *similarity* while *clustering* and *tree maps* can depict thematical hierarchy.

What is the most effective layout strategy, though?

While a topic model's computational complexity and output interpretability are often tested and evaluated in its development and application (Cheng, et al., 2020) or in dedicated studies (Egger & Yu, 2022), the produced visualisations in this review have only been comparatively evaluated by their authors (Skeppstedt, et al., 2018) and are behind in user evaluation, which usually involves less than 10 participants or is only preliminary (Chuang, et al., 2012; Havre, et al., 2002; Chaney & Blei, 2012).

Researchers state the importance of displaying many topics in a *condensed, visually appealing* way (Padilla, et al., 2014; Sievert & Shirley, 2014; Chuang, et al., 2012). The more substantial user evaluations examined in this review were aimed at topic interpretability by keywords (Sievert & Shirley, 2014) or at the efficacy of a proposed dashboard solution to a niche geo-based problem (Choi, et al., 2018).

With layout being a prime concern when designing topic visualisations, we can conclude that there is room to explore the *efficacy* of different 2D space utilisation strategies, particularly using user evaluation. This will in turn answer our final research question:

Q.3. How efficient are [different topic modelling visualisation methods] in portraying [the underlying thematic structure of the corpus] to the user?"

We will attempt to find an answer to this problem in the following chapters of this report, by targeting our final objective O3., i.e. to devise a user-based experiment evaluating the quality of topic modelling visualisations in terms of usability and user preference.

Chapter 3

Methodology

In this chapter, we describe the experimental design developed for this study by first elaborating on the implementation of visualisation interfaces used in the usability testing and second explaining the study protocol.

3.1. Hypotheses

In Chapter 2, Section 2.2. and Section 2.3. we investigated how inter-topic relationships can be expressed through many different layouts. Our research question – **Q2** considers the underlying thematic structure of the corpus, which is generally portrayed using relationship-based visualisations.

Semantically speaking, topics can be grouped into more abstract themes. This also might be a part of our natural ability to process information. *Cognitive chunking* is the process of committing and retrieving a smaller unit of information to and from memory as a larger, already familiar unit (Thalmann, et al., 2019). *The Gestalt princip*les (GPs) of perceptual organisation, on the other hand, inform us on how humans visually perceive groups of items as a singular object (Wagemans, et al., 2012).

Different layouts of the topic space might adhere to these principles differently. This in turn can directly influence cognitive chunking and therefore the interpretation of the thematic structure of the corpus.

In the previous chapter, we identified *LDAvis (Fig. 3)* as the most popular topic visualisation tool used by researchers. We also identified the need to develop more user-friendly interface designs, particularly when findings are being presented to an audience, whose professional expertise lies outside the field of STEM disciplines.

LDAvis uses dimensionality reduction to project topics in the 2D space. The topic similarity is encoded in the distance between the marks on the screen, therefore, the GP of *proximity* would play a key role in recognising groupings of topics, i.e. the larger themes.

The Bubble Treemap, which we examined in Section 2.3.3. on the other hand (*Fig. 7)*, uses *proximity* only to position similar topics next to each other, with the distance between marks being collapsed so everything is packed in one big cluster, thus conforming to GP of *element connectedness*. It also adheres to the Gestalt principle of *common region*, by displaying a border around the pre-defined clusters.

Other relationship-based layouts might adhere to some of these and additional Gestalt principles, however, to narrow down the scope of the experiment into a manageable timeframe we will focus on the most popular method as well as the other example we have reviewed of a relationship-based layout, which is also used as a part of an interactive interface in practice.

Based on the above, we can formulate several hypotheses for this experiment:

H1. *If a cluster-based layout facilitates interpretability and theme extraction better than a projection-based approach, then an interactive dashboard implementing the Bubble Treemap representation of the topic space would be preferred over the standard LDAvis implementation by evaluating users.*

H2. *If a projection-based layout emphasises truthfulness by least distorting inter-topic relationships, then evaluating users would cite an interactive dashboard implementing the LDAvis representation of the topic space as more truthful over one that does not use a projection-based approach.*

H3. *If users interacting with both a projection-based layout and a cluster-based layout are non-experts in STEM disciplines, then they would prefer higher interpretability over higher truthfulness.*

To test these hypotheses, we designed a mixed-method user study where participants would express their views on the interfaces incorporating the two layout strategies and their usefulness.

3.2. Interfaces Development

3.2.1. Data Preparation

The *20 Newsgroups* dataset available as part of the SciKit-Learn Python library was chosen as input for the topic extraction model. The sheer number of documents (around 20k) benefits topic coherence when working with a statistical probability-based model, such as LDA.

Additionally, the sources of these documents are "newsgroups", which were an early implementation of discussion forums, so the topics exhibited by these documents would be broad enough to be interpreted by anyone.

Finally, the constructing definitions of newsgroups helped to produce output with a well-defined thematic structure. The naming convention of a particular newsgroup doubles as its web address in the following manner: *Broad Theme* >> *Narrow Theme* >> *Specific Topic*, e.g. *rec.sport.hockey*.

We performed LDA topic modelling using the SciKit-Learn Python library to generate the relevant data structures required as input for the interactive visualisations. The required number of topics $\mathbf{K}$ was set to 20, corresponding to the number of newsgroups the documents were gathered from.

A standard topic modelling pipeline was followed: *documents* were vectorised, and *words* were lemmatised before fitting the LDA model to the processed corpus. To improve topic coherence the list of stop-words (colloquialisms, not factored into the model) was extended gradually until no such words could be easily detected in the top-terms lists for each topic.

While generally the produced outcome follows the "ground truth" thematic structure of the corpus, some newsgroups were absorbed into others and several small incoherent topics were generated. We used the same topic model output data for all interfaces to minimise confounding variables.

3.2.2. LDAvis

All three interfaces started as clones of an *LDAvis* JavaScript dashboard exported from the implementation of *pyLDAvis* – an open-source Python package. We did this to keep the relevance score functionality which orders the topic terms with higher interpretability.

pyLDAvis works by computing visualisation data and then plugging it into a visualisation dashboard produced using D3.js. This data is stored in a JSON format and includes the topic-term distributions to encode the size of the topic circles alongside the topics' display coordinates.

We removed the top control panels to allow users to focus on the layout itself, with the relevance control value – λ, internally set to 0.6 as the original paper authors recommend.

Hovering over a keyword would show its distribution across the corpus, which would help interpret topic relationships, so this was also disabled.

Finally, scientific terminology in labels was swapped with more colloquial language as our sample of participants would be pooled from the general population (Appendix A: Fig. J, *Fig. 10)*.

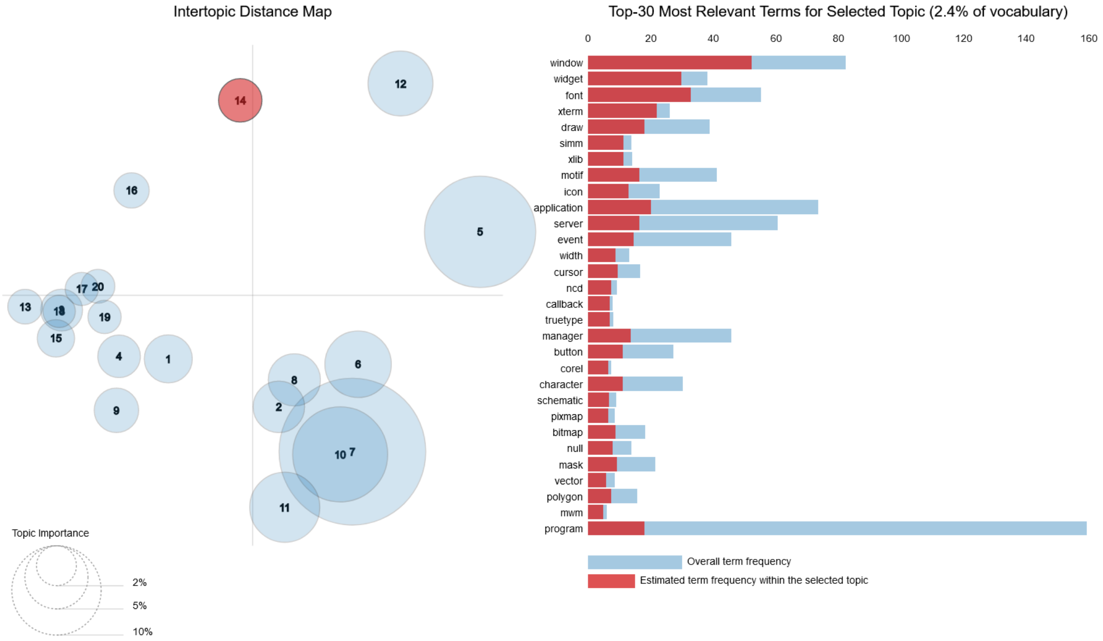


FIGURE 10: FINAL LDAVIS-BASED INTERFACE

3.2.3. Bubble Treemap

To produce the Bubble Treemap interface (henceforward referred to as *BubbleMap*) *agglomerative clustering* based on *cosine distance* between topic vectors was performed to generate the data structures needed as input for the visualisation. This was done in Python using the LDA output we utilised when producing the *LDAvis* dashboards. The data structures were exported as JSON data files.

A minor difference in approach when calculating similarity between topic vectors is that *LDAvis* calculates similarity based on the topic-term distribution. The *BubbleMap* implementation does so based on the topic-document distribution. To reflect that, the sizes of the marks are calculated based on these distributions and are slightly different between versions, however, the general thematic structure in terms of topic sizes portrayed remains the same.

The produced JSON file was loaded into an existing Bubble Treemap implementation, provided by personal communication by Dr. Le Bras. The resulting *BubbleMap* (Appendix A: Fig. K, *Fig. 12*) was integrated into the modified *LDAvis* dashboard, replacing the projection mapping.

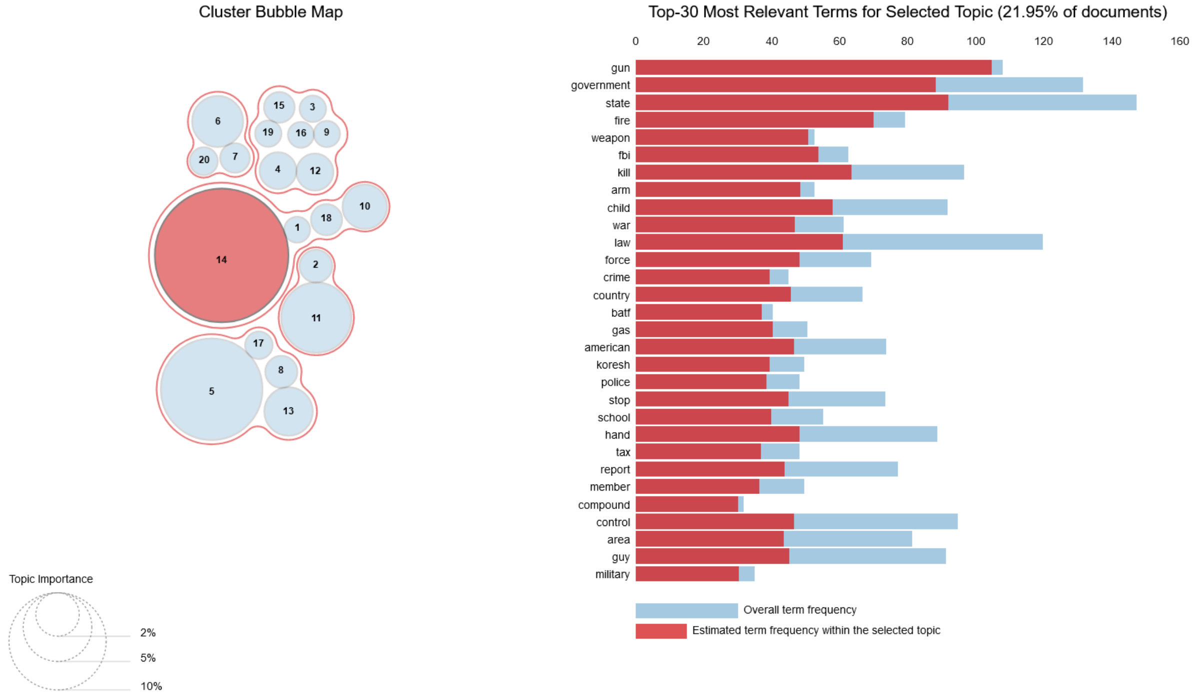


FIGURE 11: FINAL BUBBLEMAP INTERFACE

3.2.4. Control

To train participants in theme abstraction we produced a control interface (Appendix A: Fig. L, *Fig. 12)* which adhered to none of the GPs described in *Section 3.1*. Within the control visualisation, topic size encoding was removed, and topic marks were ordered in a 4 x 5 grid, thus eliminating our independent variables. This was achieved by altering the JSON data file loaded by the display. Any topic analysis would have to be performed only by looking at the list of keywords for each topic.

Finally, the topic number labels were modified between all three interfaces to mitigate the learning effects between conditions (Topic 14 has different keywords in *Fig. 10, 11* and *12*).

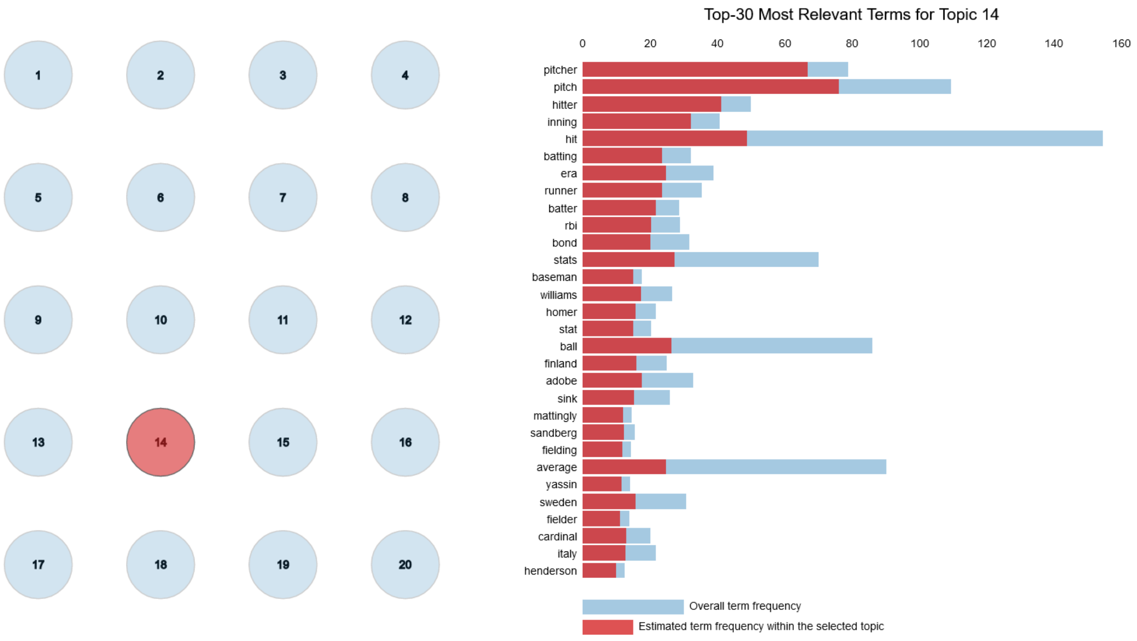


FIGURE 12: FINAL TRAINING INTERFACE

3.3. Study Design

To evaluate the usability of interfaces we recruited participants with little to no understanding of topic modelling and data science. We chose a within-participant approach for the experiment to reduce the number of participants required for a substantial study. Our experiment required users to perform several tasks investigating the thematic structure of a corpus.

3.3.1. Phase 1 - Usability Testing

Amar, et al., (2005) have used *affinity diagramming* to group similar questions often used in evaluating information visualisation systems. These groups include *retrieving a value, filtering, finding extreme cases, clustering* and *correlating cases*. We formulated usability tasks based on their research (Table 2.).

Table 2. Usability Testing Tasks

Stage	Task No.	Time	Task	Task Type
Training Control Interface	T0a.	4 min	Can you identify all topics that fall under the category of "Computer Programming"?	**Filter**
	T0b.	2 min	Out of these "Computer Programming" topics, which two are the most similar?	**Filter**
First Condition LDAvis & BubbleMap	T1a.	4 min	Can you identify all themes that topics can fall under in this collection of documents?	**Cluster**
	T1b.	2 min	Can you identify the biggest theme that topics can fall under in this collection of documents?	**Find Extremum**
	T1c.	4 min	Can you identify the two most similar topics in this whole collection of documents?	**Filter**
	T1d.	2 min	Can you identify the topic that mentions "Greece"?	**Retrieve Value**
Second Condition LDAvis & BubbleMap	T2a.	2.5 min	Given this new representation would you change any of the themes (categories) that you previously mentioned? Why or why not?	**Cluster**
	T2b.	1 min	Given this new representation would you change which the biggest theme (category) in this collection of documents is? Why or why not?	**Find Extremum**
	T2c.	2.5 min	Can you identify the two most similar topics in this representation of the collection of documents?	**Filter**
	T2d.	2 min	Can you identify the topic that mentions "yoghurt"?	**Retrieve Value**

TABLE 2: USABILITY TESTING TASKS

We designed sessions to last around 30 minutes maximum to appeal to a broader group of potential participants and to make the whole experience less cognitively strenuous. Time limits were devised, and tasks were finalised after a pilot run of the testing phase with two participants.

From this run, we concluded that users would learn more about the corpus as they completed more tasks. This is why we revised some of the second condition tasks to require the participant to think about the new mapping of the topic space presented and reflect on how it affects their interpretation of the corpus.

3.3.2. Phase 2 - Qualitative Evaluation

The *IBM Post-Study System Usability Questionnaire* (PSSUQ) is a 19-item instrument for assessing user satisfaction with system usability via *psychometric methods*. The goal of psychometrics is to establish the quality of psychological measures in terms of *reliability*, *validity* and *sensitivity to experimental manipulations* (Lewis, 1995).

The questionnaire is given to participants after they have completed all scenarios in a usability study. The questionnaire requires participants to rate statements out of 7 and justify their choice.

We used the PSSUQ to form our usability evaluation questionnaire (Table 3.). In our case, however, we are evaluating an interactive visualisation dashboard, we are not evaluating one of the complex systems this questionnaire was developed for. Additionally, to fit the entire run of the session within the desired time limit, the PSSUQ was condensed[5] down to 8 questions.

The scoring was also combined to simplify the protocol. Instead of rating each interface on a scale participants were asked to choose one interface over the other and justify their choice. Neutral options for answers such as "equally both" and "neither nor" were also added to allow for positive-neutral and negative-neutral answers.

[5] some questions regarding the complexity of the system were omitted as non-applicable.

Table 3. Evaluation Questionnaire
UE1. **Which representation of the topic space would you say is simpler?**
Based on PSSUQ 1, PSSUQ 2 and PPSUQ 12.
UE2. **Which interface allowed you to complete the tasks more effectively, (i.e. with fewer mistakes in the process)?**
Based on PSSUQ 3.
UE3. **Which interface allowed you to complete the tasks more quickly?**
Based on PSSUQ 4.
UE4. **Which interfaced allowed you to complete the tasks more efficiently, (i.e. with less repeated actions, like observing a particular topic multiple times)?**
Based on PSSUQ 5.
UE5. **Which interface felt more comfortable and pleasant to use?**
Based on PSSUQ 6, PSSUQ 16 and PSSUQ 17.
UE6. **Which representation of the topic space would you say feels more truthful?**
Several questions of the original PSSUQ are considering error handling and recovery. In development, we observed a few artefacts present in both interfaces. **LDAvis**: the multi-dimensional scaling would project topics next to each other which might lead the user into thinking they are similar, when it could be argued that this is not the case. **BubbleMap**: 7 small topics were clustered together when semantically they could have been put into 3 different thematic groupings.
UE7. **Which interface do you think explains more about the collection of documents?**
Another question that summarises several of the original PSSUQ. This time it is regarding the ability of the interface to convey information required by the user to complete their tasks.
UE8. **Which interface do you prefer, overall?**
Based on PSSUQ 19.

TABLE 3: QUALITATIVE EVALUATION QUESTIONNAIRE

Most of the questions target different aspects of interface usability and as such are investigating hypothesis H1. However, UE6. and UE7. are concerned with the veracity of the visual representation, thus seeking to prove or disprove hypotheses H2. and H3.

The next chapter will briefly describe the procedure of undertaking this study and the findings gathered upon completion.

Chapter 4

Results

4.1. Procedure Overview

4.1.1. Testing Environment and Risk Assessment

We conducted live sessions to observe participants' interaction with the interfaces and control the experimental environment better. A location within the limits of central Edinburgh was secured on the premises of a hospitality venue to increase appeal to potential participants by convenience (Appendix B).

We experimented during the usual off-peak business hours in a private room, using a Dell Vostro 3510 15" laptop computer and a standard USB office mouse. The interfaces were loaded through Visual Studio Code IDE, on a local server and displayed using Mozilla Firefox v. 128.0 running on Windows 10 Pro 64-bit.

No physical risk assessment was required as the venue has a fire safety policy.

However, ethical issues regarding participants giving informed consent to participate, volunteering their data as well as being exposed to language discussing controversial topics (such as political and religious violence) required measures to be put in place both during recruitment and in the set-up stage of the individual session to avoid causing them harm and distress.

4.1.2. Participant Recruitment

The experiment was advertised using the hospitality business' internal communication channels as this industry attracts people from various backgrounds, which allowed us to generate a pool of participants representative of the general public, specifically regarding the understanding of topic modelling and data science.

Moreover, the venue employs more than 80 staff, which sped up the recruitment process. It started at the beginning of July 2024, with testing planned between 15.07 and 29.07. After a person expressed interest in participating, they would inform us of their availability. Appointments were made and an email confirmation of the booking was sent out which also contained the Information Sheet (Appendix C.) and Consent Form (Appendix D.).

These documents outlined the experiment protocol and the data collection and participation withdrawal policies: the demographic data collection is optional, the participant input would be entirely anonymised, and they can withdraw from the experiment at any time without giving any reason.

By the fourth day of testing 17 members took part in the evaluation. Preliminary analysis of the results showed we had reached saturation with participant input. We decided to only run the experiment until the end of the same week.

In total 22 participants took part in the study. The average age of participants was 29.5 years old, 12/22 identified as male and 10/22 as female. About 32% of participants have obtained a degree in higher education. The average scores of the self-reported prior knowledge query were 5.09 confidence in computer usage, 4.46 for semantic text analysis skills, 4.14 confidence in interactive data charts, and 2.41 for computer programming knowledge. The scale to rate oneself was described to participants as 0 corresponding to having no knowledge of the matter and 7 being proficient in it.

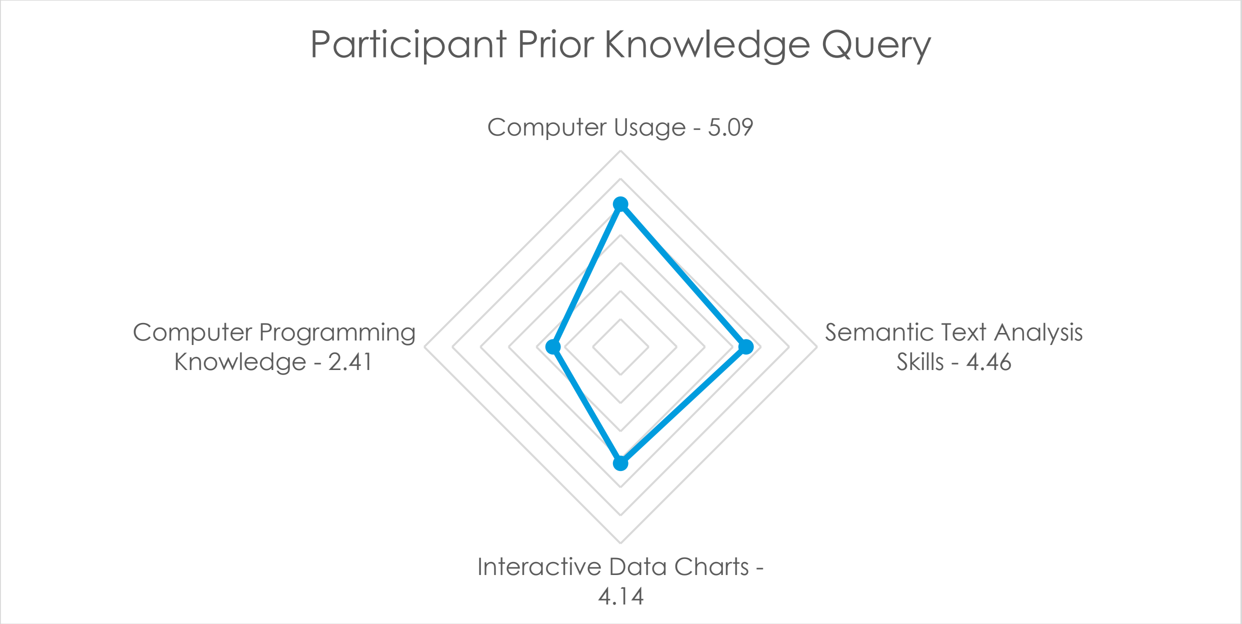


FIGURE 13: PARTICIPANT PRIOR KNOWLEDGE

4.1.3. Testing Procedure

As mentioned previously learning effects were identified as an issue in the pilot run. Ten out of 22 participants also acknowledged this in their thematic analysis.

We considered utilising a second dataset to eliminate these effects. However, the desired properties of such data, i.e. number of entries, length of documents, thematic structure and content were too specific to produce similar conditions.

We chose a within-participant approach in conducting this experiment, which would also mitigate the impact of these effects on the results. Participants were randomly split into two even groups – A and B, where the order of interaction between conditions was swapped. Group A were shown the *LDAvis* variant first, while Group B started with *BubbleMap*.

Each participant, at the set-up stage, were required to first sign the Consent Form, given a unique participant number and read a script (Appendix E.) which gave them more detailed information about the course of the session and some understanding about how these interfaces were produced.

At this stage, it was reiterated that the corpus contains controversial themes, and they can terminate their involvement at any point with no reason provided.

During the training and testing stages, participants were read aloud the task and the time limit they had to complete it. The researcher leading the session assisted them by taking notes on paper, allowing participants to concentrate on the thematic analysis. For that same reason, the Thinking-aloud technique was not enforced, but any comments made during this stage were noted down and added to the qualitative evaluation input.

In the evaluation stage, participants were interviewed using the questionnaire in Table 3. Their answers were recorded on paper, alongside input from the testing phase.

4.2. Phase 1 - Usability Testing

The themes discovered were standardised among participants, e.g. various users mentioned *medicine, health/diet, and treatment* as a theme which were processed under the common label *health*. Topic label integers were all reverted to their original state.

In the training phase, most participants named between 4 and 6 topics to fit under the "Computer Programming" theme within a widely varied time. The quickest being 60s and the longest 244s.

In terms of similarity 3/11 and 4/11 members of Groups A and B respectively would name a pair of topics to be most similar to each other and would later pair up one of these topics with another one from the same theme. This indicates the participant's expertise in the field parallel to the self-reporting in the set-up stage.

4.2.1. Group A

Most participants named 6 - 7 themes when presented with the *LDAvis* interface. Four participants out of 11 named 7 themes, while 3/11 users said there were 6 themes. One of the four, however, labelled "computing" and "internet" as two separate themes. It could be argued that they are the same.

Only one participant named 5 themes when presented with the projection mapping. However, when the cluster *BubbleMap* was shown, the number of participants who discovered 5 themes grew to 4. In general, when asked if they wanted to amend their previous answer, every but one participant were happy to do so. This user, however, had no interaction with the second interface.

This resulted in people naming between 5 – 6 themes in total, which shows a slight decrease in the number of thematic groupings uncovered. Six out of 11 participants mentioned the evident grouping of topics in clusters as a reason for this change, 1 additionally mentioning they have eventually learned more about the corpus. Three more users mentioned the learning effect as the sole reason for amending their answers.

Regarding choosing the biggest theme in the corpus, just over half of the participants changed their answers when presented with *BubbleMap*. A substantial number of users - 8/11 mentioned the visual clusters as reasons to change or confirm their original beliefs.

Regarding topic similarity, only 2/11 users named the same pairs of topics between both conditions. However, when presented with the *BubbleMap* participants were generally more consistent with naming pairs of similar topics as evidenced by the summary table below (Table 4.).

Table 4. Group A – Task Results							
Number of Themes			**Most Similar Topics**		**Locate Keyword**		
# themes	**LDAvis**	**BubbleMap**	**LDAvis**	**BubbleMap**	**time**	**LDAvis**	**BubbleMap**
four	2	-	2, 6	3, 9	Under 10s	5/11	2/11
five	1	4	3, 10	3, 5	10 – 40s	-	6/11
six	3	4	10, 16	10, 16	40 – 80s	4/11	1/11
seven	4	2	6, 11	10, 16	80s +	2/11	2/11
eight	-	1	2, 6	10, 16			
nine	1	-	1, 3	4, 14			
ten	-	-	11, 15	4, 14			
eleven	-	-	10, 16	9, 16			
			13, 20	10, 16			
			3, 10	3, 9			
			3, 9	3, 9			

TABLE 4: GROUP A - TASK RESULTS

Finally, when tasked to locate the keyword "greece" 5/11 users managed to do so in under 10s when presented with *LDAvis*, 4/11 took between 40 – 80s and 2 users managed in more than 80s.

When presented with the *BubbleMap*, only 2/11 succeeded in finding the word "yoghurt" in under 10s, 6/11 took between 10 – 40s to do so, one did in 60s and 2 took more than 80s.

4.2.2. Group B

Most participants named 5 – 6 themes when presented with the *BubbleMap* interface. Five out of 11 decided there were 5 groupings, corresponding to the number of visible clusters, indicated by a red border. Three out of 11 said there were 6.

Only 2 participants who discovered 5 themes under the first condition maintained their original beliefs when this group saw the projection mapping. However, 1 of these two mentioned that the *LDAvis* layout was confusing and the other had no interaction with the interface.

A total of 5/11 participants were happy with their initial input. Nine group members mentioned learning effects as reasons for both changing and maintaining their beliefs, i.e. they have either learned enough about the corpus with their interface interaction under the first condition or through more topic analysis have decided to amend the number of topics and their labels. Only 1 participant split the theme "American culture" into "war" and "gun culture", based on the relative distance of topic marks on screen.

Regarding choosing the biggest theme in the corpus only 1 participant made a partial change (as they split their named biggest theme into two). Ten out of the eleven users were content with their initial answer, 5 of them mentioning topic 1 (labelled "7" in *LDAvis,* has the biggest mark) as a reason for confirming their beliefs and 2 additionally mentioning the overlap between topics 1 and 3 (labelled "7" and "10" in *LDAvis*) as well as the size of the circles on the screen.

One user, who named "computing" as the biggest theme mentioned the fact that topic 5 (labelled "6" in *LDAvis* and "10" in *BubbleMap)* is not constrained into a cluster and so can be added to this category. This topic is likely to be derived from a newsgroup about encryption.

As for topic similarity, 4/11 users named the same pairs of topics between both conditions. Regarding selecting the same pairs consistently between participants, we cannot say that either interface had a noticeable effect on this group as the count of pairs of similar topics was 8 for both stages.

Finally, when tasked to locate the keyword "greece" 5/11 users managed to do so in under 10s when presented with the BubbleMap, 5/11 took between 10 – 40s and one user managed in more than 40s.

When presented with *LDAvis*, only 1/11 succeeded in finding the word "yoghurt" in under 10s, 3/11 took between 10 – 40s to do so, 4 did in between 40 – 80s and 3 took more than 80s. The results for topic similarity pairs and locating keyword summaries can be seen in the table below (Table X.).

Table 5. Group A – Task Results							
Number of Themes			**Most Similar Topics**		**Locate Keyword**		
# themes	**LDAvis**	**BubbleMap**	**LDAvis**	**BubbleMap**	**time**	**LDAvis**	**BubbleMap**
four	1	-	3, 10	3, 10	Under 10s	1/11	5/11
five	2	5	1, 3	3, 9	10 – 40s	3/11	5/11
six	4	3	2, 11	3, 9	40 – 80s	4/11	1/11
seven	2	1	2, 6	2, 6	80s +	3/11	-
eight	-	-	2, 6	2, 6			
nine	1	1	2, 11	6, 11			
ten	1		14, 16	1, 2			
eleven	-	1	13, 20	4, 14			
			6, 15	6, 15			
			7, 8	10, 16			
			7, 8	2, 6			

TABLE 5: GROUP B - TASK RESULTS

4.2.3. Results Summary

When presented with 5 visually separated clusters, users tend to name 5 themes, even though our exploratory data analysis indicates that there are likely more themes than that. A total of 4 participants labelled 1 cluster as "random", as it contained topics formed from the *sports*, *auto* and *science* newsgroups.

Group B, tended to stick to their answers when presented with the alternative projection mapping, indicating that the cluster *BubbleMap* has a greater "teaching" capability.

Another possible proof for this claim might be that more people named the same pairs of similar topics in this group between conditions, possibly because they have an easier time committing them to memory.

Almost all users named topics in the same cluster to be most similar in both groups, one member of Group B seemed to have selected their answer based solely on the sizes of the circles, as one topic is clearly about "computing" while the other is about "politics".

In Group A, one member has made a similar non-sensical pairing, however, the reason for this is not apparent. Another user has paired topics in neighbouring clusters, which appear next to each other on the screen. One seems to be about the Middle East and the other about religion. We can say this is a valid interpretation of the thematic structure.

With *LDAvis* one member of each group pointed to topics 1 and 3 (labelled "7" and "10") as being most similar, however, it is likely they were formed from different newsgroups – *politics and religion.*

The marks of these two topics are quite large and overlap on screen, one appearing to "contain" the other. It could be argued how similar these topics are or if they should be in the same thematic category due to the subjectivity and nature of semantic analysis, but we believe that they should have been separated.

Two users in Group B confirmed that the biggest theme was "war", or "international relations" based on this visual cue. Moreover, 16/22 participants seem to have picked their most similar pair either because the topic marks are closest to each other on the screen or in the case of the four "computing" topics, the marks are within this cluster in the upper half of the map.

When looking for a particular keyword, we have to acknowledge the fact that while both "greece" and "yoghurt" are contained in topics of similar sizes and are located in similar positions in the word list, the former is part of a larger theme, and the latter is part of a small one. We also need to consider that finding the topic number for the word "yoghurt" is the last task of the testing phase, therefore fatigue might come into play. Despite, those two factors, people tend to be quicker using the *BubbleMap* in both stages of the testing phase.

The results of this phase indicate empirically that *BubbleMap* is easier to use for our selected group of participants. However, both interfaces could lead to some errors in interpretation.

The results of the qualitative evaluation phase, however, will prove or disprove the hypotheses posed in the previous chapter.

4.3. Phase 2 – Qualitative Evaluation

As with task input, the comments participants gave when interviewed were standardised, to summarise and interpret them better in addition to observing the answers to the questionnaire.

For example, users mentioned that one interface felt "less restricted", "painted a broader picture" and "allowed you to rely more on your judgment". These comments were abstracted with the label *unrestricted*. Comments that duplicated answers from the questionnaire were omitted in the analysis. A summary of the standardised comment counts can be found at the end of this section (Table 8.).

4.3.1. Group A

Every member of Group A preferred *BubbleMap* over LDAvis when answering the first 5 questions of the interview. However, Question UE6., asking which of the two interfaces feels more truthful, has 7/11 answers in favour of *LDAvis*, 2/11 positive neutral answers and 2/11 in favour of the BubbleMap.

Question UE7. asks which of the two explains more about the corpus, the answers being split about halfway with 5/11 in favour of *LDAvis* and 6/11 for the *BubbleMap*. Four of the former also think LDAvis is more truthful.

Table 6. Group A – Qualitative Evaluation

ue1	ue2	ue3	ue4	ue5	ue6	ue7	ue8	comment type		testers, who made at least one comment
B	B	B	B	B	L	L	**B**	LDAvis positive comment:		7/11
B	B	B	B	B	B	B	**B**	LDAvis negative comment:		11/11
B	B	B	B	B	=	B	**B**			
B	B	B	B	B	=	B	**B**	BubbleMap positive comment:		11/11
B	B	B	B	B	L	L	**B**	BubbleMap negative comment:		5/11
B	B	B	B	B	L	B	**B**			
B	B	B	B	B	B	L	**B**	**legend:**		
B	B	B	B	B	L	B	**B**		N	neither one
B	B	B	B	B	L	L	=		=	equally both
B	B	B	B	B	L	L	**B**		L	LDAvis
B	B	B	B	B	L	B	**B**		B	BubbleMap

TABLE 6: GROUP A - QUALITATIVE EVALUATION

Only 1 participant said they preferred both interfaces the same overall, depending on the use case – LDAvis being more data-focused and the BubbleMap being user-focused. All other members of the group quoted the BubbleMap as their favourite in general.

All these direct answers are supported by the number of positive and negative comments made about each interface (Table 6.). Seven users had something positive to say about *LDAvis* and 5/11 had something negative to say about the *BubbleMap*. All users had negative comments about the former and positive about the latter.

Four participants in this group rendered LDAvis as *unrestricted*, 5/11 said it feels more *accurate*, 2/11 said that *proximity gives a good indication of similarity* and 1/11 that this mapping is more *familiar*. A different user recognised that this is a 2D mapping but were confused with the encoding of the axes. Additionally, 6/11 said that the *overlap* of topic marks hinders the user experience, 5/11 said interacting with LDAvis was *more work* and *frustrating*, and 8/11 said it was *messy*.

Nine users in this group mentioned that the *evident grouping* is helpful when interpreting the themes in the BubbleMap interface, 7/11 said this interface was more *navigable, 4/11* described it as more *organised,* and 7/11 specifically enjoyed the that it is *compact*. Four out of eleven said this interface was more *memorable,* 7/11 described it as *neat,* and 4/11 said it was *intuitive* to use, i.e. it gives you direction. Other occasional comments rendered it more *readable* and *aesthetically pleasing*.

Five participants in Group A said that while the evident grouping in the *BubbleMap* was helpful, it could also be misleading. One user described it as *constrained* negatively.

Finally, 3 people in this group indicated that the two interfaces would suit different use-case scenarios.

4.3.2. Group B

In Group B people leaned a little more towards *LDAvis* when answering the first 5 questions, but generally still preferred the *BubbleMap* over it (Table 7.). Two participants answered question UE4. which asks them to pick the more efficient interface by naming *LDAvis*, however, they also stated that they believed this was so, because they saw that interface as a second condition. The other two who gave the same answer were the only two participants who chose *LDAvis* as their preferred interface overall.

Answers to question UE6. are the same as from the previous group. Question UE7., however, has some neutral positive answers and fewer *LDAvis* choices given as answers.

Table 7. Group B – Qualitative Evaluation

ue1	ue2	ue3	ue4	ue5	ue6	ue7	ue8	comment type		testers, who made at least one comment
B	B	B	B	B	L	L	**B**	LDAvis positive comment:		9/11
B	B	B	B	B	L	B	**B**	LDAvis negative comment:		8/11
B	B	B	L	B	B	=	**B**			
B	B	B	L	B	L	B	**B**	BubbleMap positive comment:		11/11
B	B	B	B	L	L	B	**B**	BubbleMap negative comment:		6/11
B	B	B	B	B	B	=	**B**			
B	B	L	B	B	=	=	**B**	**Legend**		
L	N	=	L	L	=	B	**L**		N	neither one
B	B	B	B	=	L	L	**B**		=	equally both
B	B	B	B	B	L	B	**B**		L	LDAvis
N	L	L	L	B	L	=	**L**		B	BubbleMap

TABLE 7: GROUP B - QUALITATIVE EVALUATION

From the table above we can see that within this group the number of participants who gave positive comments about *LDAvis* has increased by 2 and the number that gave negative comments about the *BubbleMap* has increased by 1. The number of negative comments about *LDAvis* has decreased by 3.

Four participants said that *LDAvis* felt more *accurate* than the *BubbleMap*. Moreover, two participants said themes were more *interpretable,* 2/11 said this mapping was more *familiar,* 3/11 said *proximity gives a good indication of similarity, 1/11* described it as *neat* and 2/11 found LDAvis *aesthetically pleasing*.

Finally, 2/11 said that the overlap of the marks was a better representation of the data, one of them adding that it was also easier to compare topics when they overlapped. Another user also mentioned it was easier for him to *compare* topics with LDAvis.

The overlap was hindering the user experience in this group for 7/11 participants. This led to one of them also saying the projection mapping was *more work* and another saying it was *misleading*. Three users found *LDAvis messy* and 4/11 described it as *frustrating*.

Eight participants mentioned that the *evident grouping* is helpful when interpreting the themes using the *BubbleMap* interface in this group, 6/11 said this interface was more *navigable, 3/11* described it as *organised, and* 4/11 specifically enjoyed the that it is *compact*. Between 1 – 3 users in this group described *BubbleMap* as *memorable, neat, intuitive, aesthetically pleasing* and more *readable.*

In Group B 2/11 participants said the grouping might be *misleading*. One said it was *more work* and 2/11 found it *frustrating*. One did not find any indication of similarity between topics and another one said it was not obvious what the red lines represent.

One additional user separated the interfaces depending on the use case scenario.

Table 8. Standardised Comments Summary

comment type	standardised comment	A count	B count	Total	% / Total
LDAvis positive	*unrestricted*	4/11	-	4/22	18.18
	accurate	5/11	4/11	9/22	40.91
	proximity -> similarity	2/11	3/11	5/22	22.73
	familiar	1/11	2/11	3/22	13.64
	aesthetically pleasing	-	2/11	2/22	9.09
	interpretable themes	-	2/11	2/22	9.09
	neat	-	1/11	1/22	4.55
	easier to compare topics	-	2/11	2/22	9.09
	overlap -> interpretable	-	2/11	2/22	9.09
LDAvis negative	*overlap is hindering*	6/11	7/11	13/22	59.09
	more work	5/11	1/11	6/22	27.27
	frustrating	5/11	4/11	9/22	40.91
	messy	8/11	3/11	11/22	50
	misleading	-	1/11	1/22	4.55
BubbleMap positive	*evident grouping -> interpretable*	9/11	8/11	17/22	77.27
	navigable	7/11	6/11	13/22	59.09
	organised	4/11	3/11	7/22	31.82
	compact	7/11	4/11	11/22	50
	memorable	4/11	3/11	7/22	31.82
	neat	7/11	2/11	9/22	40.91
	intuitive -> direction	4/11	2/11	6/22	27.27
	readable	4/11	3/11	7/22	31.82
	aesthetically pleasing	3/11	1/11	4/22	18.18
BubbleMap negative	*misleading*	4/11	2/11	6/22	27.27
	irrelevant topics	1/11	-	1/22	4.55
	constrained	1/11	-	1/22	4.55
	frustrating	-	2/11	2/22	9.09
	more work	-	1/11	1/22	4.55
Neutral	*use case dependent*	3/11	1/11	4/22	18.18

TABLE 8: STANDARDISED COMMENTS SUMMARY

4.2.3. Results Summary

We can see a decrease in the number of positive comments about the *BubbleMap* and negative ones about *LDAvis* parallel to an increase in the number of positive comments about *LDAvis* in Group B compared to Group A.

In the interview stage, Group A chose *LDAvis* 12 times when answering the questions compared to 19 times for Group B. The number of neutral answers also increased from 2 to 10. The two participants who preferred *LDAvis* overall are a part of Group B.

Since Group B is the group that saw *LDAvis* last, their perception of the interface could have been impacted by the learning effect. Even so, they still picked the *BubbleMap* as the better interface 48 times in total, compared to 19 for *LDAvis* when answering the questions.

The *BubbleMap* was also the overall preferred interface as self-reported by participants when answering UE8. – 19/22 times, compared to 2/22 for *LDAvis*. The only evident aspect where *LDAvis* has an advantage is the perceived truthfulness of the interface. When we look at the answers given to question UE6. LDAvis is picked as the answer 14/22 times, compared to 4/22 for the BubbleMap.

Question UE7. is more abstract, as inter-topic similarities are just one of the features of corpus structure. The results from this question are mixed: 7/22 said LDAvis explains more, 11/22 said the BubbleMap does and 4/22 gave a neutral answer. This suggests that the participants in this study value clustering topics more than seeing the topic relationships encoded in distance.

On the other hand, 27.27% of participants said that the *BubbleMap* can be *misleading*, while only 4.55% said that about *LDAvis*. Nearly forty-one per cent said it was *accurate*, a standardised label depicting truthfulness. No one described the *BubbleMap* in such a way.

The most common positive comment (77.27%) is that the *evident grouping* in the BubbleMap makes the representation more *interpretable*. The most common negative comment (59.09%) is that the *overlap* of marks in *LDAvis hinders the* user experience.

Chapter 5

Discussion

In *Section 3.1* we set out three hypotheses to be investigated by the qualitative evaluation of the user experience in interacting with the *LDAvis* and *BubbleMap*-based interfaces. Twenty-two participants divided into two groups answered 8 questions and gave additional comments about their interactions with the dashboards.

The results from *Phase 1 – Usability Testing* indicate that the BubbleMap interface is more intuitive as people are more consistent with their answers and quicker in performing some tasks. The results from *Phase 2 – Qualitative Evaluation* further support this claim, even though *LDAvis* is rated slightly higher by Group B, possibly due to learning effects and the order of presentation.

The tasks we developed were focused on interpreting the thematic structure of the corpus. Users *filtered, clustered, found the extremities* and *retrieved values* of topics and keywords based on the more abstract themes.

5.1. Results Analysis

5.1.1. Thematic Structure Interpretability - H1.

Our *hypothesis H1*. suggests that a cluster-based approach such as our produced *BubbleMap* would facilitate the performance of these tasks better and so would be the preferred method for thematic analysis, over the alternative – a projection-based layout.

We can conclude that this hypothesis is *true* based on our post-testing evaluation. Most participants (19/22) answered question UE8., which is regarding overall preference by naming the *BubbleMap* interface. Moreover, the *BubbleMap* was preferred when answering the first 7 questions: 111/154 answers in favour of the cluster-based approach, 31/154 for *LDAvis* and 12/154 neutral answers.

This is also supported by the number of comments about both interfaces: *BubbleMap* has positive comments from more users and negative ones from fewer. Finally, the analysis of the answers to UE7. suggests that participants prefer evident clustering, over inter-topic relationship encoding.

5.1.2. Perceived Truthfulness - H2.

Hypothesis H2. revolves around the fact, that a projection-based mapping aims to retain inter-topic relationships by encoding them within the distance between the topic marks, thus emphasising *truthfulness*. We targeted this hypothesis with questions UE6. and UE7., the first directly asking which interface felt more truthful and the second asking participants which interface explained more about the corpus.

Sixteen participants in the study said LDAvis felt more truthful with sporadic comments that "raw data is messy" or another similar sentiment. These were standardised under the label *accurate* (9/22 mentions), thus also proving H2 to be *true*.

The fact that *LDAvis* feels more truthful to users is further supported by the number of people that labelled either interface as *misleading*: 1/22 for LDAvis and 6/22 for the *BubbleMap*.

The same number of participants – 6/22 said that the BubbleMap is more *intuitive*, further clarifying that it "gives you direction" and 59% praised the cluster-based approach for its *navigability*. From our usability testing, we can see that people are more inclined to name 5 themes in total, corresponding to 5 clusters on the screen, even though it could be argued that the actual themes in the collection are more than that.

In development, we had to decide how many clusters to produce in the visualisation, based on cluster density. Some users have mentioned that a benefit of *BubbleMap* is that "some of the work is done for you", which can explain why people perceive *LDAvis* as more truthful. I*ntroducing bias* should be under consideration when producing BubbleMap-based interfaces.

5.1.3. Non-expert User Preference – H3.

Our last *hypothesis H3*. considered the use-case scenario and users' background expertise. We asked our participants to self-report some demographic data, one query was regarding their confidence in their computer programming knowledge. With an average score of 2.41/7, we can conclude that this is a group of *non-expert* users.

By proving hypotheses H1. and H2. as well as semantically analysing the types of comments participants gave in their qualitative evaluation, we can see that, while around 40% of respondents recognised *LDAvis* as *accurate,* the same number of participants liked the fact that the BubbleMap is *neat* and even a bigger part like that it is *navigable, compact* and say that the *evident grouping* is helpful.

We can conclude that H3. is *true,* i.e. *non-expert users prefer interpretability over usability.*

5.2. Study Strengths and Limitations

When reviewing the relevant literature, we noted the lack of substantial studies on the usability of topic model visualisation designs. The experiment we conducted included over 20 participants, even though we decided to cut the testing short as we reached input saturation.

The pool of participants was limited to people with little prior knowledge of computer programming. Even though we can say that non-expert users prefer the cluster-based approach, we only have indications about the perceived truthfulness of those interfaces.

Our survey showed that the projection-based approach feels more truthful, and 3/22 participants mentioned that this type of interface would be better for performing semantic analysis. A parallel comment made by one of these users about the *BubbleMap* was that it is better suited for browsing the corpus.

There is misleading information in both interfaces. However, LDAvis was seldom pointed out for this. A more technical user might discover this insight.

Finally, we limited the analysis of the usability testing phase to only answers to task prompts and timings to complete the tasks. Recorded mouse input, for example, would have provided input data for more in-depth analysis.

Despite its limitations, we hope this study is a good addition to the conversation about the usability evaluation of ML output visualisations and the ability of different audiences to perceive that information.

Chapter 6

Conclusion

In this dissertation, we wanted to investigate the usability and interpretability of two topic model visualisation interfaces: *LDAvis*, a projection-based layout, and *BubbleMap*, a cluster-based layout. Our research aimed to determine which interface is more effective for non-expert users considering three directions: thematic structure interpretation, perceived truthfulness, and overall user preference.

6.1. Key Findings

6.1.1. Thematic Structure Interpretability

We confirmed that the *BubbleMap* interface – a cluster-based approach to topic-space mapping is preferred over *LDAvis, the projection-based strategy* when users try to interpret thematic structures. This was evident from the fact that the majority of participants found the *BubbleMap* easier to use, more intuitive, and quicker for task completion. Participants appreciated the evident grouping of topics, which made the thematic analysis more accessible and reduced cognitive load.

6.1.2. Perceived Truthfulness

While the *BubbleMap* was preferred for its usability, *LDAvis* was perceived as more truthful. The projection-based approach retained inter-topic relationships, which participants associated with accuracy. This was particularly important for users who prioritised the fidelity of the data representation over ease of use. Despite some participants pointing out the limitations of *LDAvis*, such as the overlap of topic marks, it was still seen as providing a more accurate depiction of the corpus.

6.1.3. Non-Expert User Preference

The study supported the hypothesis that non-expert users prefer interfaces that offer higher interpretability over those that emphasise truthfulness. The *BubbleMap*, with its organised and navigable layout, was favoured by participants with limited programming knowledge. This preference highlights the importance of catering to the target audience's needs in designing visualisation tools, particularly when conveying complex, highly-dimensional data, such as thematic information.

6.2. Implications and Future Research

The results of this study suggest that when designing visualisations for non-expert users, it is crucial to prioritise usability and interpretability. While accuracy and truthfulness are important, they should not come at the expense of user experience, especially for audiences with limited technical backgrounds.

The findings also indicate that different visualisation tools may be more or less suitable depending on the tasks or user groups, *LDAvis* being more appropriate for detailed semantic analysis and *BubbleMap* for broader thematic exploration.

This study had several limitations, including a relatively small sample size and a pool of non-expert participants. Future research could explore the usability of these interfaces with expert users or in different contexts, such as specific industries or academic disciplines. Additionally, more detailed interaction data, such as mouse movements and click patterns, could provide deeper insights into how users navigate these interfaces.

6.3. Dissertation Conclusion

This dissertation contributes to the growing research on the usability of machine learning output visualisations. By comparing *LDAvis* and *BubbleMap*, we have highlighted the trade-offs between interpretability and truthfulness and provided insights into how non-expert users interact with these tools.

As visualisation technologies continue to evolve, it is essential to keep user experience at the forefront of design considerations, ensuring that complex data is presented in a way that is both accessible and accurate.

References

Amar, R., Eagan, J. & Stasko, J., 2005. *Low-Level Components of Analytic Activity in Information Visualization.* s.l., IEEE Computer Society, pp. 111-117.

Angelov, D., 2020. Top2Vec: Distributed Representations of Topics. s.l.:s.n.

Asghari, M., Sierra-Sosa, D. & Elmaghraby , A. S., 2020. A topic modelling framework for spatio-temporal information management. *Inf Process Manag,* 57(6).

Blandford, A., Cox, A. L. & Cairns, P., 2008. Controlled experiments. In: *Research Methods for Human-Computer Interaction.* s.l.:Cambridge University Press, pp. 1-16.

Blei, D. M., 2012. Probabilistic topic models. *Communications of the ACM,* 55(4), pp. 77-84.

Blei, D. M., Ng, Y. A. & Jordan, I. M., 2003. Latent Dirichlet allocation. *J. Mach. Learn. Res.,* Volume 3.

Buenaño-Fernandez, D., González, M., Gil, D. & Luján-Mora, S., 2020. Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach. *IEEE Access,* Volume 8, pp. 35318-35330.

Byron, L. & Wattenberg, M., 2008. Stacked Graphs – Geometry & Aesthetics. *IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS,* 14(6), pp. 1245-1252.

Chaney, A. J.-B. & Blei, D. M., 2012. *arXiv physics papers Topics.* [Online] Available at: https://www.cs.mcgill.ca/~isavov/arxiv_demo/browse/topic-presence.html
[Accessed 01 04 2024].

Chaney, A. J.-B. & Blei, D. M., 2012. Visualizing Topic Models. Proceedings of the International AAAI Conference on Web and Social Media.

Cheng, K. S. et al., 2020. TopExplorer: Tool Support for Extracting and Visualizing Topic Models in Bioengineering Text Corpora. Chicago, IL, USA, s.n., pp. 334 - 343.

Choi, M. et al., 2018. *TopicOnTiles: Tile-Based Spatio-Temporal Event Analytics.* Montreal, QC, Canada, Association for Computing Machinery.

Chuang, J., Manning, C. D. & Heer, J., 2012. *Termite: visualization techniques for assessing textual topic models.* Capri Island, Italy, Association for Computing Machinery, p. 74–77.

Egger, R. & Yu, J., 2022. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology,* Volume 7.

Gharavi, A., Le Bras, P. & Chantler, M., 2022. *Topic Map of UKRI Software Research Grants.* Edinburgh: Strategic Futures Lab. Provided by Pierre Le Bras, with permission from Mike Chantler, personal communication.

Gortler, J., Schulz, C., Weiskopf, D. & Deussen, O., 2018. Bubble Treemaps for Uncertainty Visualization. *IEEE Trans Vis Comput Graph,* 24(1), pp. 719-728.

Goswami, A., Mohapatra, P. & Zhai, C., 2019. *Quantifying and visualizing the demand and supply gap from e-commerce search data using topic models.* San Francisco, Association for Computing Machinery, pp. 348-353.

Greene, D. & Cunningham, P., 2005. *Producing Accurate Interpretable Clusters from High-Dimensional Data,* Dublin, Ireland: Trinity College Dublin, Department of Computer Science.

Griffiths, T. L. & Steyvers, M., 2004. Finding scientific topics. *Proceedings of the National Academy of Sciences of the United States of America,* Volume 101, pp. 5228-5235.

Grootendorst, M., 2022. BERTopic: Neural topic modelling with a class-based TF-IDF procedure. s.l.:arXiv.

Haidar, A. A., Yang, B. & Ganascia, J.-G., 2016. *Visualizing the First World War Using StreamGraphs and Information Extraction.* Lisbon, Portugal, s.n., pp. 290-293.

Han, S., Ye, S. & Zhang, H., 2020. Visual exploration of Internet news via sentiment score and topic models. *Comp. Visual Media,* Volume 6, pp. 333-347.

Havre, S., Hetzler, E., Whitney, P. & Nowell, L., 2002. ThemeRiver: visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics,* 8(1), pp. 9-20.

Helldin, T., Steinhauer, J. H., Karlsson, A. & Mathiason, G., 2018. *Situation Awareness in Telecommunication Networks Using Topic Modeling.* Cambridge, UK, s.n., pp. 549-556.

Koylu, C., 2019. Modeling and visualizing semantic and spatio-temporal. *International Journal of,* 33(4), pp. 805-832.

Kumari, R., Jeong, J. Y. & Choi, K., 2021. Topic modelling and social network analysis of publications and patents in humanoid robot technology. *Journal of Information Science,* 47(5), pp. 658-676.

Le Bras, P. et al., 2020. Visualising COVID-19 Research. *ArXiv.*

Lee, J. H. & Ostwald, M. J., 2024. Latent Dirichlet Allocation (LDA) topic models for Space Syntax studies on spatial experience. *City, Territory and Architecture,* Volume 11.

Liu, H. et al., 2020. Mapping the technology evolution path: a novel model for dynamic topic detection and tracking. *Scientometrics,* Volume 125, pp. 2043-2090.

Mabey, B., 2015. *pyLDAvis.* [Online] Available at: https://pypi.org/project/pyLDAvis/ [Accessed 30 03 2024].

Martin, M. E. & Schuurman, N., 2017. Area-Based Topic Modeling and Visualization of Social Media for Qualitative GIS. *Annals of the American Association of Geographers,* 107(5), pp. 1028-1039.

Maskat, R., Shaharudin, S., Witarsyah, D. & Mahdin, H., 2023. A Survey on Forms of Visualization and Tools Used in Topic Modelling. *International Journal on Informatics Visualization,* 7(2).

Odlum, M. et al., 2020. Application of Topic Modeling to Tweets as the Foundation for Health Disparity Research for COVID-19. *Stud Health Technol Inform,* Volume 272, pp. 24-27.

Padilla, S., Corne, D. W., Methven, T. S. & Chantler, J. M., 2014. Hot topics in CHI: trend maps for visualising research. In: *CHI '14 Extended Abstracts on Human Factors in Computing Systems.* Toronto, Ontario, Canada: Association for Computing Machinery, p. 815–824.

Sievert, C. & Shirley, K., 2015. *emo page for LDAvis: A method for visualizing and interpreting topics.* [Online] Available at: http://www.kennyshirley.com/LDAvis/ [Accessed 30 03 2024].

Sievert, C. & Shirley, K. E., 2014. *LDAvis: A method for visualizing and interpreting topics.* Baltimore, Maryland, USA, Association for Computational Linguistics, pp. 63-70.

Skeppstedt, M., Kucher, K., Stede, M. & Kerren, A., 2018. Topics2Themes: Computer-Assisted Argument Extraction by Visual Analysis of Important Topics. Paris, France, s.n., pp. 9-16.

Thalmann, M., Souze, A. S. & Oberauer, K., 2019. How does chunking help working memory?. *J Exp Psychol Learn Mem Cogn,* 45(1), pp. 37-55.

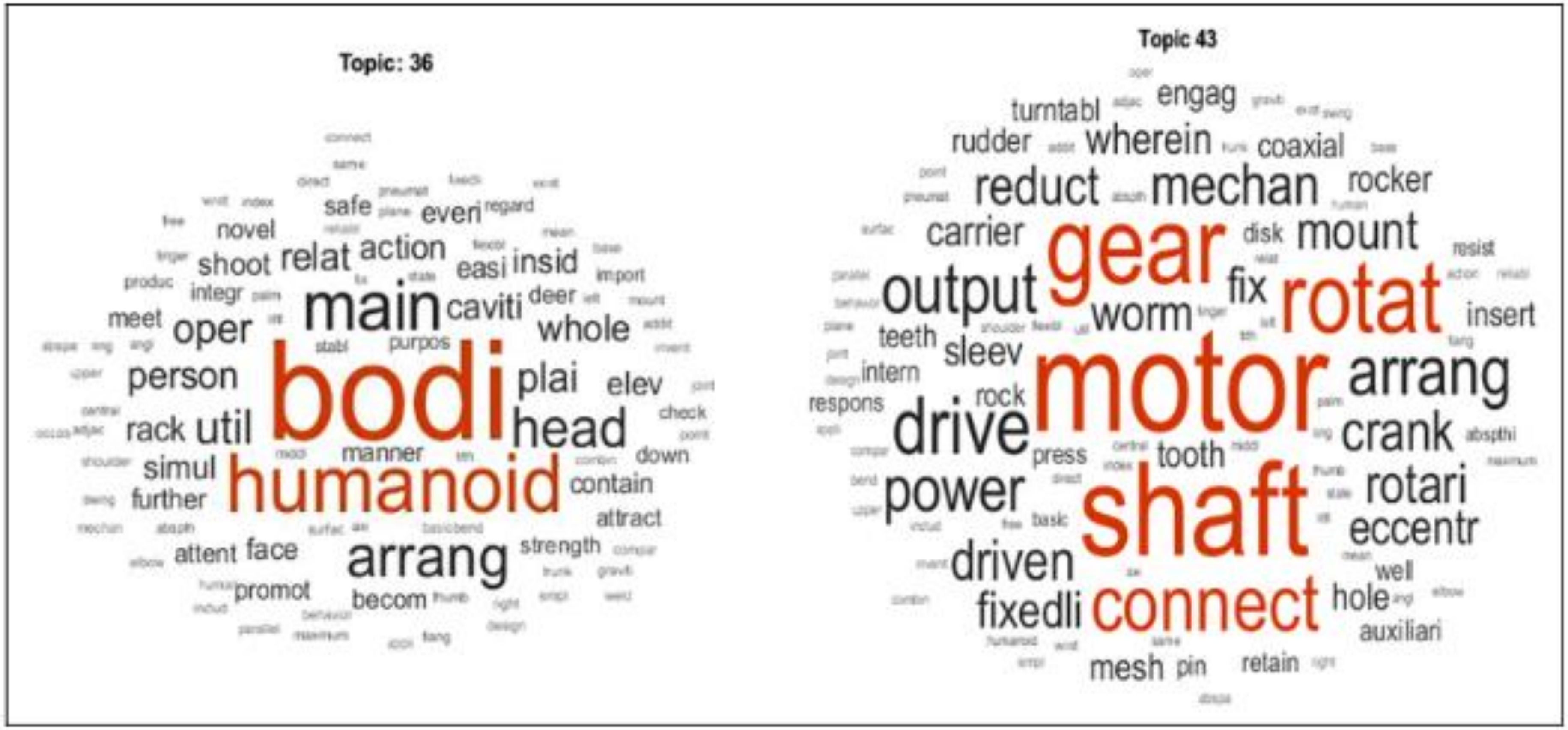
Wagemans, J. et al., 2012. A Century of Gestalt Psychology in Visual Perception I. Perceptual Grouping and Figure-Ground Organization. *Psychol Bull,* 30 Jul, 138(6), pp. 1172-1217.

Won, J., Kim, K. & Sohng, K.-Y., 2021. Trends in Nursing Research on Infections: Semantic Network Analysis and Topic Modeling. *Int. J. Environ. Res. Public Health,* 18(13), p. 6915.

Zhang, T. et al., 2020. Multi-Dimension Topic Mining Based on Hierarchical Semantic Graph Model. *IEEE Access,* Volume 8, pp. 64820-64835.

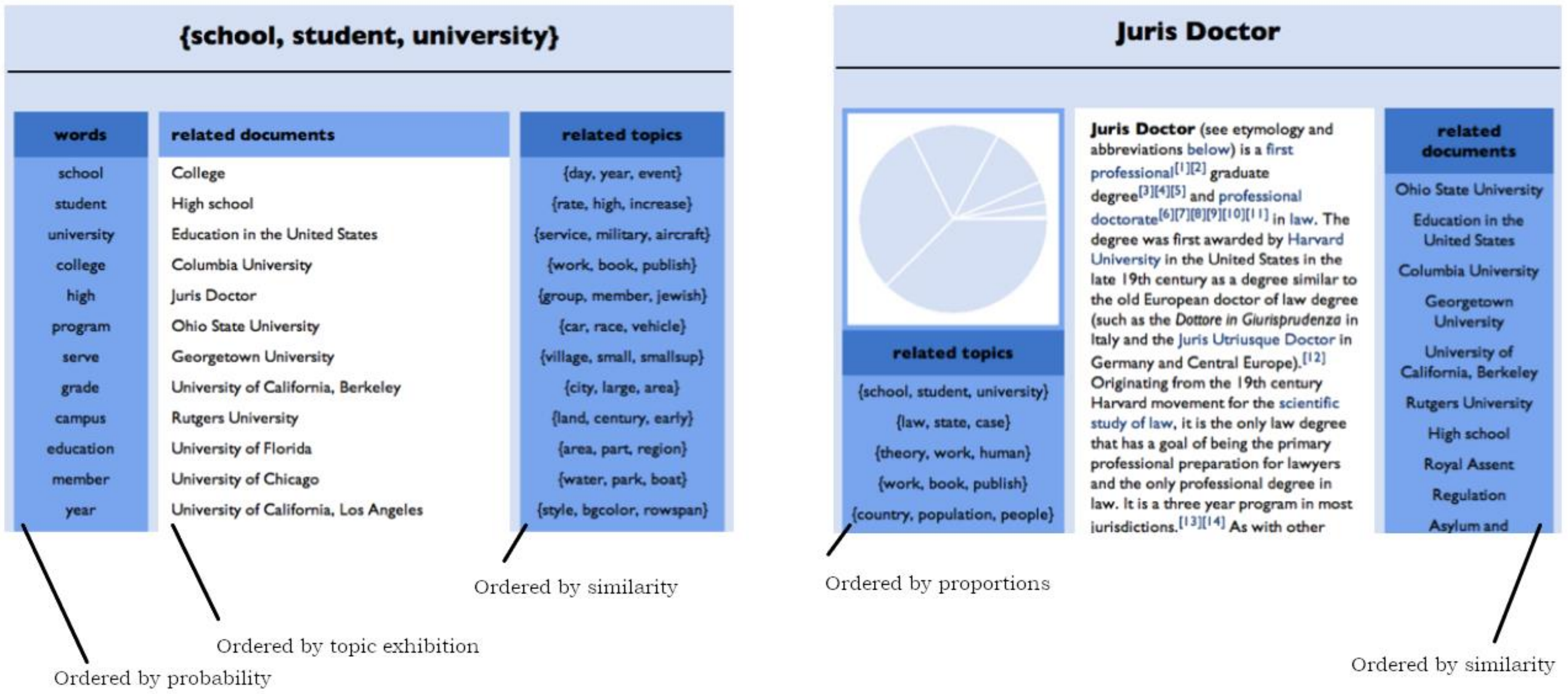
Appendix A: Topic Modelling Visualisation Figures

Figure A: Word Clouds



APPENDIX A - FIGURE A: CENTRE-ORIENTED WORD CLOUDS OF PRE-PROCESSED DATA (KUMARI, ET AL., 2021)

Figure B: List-based Navigational Tool



{school, student, university}

words	related documents	related topics
school	College	{day, year, event}
student	High school	{rate, high, increase}
university	Education in the United States	{service, military, aircraft}
college	Columbia University	{work, book, publish}
high	Juris Doctor	{group, member, jewish}
program	Ohio State University	{car, race, vehicle}
serve	Georgetown University	{village, small, smallsup}
grade	University of California, Berkeley	{city, large, area}
campus	Rutgers University	{land, century, early}
education	University of Florida	{area, part, region}
member	University of Chicago	{water, park, boat}
year	University of California, Los Angeles	{style, bgcolor, rowspan}

Juris Doctor

related topics

{school, student, university}

{law, state, case}

{theory, work, human}

{work, book, publish}

{country, population, people}

Juris Doctor (see etymology and abbreviations below) is a first professional[1][2] graduate degree[3][4][5] and professional doctorate[6][7][8][9][10][11] in law. The degree was first awarded by Harvard University in the United States in the late 19th century as a degree similar to the old European doctor of law degree (such as the *Dottore in Giurisprudenza* in Italy and the Juris Utriusque Doctor in Germany and Central Europe).[12] Originating from the 19th century Harvard movement for the scientific study of law, it is the only law degree that has a goal of being the primary professional preparation for lawyers and the only professional degree in law. It is a three year program in most jurisdictions.[13][14] As with other

related documents

Ohio State University

Education in the United States

Columbia University

Georgetown University

University of California, Berkeley

Rutgers University

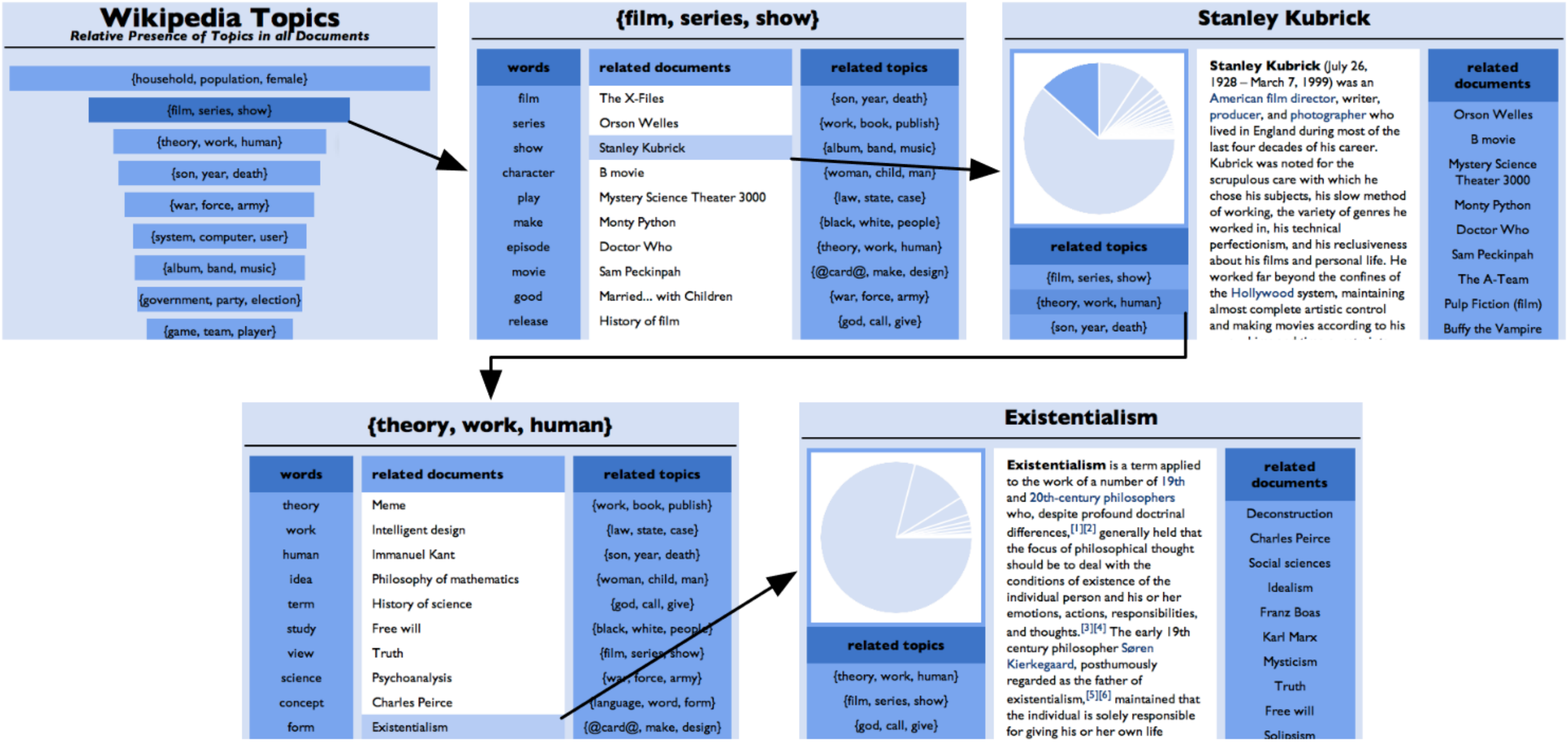
High school

Royal Assent

Regulation

Asylum and

I) TOPIC AND DOCUMENT VIEW



Wikipedia Topics

Relative Presence of Topics in all Documents

{household, population, female}
{film, series, show}
{theory, work, human}
{son, year, death}
{war, force, army}
{system, computer, user}
{album, band, music}
{government, party, election}
{game, team, player}

{film, series, show}

words	related documents	related topics
film	The X-Files	{son, year, death}
series	Orson Welles	{work, book, publish}
show	Stanley Kubrick	{album, band, music}
character	B movie	{woman, child, man}
play	Mystery Science Theater 3000	{law, state, case}
make	Monty Python	{black, white, people}
episode	Doctor Who	{theory, work, human}
movie	Sam Peckinpah	{@card@, make, design}
good	Married... with Children	{war, force, army}
release	History of film	{god, call, give}

Stanley Kubrick

Stanley Kubrick (July 26, 1928 – March 7, 1999) was an American film director, writer, producer, and photographer who lived in England during most of the last four decades of his career. Kubrick was noted for the scrupulous care with which he chose his subjects, his slow method of working, the variety of genres he worked in, his technical perfectionism, and his reclusiveness about his films and personal life. He worked far beyond the confines of the Hollywood system, maintaining almost complete artistic control and making movies according to his

related topics

{film, series, show}
{theory, work, human}
{son, year, death}

related documents

Orson Welles
B movie
Mystery Science Theater 3000
Monty Python
Doctor Who
Sam Peckinpah
The A-Team
Pulp Fiction (film)
Buffy the Vampire

{theory, work, human}

words	related documents	related topics
theory	Meme	{work, book, publish}
work	Intelligent design	{law, state, case}
human	Immanuel Kant	{son, year, death}
idea	Philosophy of mathematics	{woman, child, man}
term	History of science	{god, call, give}
study	Free will	{black, white, people}
view	Truth	{film, series, show}
science	Psychoanalysis	{war, force, army}
concept	Charles Peirce	{language, word, form}
form	Existentialism	{@card@, make, design}

Existentialism

Existentialism is a term applied to the work of a number of 19th and 20th-century philosophers who, despite profound doctrinal differences,[1][2] generally held that the focus of philosophical thought should be to deal with the conditions of existence of the individual person and his or her emotions, actions, responsibilities, and thoughts.[3][4] The early 19th century philosopher Søren Kierkegaard, posthumously regarded as the father of existentialism,[5][6] maintained that the individual is solely responsible for giving his or her own life

related topics

{theory, work, human}
{film, series, show}
{god, call, give}

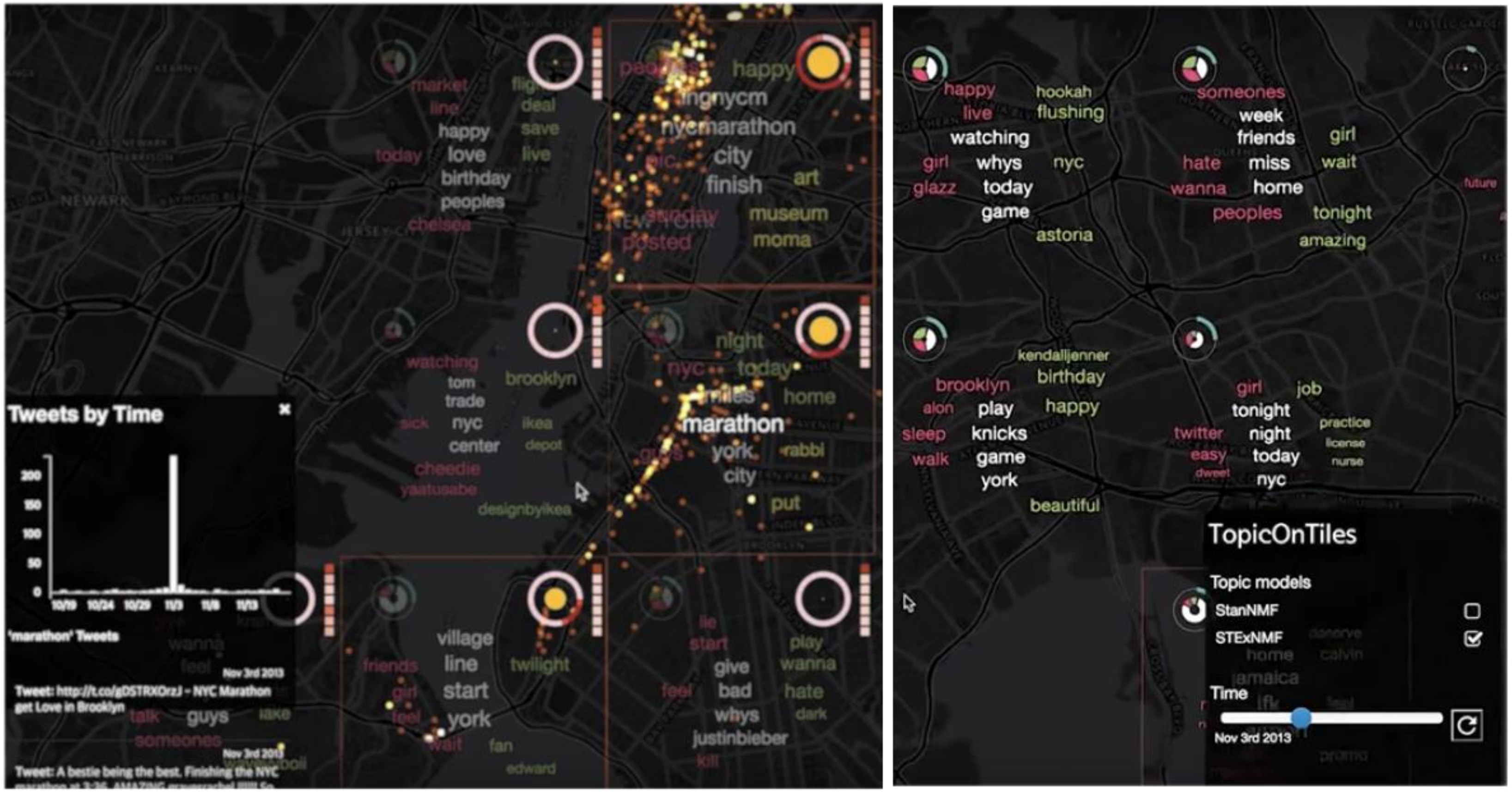
related documents

Deconstruction
Charles Peirce
Social sciences
Idealism
Franz Boas
Karl Marx
Mysticism
Truth
Free will
Solipsism

II) NAVIGATION

APPENDIX A - FIGURE B: LIST-BASED NAVIGATIONAL TOOL (CHANEY & BLEI, 2012)

Figure C: Geo-based Glyph Map



Nov 3rd 2013

Tweet: http://t.co/gDSTRXOrzJ - NYC Marathon get Love in Brooklyn

Nov 3rd 2013

Tweet: A bestie being the best. Finishing the NYC marathon at 3:36. AMAZING

StanNMF □

STExNMF ☑

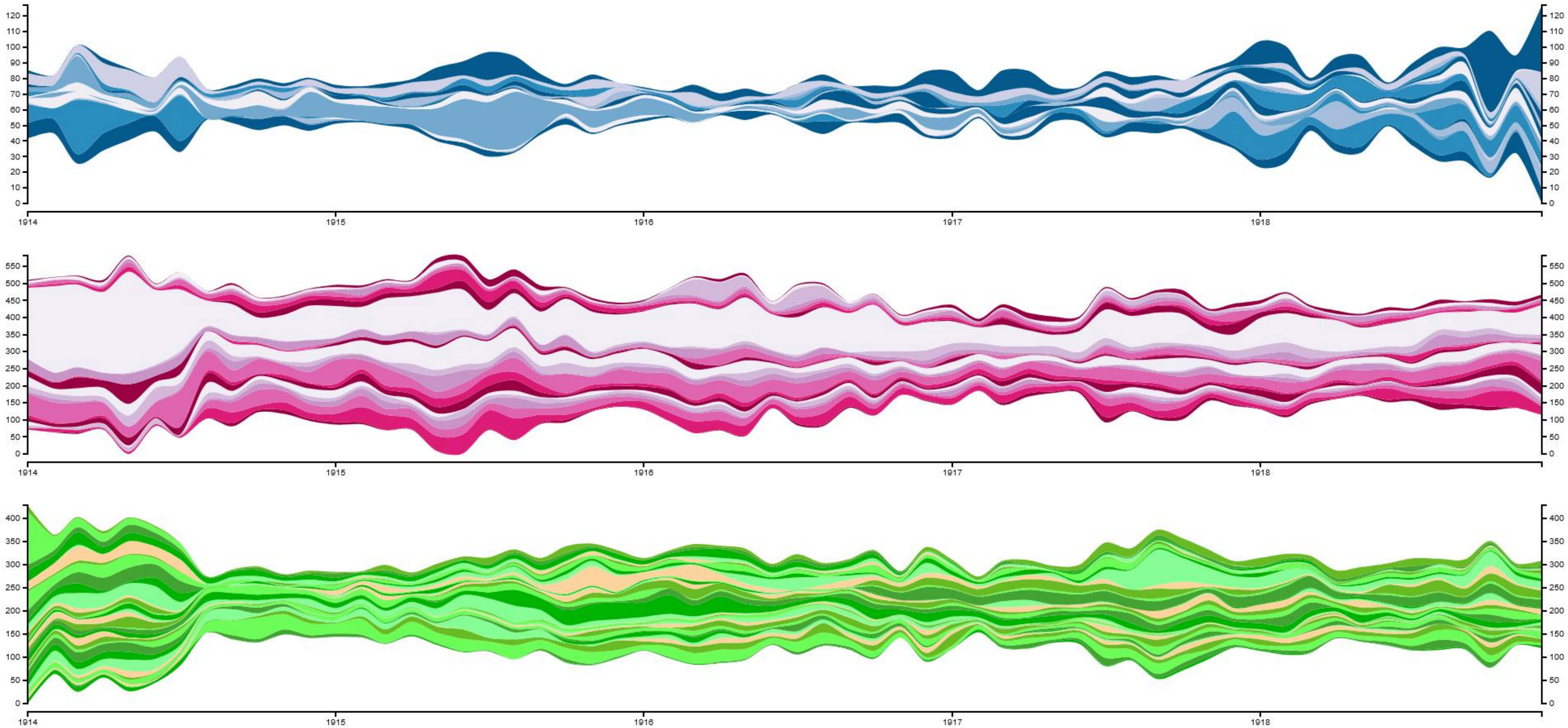
i)

ii)

Appendix A - Figure C: I) TERM-SPECIFIC GEOGRAPHICAL SPREAD;

II) MULTIPLE TOPICS IN A GRID LAYOUT AND THEIR PROBABILITIES. (CHOI, ET AL., 2018)

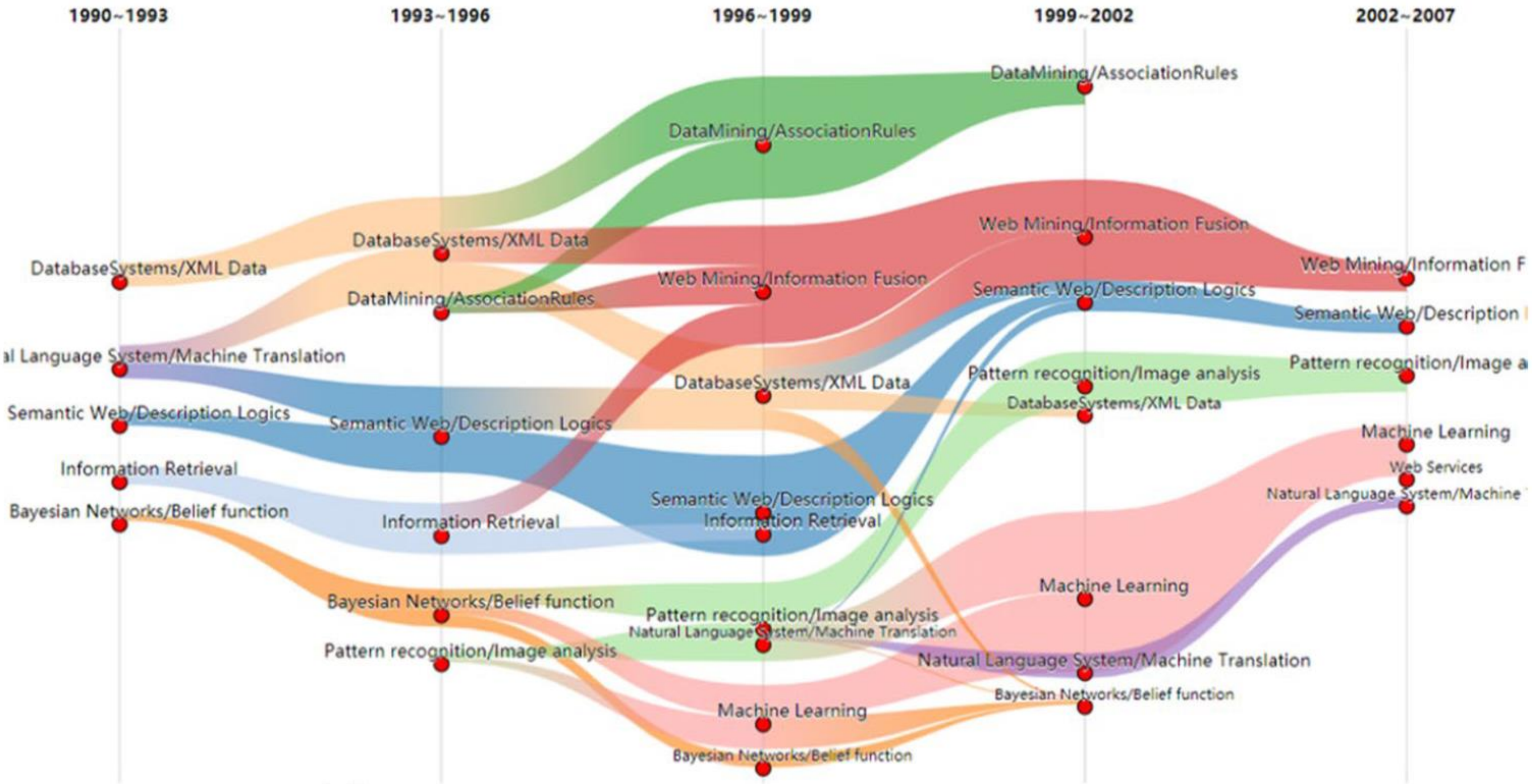
Figure D: Temporal Streamgraphs



APPENDIX A - FIGURE D: WORD WAR 1 TERMS STREAMGRAPHS (HAIDAR, ET AL., 2016)

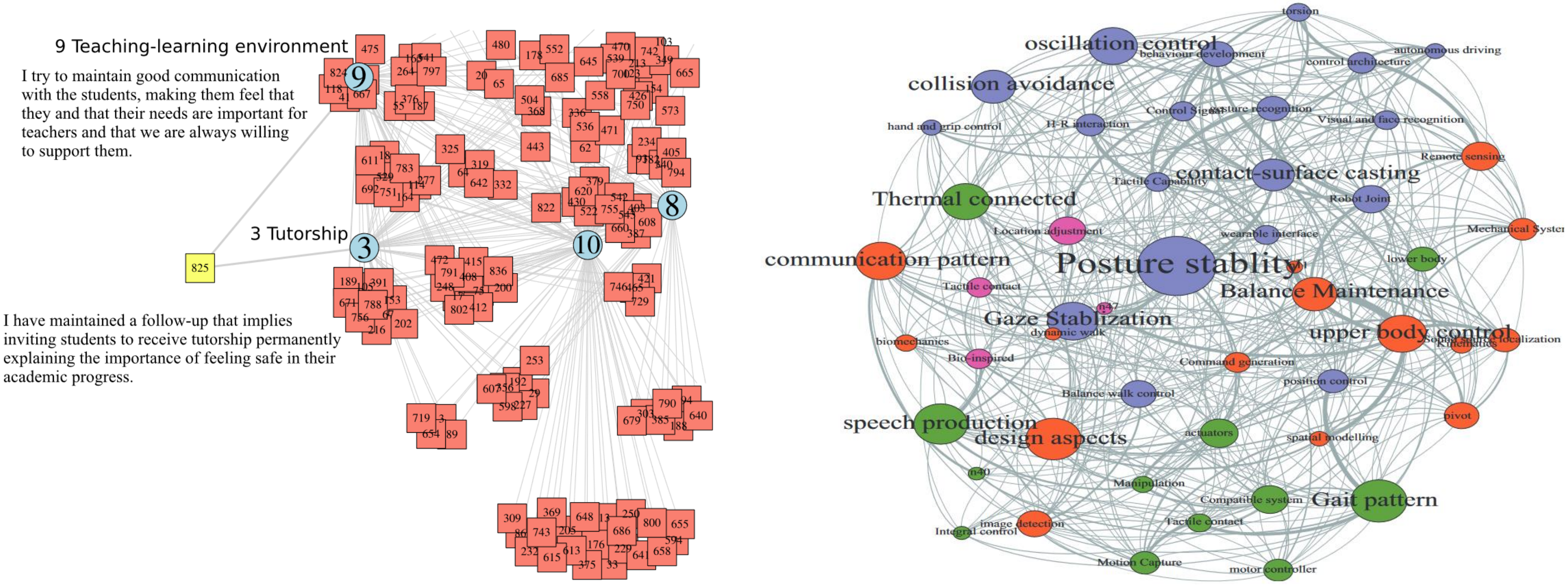
Every term has a corresponding stream where the thickness shows the term's strength at this moment in time.

Figure E: Evolution Path Map


1990~1993
1993~1996
1996~1999
1999~2002
2002~2007
DatabaseSystems/XML Data
al Language System/Machine Translation
Semantic Web/Description Logics
Information Retrieval
Bayesian Networks/Belief function
DatabaseSystems/XML Data
DataMining/AssociationRules
Semantic Web/Description Logics
Information Retrieval
Bayesian Networks/Belief function
Pattern recognition/Image analysis
DataMining/AssociationRules
Web Mining/Information Fusion
DatabaseSystems/XML Data
Semantic Web/Description Logics
Information Retrieval
Pattern recognition/Image analysis
Natural Language System/Machine Translation
Machine Learning
Bayesian Networks/Belief function
DataMining/AssociationRules
Web Mining/Information Fusion
Semantic Web/Description Logics
Pattern recognition/Image analysis
DatabaseSystems/XML Data
Machine Learning
Natural Language System/Machine Translation
Bayesian Networks/Belief function
Web Mining/Information F
Semantic Web/Description
Pattern recognition/Image a
Machine Learning
Web Services
Natural Language System/Machine

Appendix A - Figure E: Evolution Path Map of the AI and Machine Learning Research Area (Liu, et al., 2020)

Figure F: Network Diagrams



I try to maintain good communication with the students, making them feel that they and that their needs are important for teachers and that we are always willing to support them.

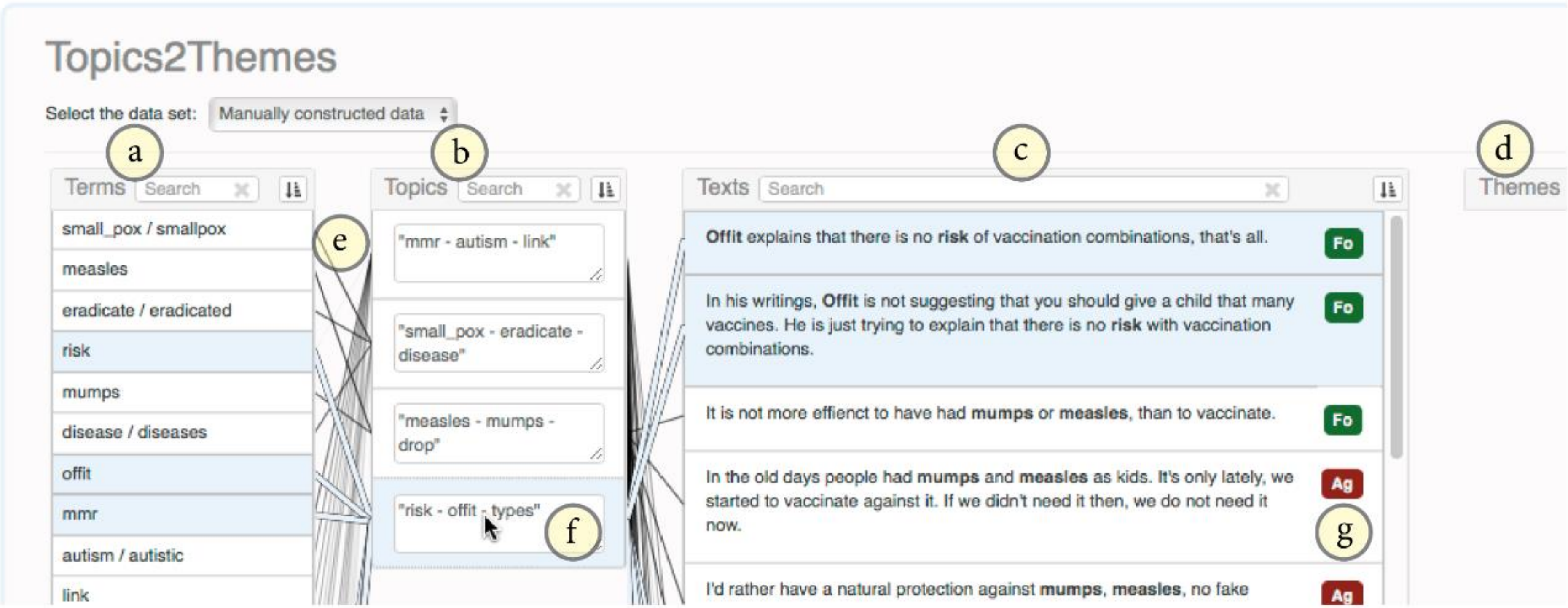
I have maintained a follow-up that implies inviting students to receive tutorship permanently explaining the importance of feeling safe in their academic progress.

i)

ii)

APPENDIX A - FIGURE F: I) PART OF DOCUMENT-TOPIC NETWORK DIAGRAM (BUENAÑO-FERNANDEZ, ET AL., 2020); II) TOPIC-CONNECTEDNESS NETWORK DIAGRAM (KUMARI, ET AL., 2021)

Figure G: Topics2Themes Dashboard



Topics2Themes

Select the data set: Manually constructed data

Terms

- small_pox / smallpox
- measles
- eradicate / eradicated
- risk
- mumps
- disease / diseases
- offit
- mmr
- autism / autistic
- link

Topics

- "mmr - autism - link"
- "small_pox - eradicate - disease"
- "measles - mumps - drop"
- "risk - offit - types"

Texts

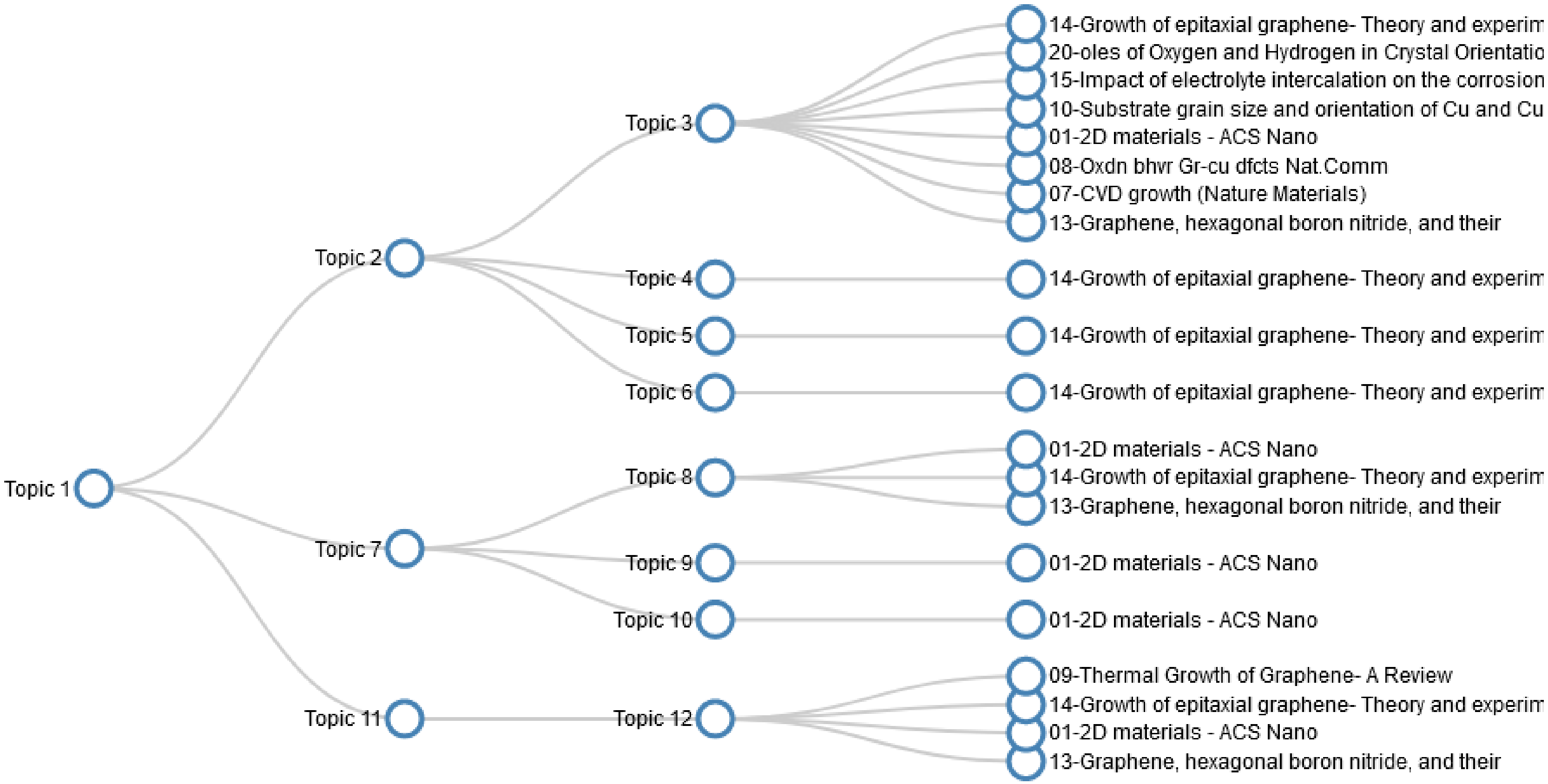
- **Offit** explains that there is no **risk** of vaccination combinations, that's all. Fo
- In his writings, **Offit** is not suggesting that you should give a child that many vaccines. He is just trying to explain that there is no **risk** with vaccination combinations. Fo
- It is not more effienct to have had **mumps** or **measles**, than to vaccinate. Fo
- In the old days people had **mumps** and **measles** as kids. It's only lately, we started to vaccinate against it. If we didn't need it then, we do not need it now. Ag
- I'd rather have a natural protection against **mumps**, **measles**, no fake Ag

Themes

APPENDIX A - FIGURE G: TOPICS2THEMES DASHBOARD (SKEPPSTEDT, ET AL., 2018)

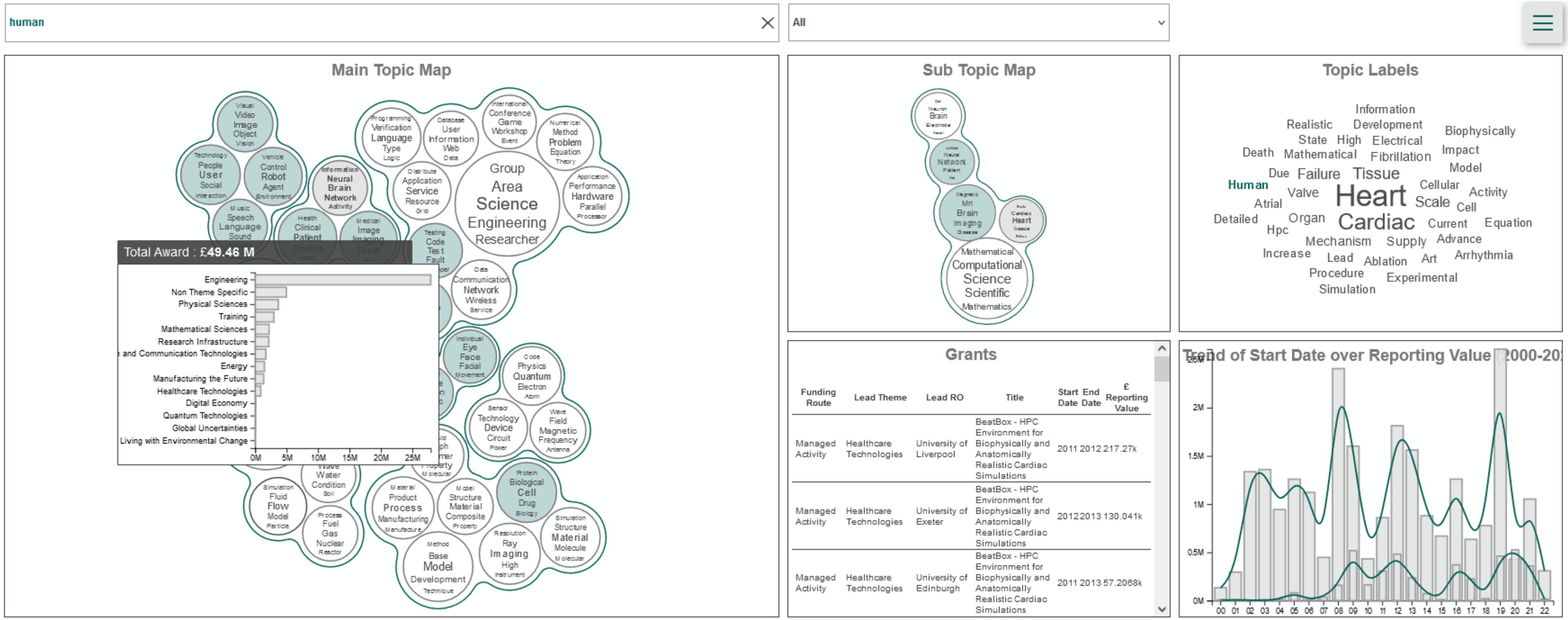
The interface includes the following components: (a–d)the panels containing lists of terms, topics, document texts, and user-created themes, respectively; (e) links between the related elements of the respective lists (e.g., terms belonging to a topic); (f) a topic highlighted by the user by hovering; and (g) a stance symbol assigned to the corresponding text.

Figure H: TopExplorer Tree Diagram


Topic 1
Topic 2
Topic 3
14-Growth of epitaxial graphene- Theory and experim
20-oles of Oxygen and Hydrogen in Crystal Orientatio
15-Impact of electrolyte intercalation on the corrosion
10-Substrate grain size and orientation of Cu and Cu
01-2D materials - ACS Nano
08-Oxdn bhvr Gr-cu dfcts Nat.Comm
07-CVD growth (Nature Materials)
13-Graphene, hexagonal boron nitride, and their
Topic 4
14-Growth of epitaxial graphene- Theory and experim
Topic 5
14-Growth of epitaxial graphene- Theory and experim
Topic 6
14-Growth of epitaxial graphene- Theory and experim
Topic 7
Topic 8
01-2D materials - ACS Nano
14-Growth of epitaxial graphene- Theory and experim
13-Graphene, hexagonal boron nitride, and their
Topic 9
01-2D materials - ACS Nano
Topic 10
01-2D materials - ACS Nano
Topic 11
Topic 12
09-Thermal Growth of Graphene- A Review
14-Growth of epitaxial graphene- Theory and experim
01-2D materials - ACS Nano
13-Graphene, hexagonal boron nitride, and their

APPENDIX A - FIGURE H: TOPEXPLORER - TOPIC-HIERARCHICAL TREE DIAGRAM (CHENG, ET AL., 2020)

Figure I: Cluster-based BubbleMap of UKRI Software Research Grants

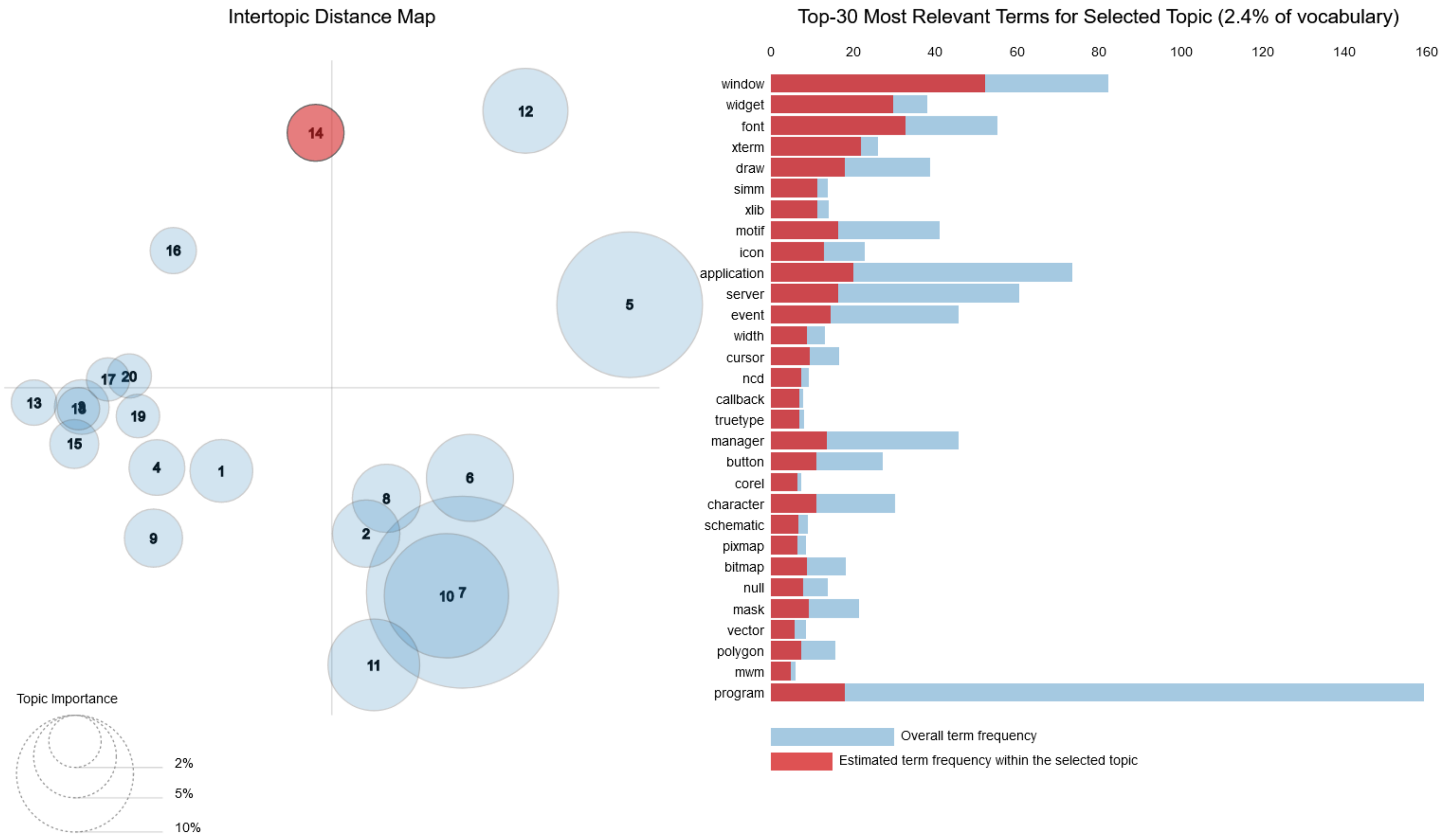

human
All
Main Topic Map
Visual Video Image Object Vision
Technology People User Social Interaction
Vehicle Control Robot Agent Environment
Information Neural Brain Network Activity
Music Speech Language Sound
Health Clinical Patient
Medical Image Imaging
Programming Verification Language Type Logic
Database User Information Web Data
International Conference Game Workshop Event
Numerical Method Problem Equation Theory
Distribute Application Service Resource Grid
Group Area Science Engineering Researcher
Application Performance Hardware Parallel Processor
Testing Code Test Fault
Data Communication Network Wireless Service
Individual Eye Face Facial Movement
Code Physics Quantum Electron Atom
Sensor Technology Device Circuit Power
Wave Field Magnetic Frequency Antenna
Water Condition Soil
Simulation Fluid Flow Model Particle
Process Fuel Gas Nuclear Reactor
Material Product Process Manufacturing Manufacture
Model Structure Material Composite Property
Protein Biological Cell Drug Biology
Simulation Structure Material Molecule Molecular
Method Base Model Development Technique
Resolution Ray Imaging High Instrument
Total Award : £49.46 M
Engineering
Non Theme Specific
Physical Sciences
Training
Mathematical Sciences
Research Infrastructure
and Communication Technologies
Energy
Manufacturing the Future
Healthcare Technologies
Digital Economy
Quantum Technologies
Global Uncertainties
Living with Environmental Change
0M
5M
10M
15M
20M
25M
Sub Topic Map
Cell Neuron Brain Electrode Neural
Artificial Neural Network Pattern
Magnetic Mri Brain Imaging Disease
Cardiac Heart Tissue Failure
Mathematical Computational Science Scientific Mathematics
Topic Labels
Information
Realistic
Development
Biophysically
State
High
Electrical
Death
Mathematical
Fibrillation
Impact
Model
Due
Failure
Tissue
Human
Valve
Heart
Cellular
Activity
Atrial
Scale
Cell
Detailed
Organ
Cardiac
Current
Equation
Hpc
Mechanism
Supply
Advance
Increase
Lead
Ablation
Art
Arrhythmia
Procedure
Experimental
Simulation
Grants
Trend of Start Date over Reporting Value (2000-20
2.5M
2M
1.5M
1M
0.5M
0M
00
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
20
21
22

Funding Route	Lead Theme	Lead RO	Title	Start Date	End Date	£ Reporting Value
Managed Activity	Healthcare Technologies	University of Liverpool	BeatBox - HPC Environment for Biophysically and Anatomically Realistic Cardiac Simulations	2011	2012	217.27k
Managed Activity	Healthcare Technologies	University of Exeter	BeatBox - HPC Environment for Biophysically and Anatomically Realistic Cardiac Simulations	2012	2013	130.041k
Managed Activity	Healthcare Technologies	University of Edinburgh	BeatBox - HPC Environment for Biophysically and Anatomically Realistic Cardiac Simulations	2011	2013	57.2068k

Appendix A - Figure I: UKRI Research Grants BubbleMap Dashboard (Gharavi, et al., 2022)

Viewports from left to right: 1. *Main Topic Map:* super-topic corpus structure; 2. *Sub-Topic Map:* sub-topics related to the user-selected super-topic; 3. *Grants:* linked documents ranked by similarity; 4. *Topic Labels:* selected topic's word cloud representation; 5. *Trend of Start Date over Reporting Value (2000-2022):* topic-related temporal data.

Figure J: LDAvis-based Testing Interface

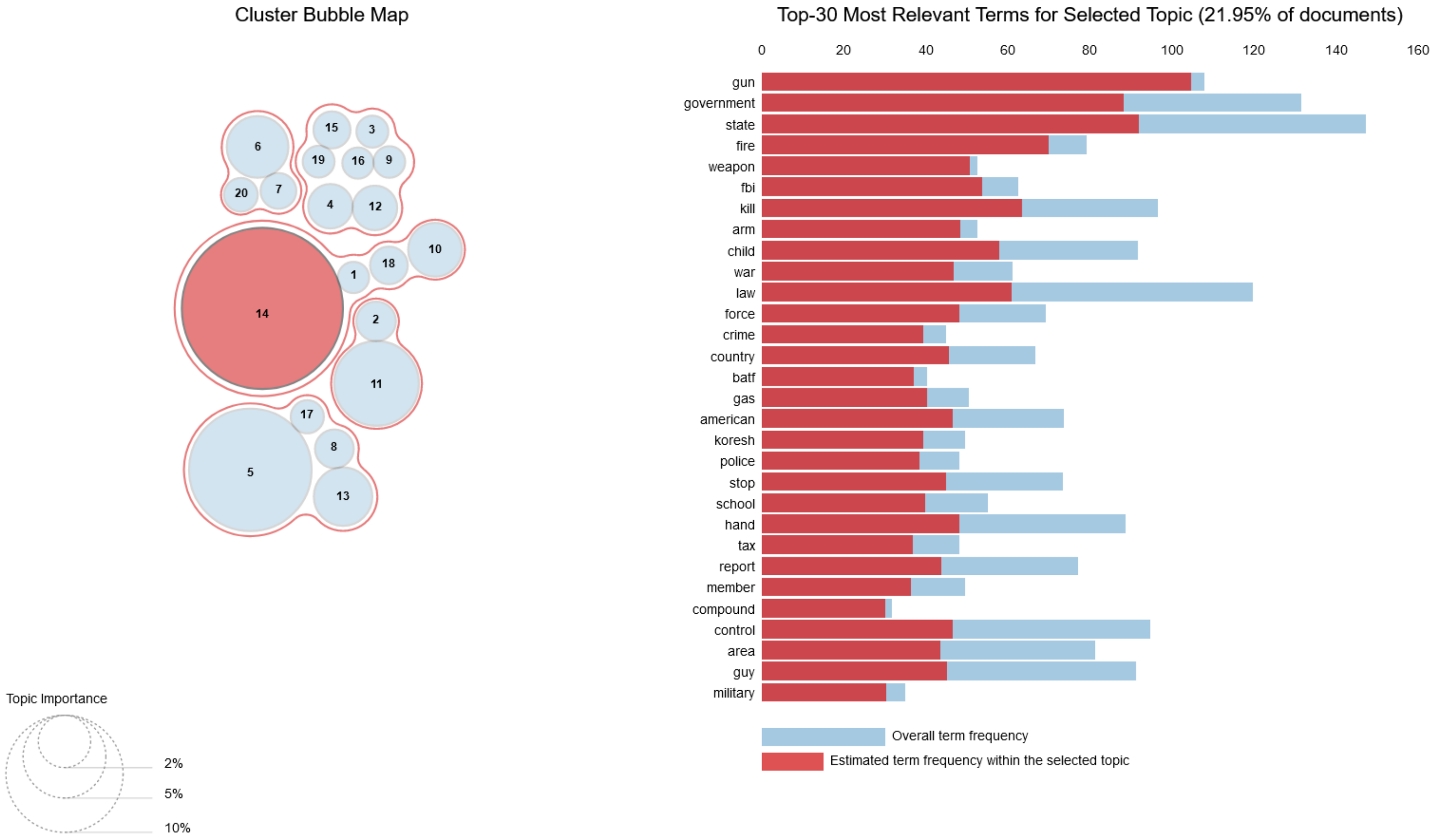


Alterations:

1. The top two panels are removed, and **λ** is internally set to 0.6;
2. *Labels are simplified.*
3. Topic numbers are randomised.
4. Term distribution functionality is removed.

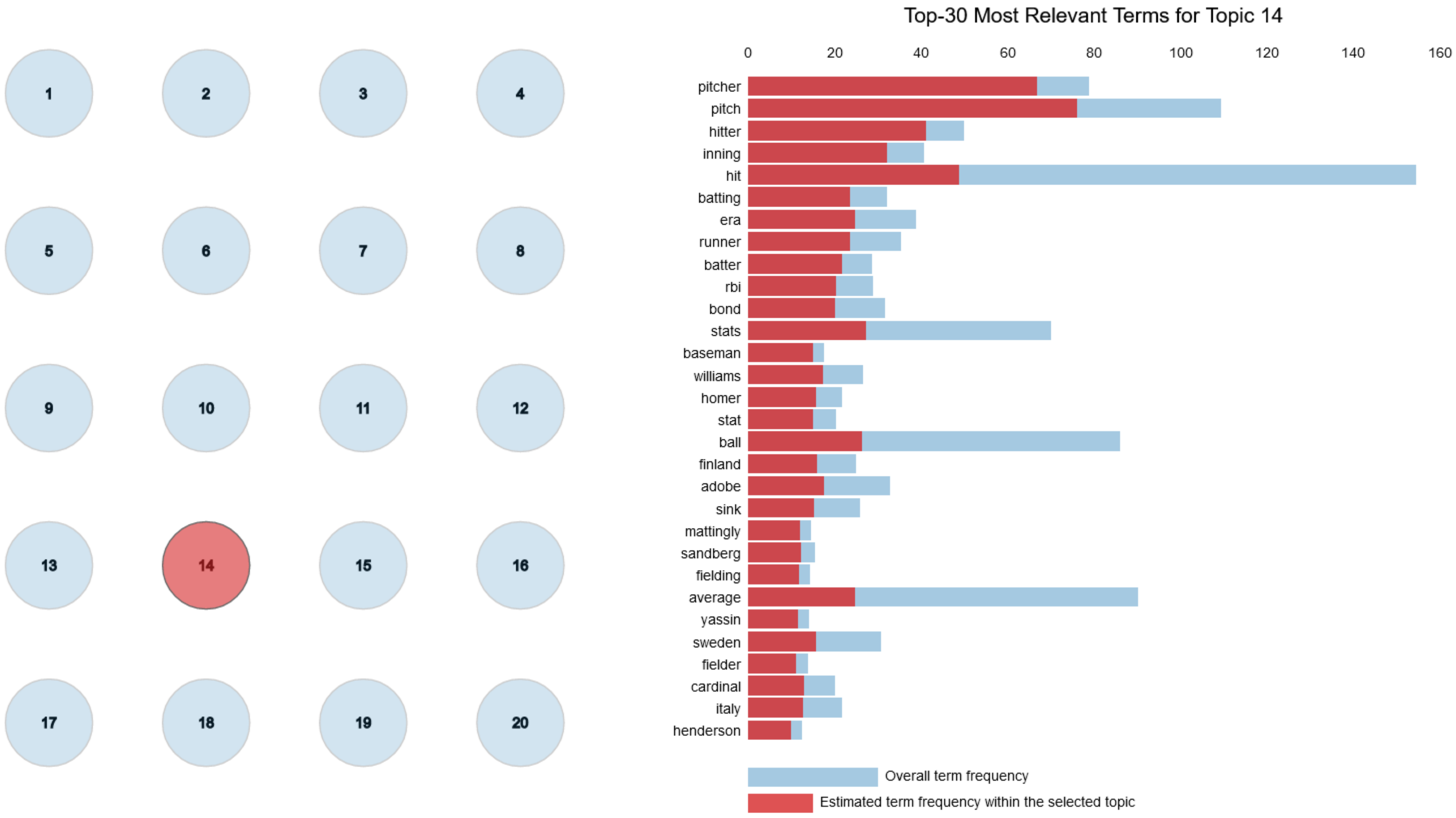
APPENDIX A - FIGURE J: FINAL LDAVIS-BASED INTERFACE

Figure K: BubbleMap Testing Interface



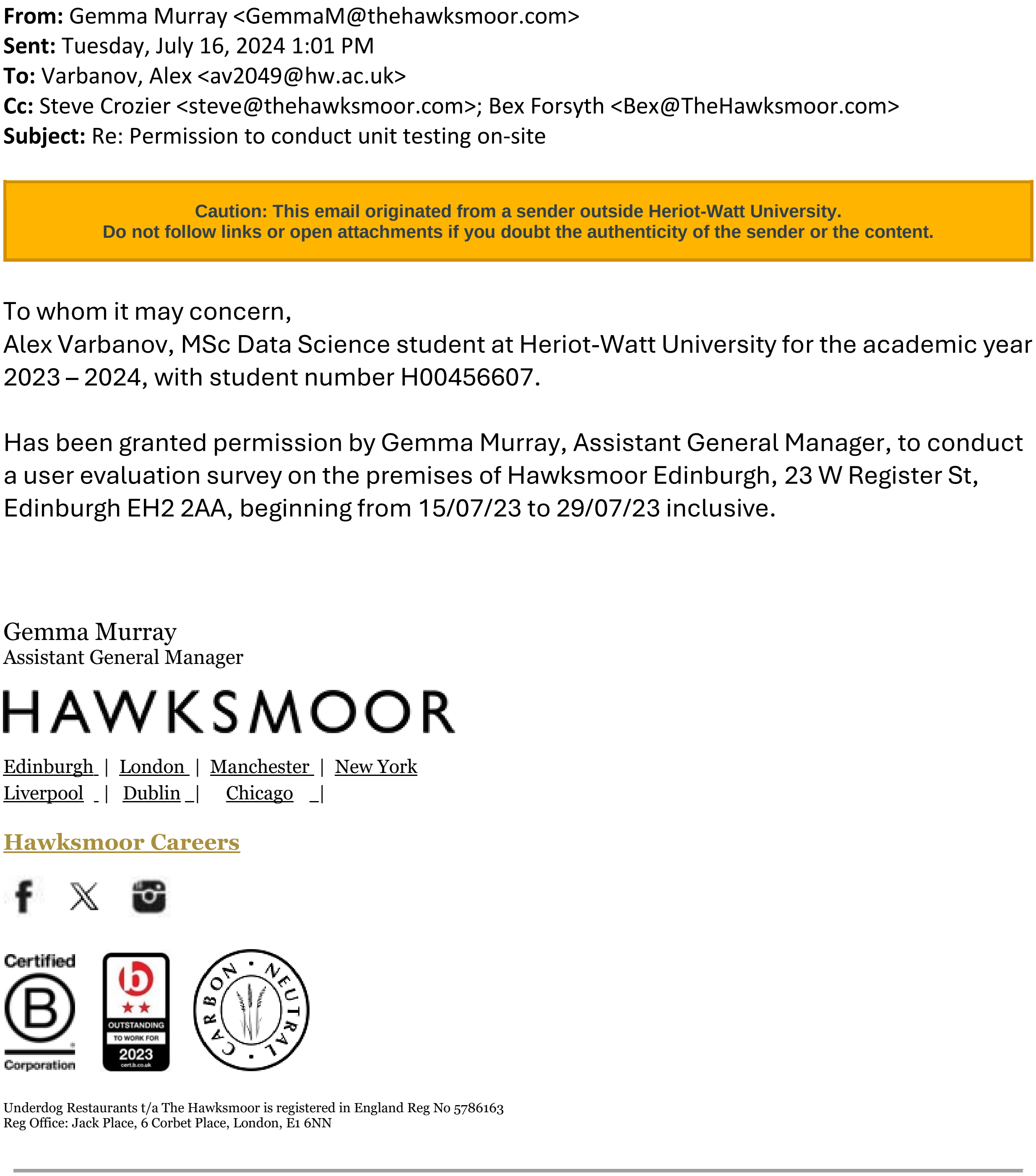
APPENDIX A - FIGURE K: FINAL BUBBLEMAP INTERFACE

Figure L: Control Testing Interface - Training


Top-30 Most Relevant Terms for Topic 14

APPENDIX A - FIGURE L: FINAL TRAINING INTERFACE

Appendix B: Email Correspondence



From: Gemma Murray <GemmaM@thehawksmoor.com>
Sent: Tuesday, July 16, 2024 1:01 PM
To: Varbanov, Alex <av2049@hw.ac.uk>
Cc: Steve Crozier <steve@thehawksmoor.com>; Bex Forsyth <Bex@TheHawksmoor.com>
Subject: Re: Permission to conduct unit testing on-site

Caution: This email originated from a sender outside Heriot-Watt University.
Do not follow links or open attachments if you doubt the authenticity of the sender or the content.

To whom it may concern,
Alex Varbanov, MSc Data Science student at Heriot-Watt University for the academic year 2023 – 2024, with student number H00456607.

Has been granted permission by Gemma Murray, Assistant General Manager, to conduct a user evaluation survey on the premises of Hawksmoor Edinburgh, 23 W Register St, Edinburgh EH2 2AA, beginning from 15/07/23 to 29/07/23 inclusive.

Gemma Murray
Assistant General Manager

HAWKSMOOR

Edinburgh | London | Manchester | New York
Liverpool | Dublin | Chicago |

Hawksmoor Careers

Underdog Restaurants t/a The Hawksmoor is registered in England Reg No 5786163
Reg Office: Jack Place, 6 Corbet Place, London, E1 6NN

From: Varbanov, Alex <av2049@hw.ac.uk>
Sent: 12 July 2024 12:01
To: Gemma Murray <GemmaM@thehawksmoor.com>
Cc: Steve Crozier <steve@thehawksmoor.com>; Bex Forsyth <Bex@TheHawksmoor.com>
Subject: Permission to conduct unit testing on site

Hi Gemma/ Steve/ Bex,

This is Alex, writing from my university email. I've spoken to Bex last month about conducting an experiment on-site for my dissertation. I'm required to attach a letter of permission to my report, stating that I have been allowed to conduct my unit testing on the premises. It would be me sitting at a corner table with my laptop + participant in the quiet lunch times during weekdays. If you can just send me back confirmation stating that I:

Alex Varbanov,
MSc Data Science student at Heriot-Watt University for the academic year 2023 – 2024,
with student number H00456607

have been granted permission by

[your name + job title]

To conduct a user evaluation survey on the premises of Hawksmoor Edinburgh, 23 W Register St, Edinburgh EH2 2AA, beginning from 15/07/23 to 29/07/23 inclusive.

Thank you so much!

A

Founded in 1821, Heriot-Watt is a leader in ideas and solutions. With campuses and students across the entire globe we span the world, delivering innovation and educational excellence in business, engineering, design and the physical, social and life sciences. This email is generated from the Heriot-Watt University Group, which includes:

1. Heriot-Watt University, a Scottish charity registered under number SC000278
2. Heriot- Watt Services Limited (Oriam), Scotland's national performance centre for sport. Heriot-Watt Services Limited is a private limited company registered is Scotland with registered number SC271030 and registered office at Research & Enterprise Services Heriot-Watt University, Riccarton, Edinburgh, EH14 4AS.

The contents (including any attachments) are confidential. If you are not the intended recipient of this e-mail, any disclosure, copying, distribution or use of its contents is strictly prohibited, and you should please notify the sender immediately and then delete it (including any attachments) from your system.

Appendix C: Information Sheet

HERIOT WATT UNIVERSITY

INFORMATION SHEET

Project title: TOPIC MODEL VISUALISATIONS: A QUALITATIVE USER EVALUATION STUDY

Objectives

For my master's thesis, I am investigating the usability of different layouts used for visualising text data. The aim of this study is to rate the effectiveness of each method. To achieve this, I am conducting this user study where participants can provide their evaluation of such interfaces.

This information sheet will provide details about the modalities of this study.

Data collection

Data collection will be in three parts. First, I will ask you to complete a consent form, with your name and signature to express your willingness to participate and understanding of the study.
Then, I will collect anonymised answers about your demographics (gender, age, education) and prior experience – **these questions are optional**.
Finally, I will collect your answers to tasks relating to the visualisation interfaces. These will be noted on paper, and you will **not** be digitally recorded.

The demographics, prior experience and answers will be anonymised and identified by a unique random ID. This ID will not be linked to your name.

Experiment Protocol

The session is timed to take about 30 minutes and will take place in typical office conditions.

You will be shown 3 visualisation interfaces in sequence. For each interface, there will be between 2 and 4 tasks to complete – tasks will target the thematic analysis of a text corpus. You will be asked to speak your answers to these tasks out loud, and I will write them down.

After this, you will be asked to answer 8 questions about your interface preference. Again, you should speak your answer out loud, and I will take note of these.

Compensation

There is no compensation granted for taking part in this study.

Withdrawal

You may withdraw from this study at any time during the session with no consequences. Any answer you would have given by then will be destroyed.

If you wish to withdraw your participation after completing the study, you must provide the participant ID given to you, so that the data you provided can be excluded from the study. Note that a request to withdraw your participation after the 10th of August 2024 will not be possible, as results would have been submitted by then as part of my master's thesis report.

By ticking the relevant box in the consent form the participant confirms that they have been
informed of the withdrawal policy described above and agree to proceed with undertaking the
experiment. They can state their wish to withdraw both verbally and in writing.

For additional details or to request the deletion of your data, you can reach out to av2049@hw.ac.uk or you can contact the project supervisor via pierre.le_bras@hw.ac.uk.

NOTE: Please find below further information on Heriot-Watt Data Protection

1) Data Protection Officer contact details: dataprotection@hw.ac.uk

2) Heriot-Watt University Privacy Notice for Research Participants: https://www.hw.ac.uk/uk/services/docs/information-governance/PrivacyNoticeResearch-V4Finalversion.pdf

3) Data controller is Heriot-Watt University

Appendix D: Consent Form

HERIOT WATT UNIVERSITY

CONSENT FORM

Project title: TOPIC MODEL VISUALISATIONS: A QUALITATIVE USER EVALUATION STUDY

Heriot-Watt University attaches high priority to the ethical conduct of research. We therefore ask you to consider the following points before signing this form. Your signature confirms that you are willing to participate in this study, however, signing this form does not commit you to anything you do not wish to do, and you are free to withdraw your participation at any time.

Please tick the initial boxes

○ **I confirm** that the purpose of this study was explained to me in sufficient detail and I have had the opportunity to consider the information, to ask questions and have these answered satisfactorily. □

○ **I confirm** that I am over 18 years old. □

○ **I understand** that my participation is voluntary and that I am free to withdraw at any time, without giving any reason, and without consequences. □

○ **I understand** that any data I give will be used in the dissemination of findings and will remain anonymous. □

○ **I feel** I am well enough, physically and mentally, to complete the tasks as described in the Information Sheet. □

○ **I agree** to take part in the above study. □

Project Student:	Project Supervisor:
Alex Varbanov	Pierre Le Bras
E-mail: av2049@hw.ac.uk	Email: pierre.le_bras@hw.ac.uk

Name: ………………………………………………………………………………………………

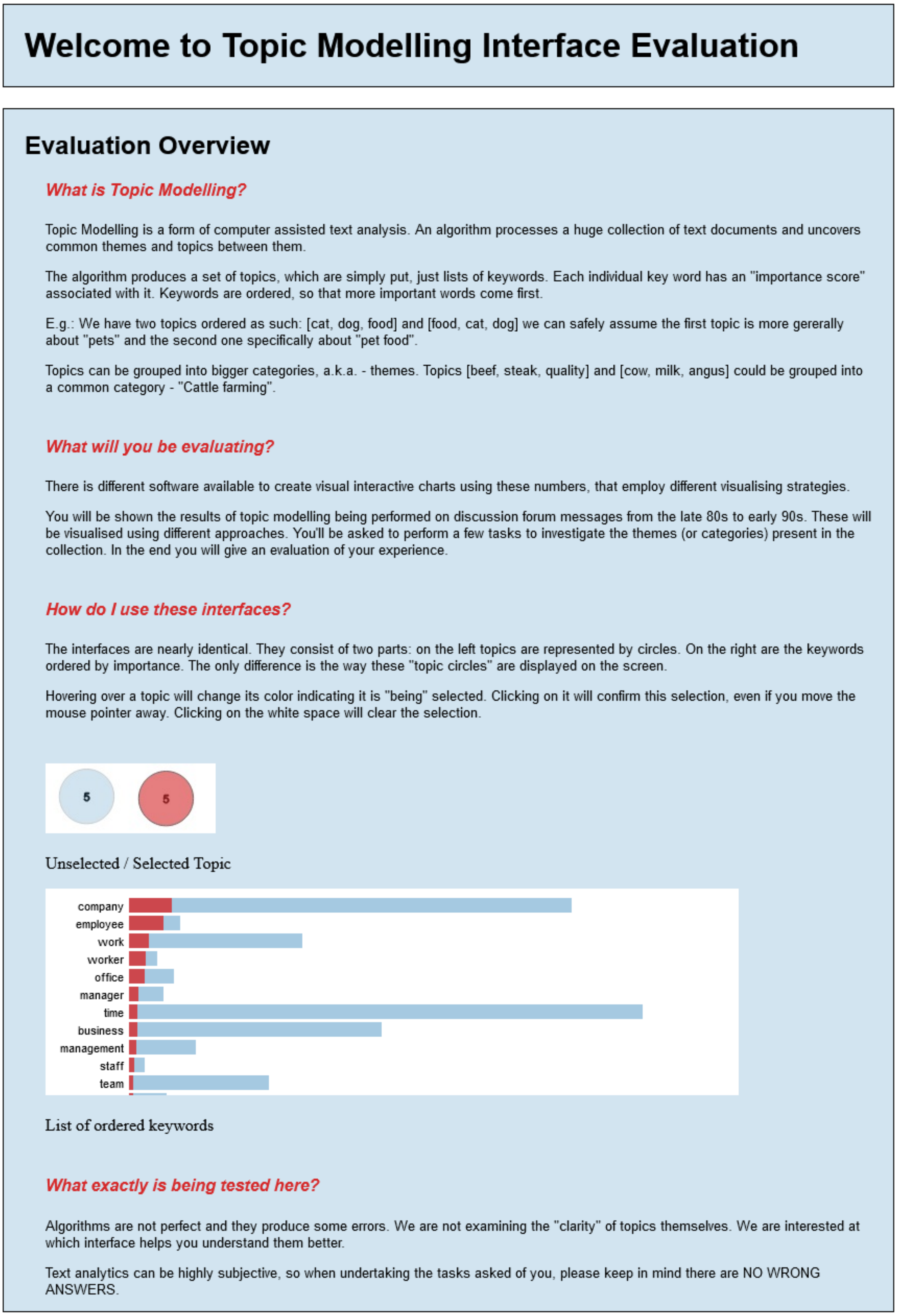
Signature: ……………………………………………………………………………………………

Date: …………………………………………………………………………………………………..

Email Address: ……………………………………………………………………………………...

I would like to receive information on the results of this study: Yes □ No □

Appendix E: Session Set-up – Introduction Script



Welcome to Topic Modelling Interface Evaluation

Evaluation Overview

What is Topic Modelling?

Topic Modelling is a form of computer assisted text analysis. An algorithm processes a huge collection of text documents and uncovers common themes and topics between them.

The algorithm produces a set of topics, which are simply put, just lists of keywords. Each individual key word has an "importance score" associated with it. Keywords are ordered, so that more important words come first.

E.g.: We have two topics ordered as such: [cat, dog, food] and [food, cat, dog] we can safely assume the first topic is more gererally about "pets" and the second one specifically about "pet food".

Topics can be grouped into bigger categories, a.k.a. - themes. Topics [beef, steak, quality] and [cow, milk, angus] could be grouped into a common category - "Cattle farming".

What will you be evaluating?

There is different software available to create visual interactive charts using these numbers, that employ different visualising strategies.

You will be shown the results of topic modelling being performed on discussion forum messages from the late 80s to early 90s. These will be visualised using different approaches. You'll be asked to perform a few tasks to investigate the themes (or categories) present in the collection. In the end you will give an evaluation of your experience.

How do I use these interfaces?

The interfaces are nearly identical. They consist of two parts: on the left topics are represented by circles. On the right are the keywords ordered by importance. The only difference is the way these "topic circles" are displayed on the screen.

Hovering over a topic will change its color indicating it is "being" selected. Clicking on it will confirm this selection, even if you move the mouse pointer away. Clicking on the white space will clear the selection.

Unselected / Selected Topic

List of ordered keywords

What exactly is being tested here?

Algorithms are not perfect and they produce some errors. We are not examining the "clarity" of topics themselves. We are interested at which interface helps you understand them better.

Text analytics can be highly subjective, so when undertaking the tasks asked of you, please keep in mind there are NO WRONG ANSWERS.