

3D gesture-based exploration and search in document collections

Angélica de Antonio, Cristian Moral, Daniel Klepel, Martín J. Abente

Facultad de Informática, Campus de Montegancedo, Univ. Politécnica de Madrid
{angelica, cmoral, dklepel}@fi.upm.es, mabente@alumnos.upm.es

Abstract. This paper describes an approach towards the interaction with 3D representations of large document collections. The goal was to provide the user with a highly dynamic environment in which even the very mapping strategy to position documents in space can be adjusted by the user depending on the specific task at hand, on his preferences, or on the context. A modification to the FDP algorithm is proposed, as well as a new gesture-based interaction paradigm in which the user can explore and search information in the collection just by simple hand movements. An experimental user evaluation was carried on to investigate the impact of the proposed approach on the precision of the mental model built by users through exploration, on the effectiveness in information search tasks, and on the general user satisfaction and perception of utility.

Keywords. 3D visualization and navigation, 3D interaction, Information Retrieval, Force-Directed Placement, Clustering, k-means

Introduction

Exploration and search of information contained in large document collections is increasingly a need, but many times also a problem. In fact, Internet offers such a huge quantity of information that it is in practice impossible to explore it exhaustively so as to fill a specific information need. In the last 30 years, many techniques have been proposed to visualize document collections. Some of them were developed for 3D virtual environments, as the third dimension increases the information that can be shown to the user. The application of one or another technique often depends on the type of documents to be visualized. For structured documents, specific visualizations highlight the structure and relations between data [1] [2], while for non-structured documents only the content can be analyzed in order to infer as much information as possible. In the last case, visualization techniques are more general and sophisticated, and then more complex [3].

This paper focuses on 3D visualization of non-structured document collections. The visual metaphor chosen, a sphere, is a simple and generic one in order to avoid overloading the visualization. Then, a collection of documents is virtually represented as a cloud of spheres. But visualization is definitely not enough to satisfy the user needs; interaction with the 3D representation is the key aspect to facilitate the user tasks.

The remainder of this document is divided into 4 sections. In Section 1 we detail the general process that must be followed to visualize a document collection in a 3D virtual environment. Section 2 deals with the interaction techniques we have proposed to

allow the user exploring and extracting information from the 3D virtual environment. To evaluate the benefits and drawbacks of our proposal, we carried out an experiment, whose details and results are presented in Section 3. Finally, Section 4 reports the conclusions we have extracted from our work and proposes future research.

1. 3D Visualization of document collections

Based on the model described in [4], the information visualization pipeline can be split into four major stages. The first one consists in gathering, cleaning and preprocessing the documents, aiming at extracting from data attributes of interest that contain relevant information. Selecting these attributes requires a deep understanding of the task that should be carried on with the data and the nature of the data itself. The second stage consists in mapping visual characteristics to each of these attributes. The most common visual characteristics are color, shape, spatial position and movement [5], and they must be selected carefully in order to exploit the natural perception abilities of the users and considering the nature of each attribute [6]. The third stage consists in the generation of a visual representation of the scene making an efficient use of the screen space while keeping the cognitive load as low as possible. The fourth step deals with the interaction between the users and the virtual representation. Interaction is a fundamental issue regarding the design of any information visualization tool. Interaction mechanisms must be designed in order to assist the user's tasks, and they can act upon any of the previous stages including: select and manipulate a dataset, navigate through the environment and perform actions affecting the data or the system itself.

1.1. Document Preprocessing

In our system, preprocess aims at determining the similarity between pairwise documents. Many complementary treatments have been proposed to undertake this task: normalization, stopwords lists, stemming [7], Part-Of-Speech Tagging [8], Named Entity Recognition [9], and many others. Which of them to use depends on the efficiency and precision required in the retrieval of information. In our case, we look for a dynamic and interactive, and hence efficient, system, even at the expense of precision.

Our preprocessing phase consists of two main parts: the lexical analysis of each document, and clustering the documents into semantic groups based on the distance measure between them. The first part allows us to translate documents into data structures treatable by computers. The process starts with the tokenization of the documents, followed by a normalization of the obtained tokens, for example by clearing accents, punctuation, email addresses, URLs and numbers, among others. All filters are applied to individual tokens, as we use a bag-of-words model, where context and order are not taken into account. Even if this causes the loss of some semantic information, the treatment is much easier and faster. Once tokens are "clean", we proceed to delete the meaningless ones (included in a stopword list), like articles, prepositions, determinants and so on. At the end, the noise, and the initial vocabulary, are reduced, fastening and the process and increasing its precision.

Before proceeding to the second step of preprocessing, we have to represent every document, and its filtered tokens, in a computational way. We have decided to use

multidimensional vectors where every dimension, also called feature, corresponds to a token, and its content is the frequency of the token in the document. This way of representing documents is known as Vector Space Model [10], allows to easily figure out the so-called cosine similarity between two documents by computing the inner product of their normalized unit vector representations. This measure of similitude, calculated for every pairwise documents, is statistical in nature as it only reflects the proportion of tokens that both documents have in common, independently of their semantic meaning. Therefore, it will not reflect faithfully if they share the same topic, but only the rate of words they share. Even if this may seem a drawback, our objective is not to obtain very precise similitude values, but to obtain good enough ones in the shortest possible time.

After this, the system assigns each document to a thematic group by means of a clustering algorithm based on the similarity between documents. In this work the k-means algorithm [11] is chosen, as it is probably the simplest clustering algorithm available and it matches our needs: the number of clusters is dynamically definable by the user and it is fast enough for small sized datasets, allowing the reconfiguration of the clusters in realtime. On the contrary, k-means is not a deterministic algorithm because of the random selection of the initial seeds, and then the algorithm doesn't always converge to the same clusters [12]. This is not acceptable for an interactive system as the user could get confused. That's why we have slightly modified the classical k-means algorithm, by executing it n times, each of them with a different random combination of initial seeds, and choosing the ones that, after executing k-means, minimize the sum of the distances between any document and the centroid of the group to which it belongs.

Finally, in this step we also extract a representative keyword for each document, by retrieving from every vector the feature with the highest value, which represents the meaningful token most frequent in the document. Again, this process is not very precise, but instead it is really simple and fast, which is our goal.

1.2. Assignment of Visual Characteristics

For this work, three attributes of interest were considered: the similarity between documents, the thematic category of each document and the keywords extracted. Taking into account the considerations mentioned in [6], three different visual characteristics were chosen in order to visually represent these attributes.

First, the collection is represented by a tridimensional cloud of spheres, each of them representing a single document. Similarity between documents is visually represented by the spatial distance between the spheres. So, the more the documents are thematically similar, the closer are the spheres. The thematic category is represented by the hue component of color, so that the same category documents have the same color. The hue values of each document are updated every time the number of clusters changes. For the visualization of the keywords, a single label is imposed on the sphere.

In order to obtain a spatial position for every document, the dimensions of the equivalent vectors must be reduced to three, but trying to keep as many similarity information between the documents as possible. In this work the Force-Directed Placement (FDP) technique [13] was used; it simulates attraction and repulsion forces among the documents depending on their similarity measures. Although an FDP approach has big scalability issues, it also has many desirable properties for our goals: good quality of the resulting layout, iterative and realtime positioning process, and ability to be extended by including other factors in the positioning process, as it will be explained in Section 2.2.

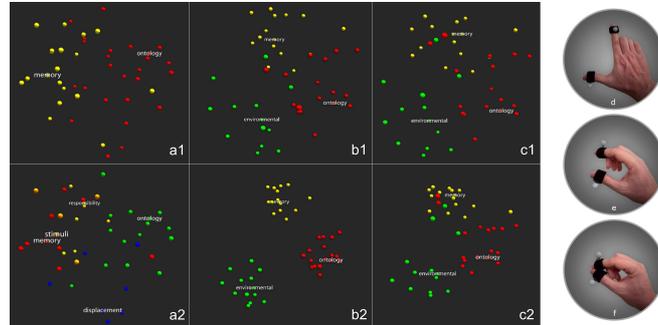


Figure 1. *a.* Changing from two groups (a1) to five groups (a2); *b.* Increasing the intra-cluster force from zero (b1) to maximum value (b2); *c.* Increasing the inter-documents force from zero (c1) to maximum value (c2); *d.* Maximal distance gesture; *e.* Medium distance gesture; *f.* Minimal distance gesture.

2. Interacting with 3D document collections

Besides the visualization of the documents in the 3D environment, it is essential to provide the user with some techniques to navigate and explore the environment. The classical interaction approach is guided by Shneiderman's famous mantra: Overview first, Zoom, Filter and Details-on-Demand [14], which describes the natural way in which users explore the information.

General interaction mechanisms can be categorized into navigation, manipulation and system control. Navigation mechanisms allow users to explore the information across the different levels of abstraction. At the overview level, it aids users to identify the potential interest areas of information, as done by the FishEye technique [15]. Manipulation mechanisms allow users to effectively interact with the elements of the virtual environment. The most common examples are selecting elements and manipulating them spatially within the environment. Finally, control system techniques deal with the modification of the visual mapping of the data or of the state of the system. An example is the removal of a set of elements from the environment or the saving of the current state of the system. These mechanisms must be designed taking into account the user's needs together with the tasks and the specific context of application. Here we present two innovative interaction techniques allowing to manipulate the visual mapping, by improving the identification of thematic groups and the visual perception of document similarities.

2.1. Modifying the Number of Document Groups

The first proposed manipulation mechanism allows users to interactively modify the number of potential thematic groups (see Figure 1.a).

The interest of this technique is based on the assumption that every user has different cognitive and perceptual capabilities, and different preferences. Some users might want to split the document collection into as many specific groups as possible, while others might prefer less and more general groups. The need for higher or lower subdivision even depends on the user task (e.g. building a mental model of the collection vs. searching for a specific document) or on the phase in the pursue of a certain goal (e.g. filter irrelevant information vs. select the best possible source to search a specific information). Our hypothesis is that combining this mechanism with some filtering mechanisms for documents can greatly increase the flexibility of the system while exploring or searching.

2.2. Adjustment of Attraction Forces

This mechanism allows users to modify the relative position of the documents. Although the spatial positioning reflects by itself the similarity between documents, we think the visual clarity of the representation can be improved. In this sense, we propose a modification to the original FDP algorithm, called FDP with Force Control (FDPFC) algorithm, where two new attraction forces, dynamically modifiable by the users, are added.

The first force modifier aims at enhancing the overview visualization. To achieve this, it modifies the spatial position of spheres so documents that belong to the same cluster get closer and are separated from the other clusters (see Figure 1.b). This modification results into a more compact and lean clusters visualization from the overview perspective. Technically, this effect is generated by adding a specific multiplier to the force that controls the attraction between any pairwise documents, and it is applied if and only if both documents belong to the same cluster.

The second force modifier aims at enhancing the inter-document similarity perception, both in the overview and the detail levels. This is accomplished by modifying the spatial position of the spheres so documents whose similarity is higher than a predefined threshold are located closer in space, independently of the previous clustering (see Figure 1.c). Then, users can determine at a glance which documents are similar even if they belong to different clusters. This effect is generated by adding another multiplier to the attraction force between pairwise documents if and only if the distance between them is lower than the mean distance between any other pairwise documents.

2.3. A Gesture-based Interaction Paradigm

As for our daily tasks we use our hands, we think that the manipulation of 3D environments is also more intuitive using them. Therefore, we propose a gestural interaction paradigm where the index and the thumb, equipped with solitary markers, are optically tracked to obtain their spatial position and the distance between them. Both the position and the distance are used to infer which of the three recognizable gestures does the user: maximal distance (thumb and index spread, see Figure 1.d), medium distance (relaxed fingers, see Figure 1.e) and minimal distance (thumb and index pinched, see Figure 1.f). To adapt distances to each user, an initial and directed training phase is carried out. Fingers are virtually represented by two 3D fingerprints (see Figure 2.a), that are immediately identified by the user as representations of his fingers because of the realtime feedback received in terms of position, inclination and rotation while the real fingers move.

In order to test our approach, we have implemented a prototype system with three types of interaction techniques: navigation, selection and control state of spheres. In the first case, we offer three navigation mechanisms allowing the user to reach any point of the environment and to visualize the collection with the desired detail: horizontal and vertical translation of the environment; rotation of the environment with respect to a fixed point; and zoom in and out within the environment. Associated to the zooming mechanisms, and to avoid cognitive overload in the visualization, keywords get visible only when the distance between the spheres they are attached to and the position of the user's avatar (fingerprints) is under a defined threshold. The second group of techniques deals with the selection of one or more documents, as our system is intended both for individual document search and global exploration of a collection of documents. The first one

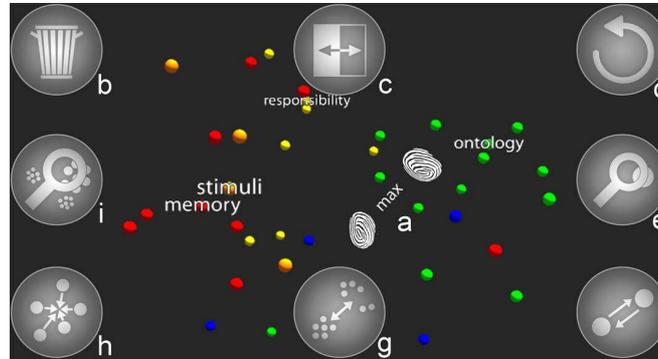


Figure 2. User interface: *a.* Virtual fingertips; *b.* Delete; *c.* Invert; *d.* Undo; *e.* Inspect; *f.* Document force adjustment; *g.* Groups adjustment; *h.* Group force adjustment; *i.* Jump

permits selecting one of the documents, whereas the latter allows selecting all the documents that belong to the same cluster. For increased flexibility, the system allows making incremental selections, that is adding new documents and/or groups to the current selection. We also provide the possibility of deselecting a group that has been previously selected. At last, the system offers eight different functionalities for manipulating the collection, implemented as pop-up widgets (see Figure 2). To avoid undesired executions of these functionalities during navigation, widgets have to remain activated by the user for at least 2 seconds in order to become operational. Most of the widgets have a binary behaviour and their associated functionality is executed just by activation. The first case is jumping between groups in order to focus the visualization over the centroid of one group. If more than one group has been created, every activation results in focusing on a new group, which is a great help when navigating and exploring over groups and not over individual documents. In addition, the system offers four control functionalities that require a previous selection of a document or a group: consult the content of a document, invert the selection, clear the selected documents or groups from the visualization or recover the last documents or groups cleared. The last three functionalities require a numerical value, so, besides activation, their widgets require the user to indicate a number, that is also provided by finger movements. This is the case of the dynamic adjustment of the number of groups for the clustering algorithm (see Section 2.1), and the adjustment of both force modifiers added in the FDPFC approach (see Section 2.2).

3. Experimental Evaluation

3.1. Experiment Design

A user evaluation was designed in order to evaluate the impact of the modifications introduced by the FDPFC and of the gestural interaction techniques proposed. An experiment was conducted in which one independent variable, the type of 3D mapping method, was manipulated with two possible levels: Force Directed Placement (FDP), and Force Directed Placement with Force Control (FDPFC). A total of 36 subjects were selected for the experiment and randomly assigned to one of the two experimental treatments. All subjects were computer science students or professors, so they were assumed to be

able to understand the abstraction applied in the 3D representation of document collections, and to quickly learn and apply new interaction techniques. Ages of the participants ranged from 21 to 40, with 25 males and 9 females. The measured dependent variables were: Precision in the mental model resulting from exploration (M), Effectiveness in search (E), and Satisfaction (S). The null hypotheses were: H01) The mean precision in the mental model obtained from exploration is the same for FDP and FDPFC ; H02) The mean effectiveness in search is the same for FDP and FDPFC; H03) The mean satisfaction is the same for FDP and FDPFC. Statistical t-tests were applied to evaluate these hypotheses.

Precision of the mental model was evaluated after an exploration task. The experimental document collection was manually composed by intentionally selecting documents within a set of predefined topics. Precision of the mental model was measured by presenting, after some minutes of free exploration, a brief thematic questionnaire with nine questions in the form "Were there any documents in the collection talking about *topic*?" A 5 levels response scale was presented with the meaning -2:"definitely no", -1:"probably no", 0:"I do not know", 1:"probably yes", and 2:"definitely yes". The reference topics for the nine questions were Ontologies, Buildings, Bridges, Animals, Pedagogy, Software Agents, Memory, Art, and Civilizations. The hypothesis was that the availability of force control mechanisms in FDPFC would allow the subjects to construct a more precise model of the thematic structure of the collection, particularly by strengthening the intra-cluster attraction force.

Effectiveness in search was measured via a search task in which the participant was asked to find within the collection a representative document talking about a particular topic, as well as two additional documents that were the closest to it in their thematic content. The best possible solution for the task was pre-calculated. Taking the inter-document distance matrix, the effectiveness in search was computed as the distance between the solution document provided by the experimental subject and the optimal solution. For the evaluation of the selected neighbors we decided to count the number of best neighbors provided. The hypothesis was that the availability of force control mechanisms in FDPFC would allow the subjects to find a higher quality set of documents, particularly by strengthening the inter-documents attraction force. We hypothesized that the increase in quality would be more significant when considering the neighbors than just by comparing the best representative.

Finally, Satisfaction was evaluated by a questionnaire at the end of the experience. 15 questions were common for both experimental groups, with 3 additional questions for the FDPFC group evaluating the perceived usefulness of the force control capabilities, and 2 additional questions for the FDP group trying to find out the usefulness they anticipated for these features. The questions offered a 5 levels response scale with the interpretation 0:"I strongly disagree", 1:"I disagree", 2:"Neutral", 3:"I agree", and 4:"I strongly agree". There were four open questions investigating previous user knowledge of document collection exploration and search tools, and requesting a general evaluation of positive and negative aspects.

In the test arrangement the subjects were standing in front of a projection screen (see Figure 3.a), with two fingertips (see Figure 3.c) tracked with an optical system (see Figure 3.b). The experiment was conducted along four days, with half an hour assigned for each subject. Each subject went through the following stages: 1) Preparation: Before entering the experimental area a document was handed over in which a general description

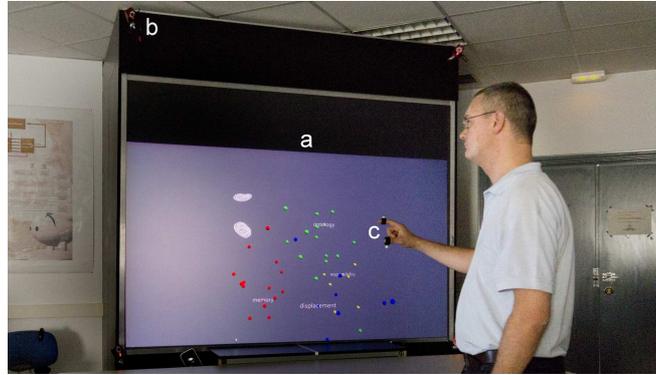


Figure 3. VR installation: *a.* Stereoscopic screen; *b.* Optical trackers; *c.* Optical markers

of the experiment's goal and procedure was presented; 2) Training: The participant was led to the projection screen and ten minutes were allocated for the participant to familiarize with the visualization display and interaction techniques by using a training collection; 3) Exploration: The test collection was loaded and the participant was instructed to freely explore it trying to get an idea of the documents' topics during five minutes; 4) Mental model evaluation: The participant was taken to a separate room and five minutes were left for completing the thematic questionnaire; 5) Search: The participant was taken again in front of the projection screen and the search task was completed in five minutes; 6) Questionnaire: Again at a separate room, the participant was requested to complete the satisfaction questionnaire and was thanked by their collaboration. No time limit was imposed at this stage.

3.2. Results and Discussion

When evaluating precision of the mental model, if a topic presented to the user is certainly included in the collection, the correct response is considered to be 2, while it should be -2 if the topic is not in the collection. Any other response is considered incorrect. The topics that were certainly in the collection (Ontologies, Bridges and Memory) were noticed by the majority of the users (with more than 10 correct responses out of 18 participants in each topic and experimental group -FDP and FDPFC-, and a mean of 13 correct responses in both cases). If we consider topics that were not included and were semantically quite far from other topics in the collection (Animals, Art and Civilizations), users found it difficult to assert with certainty that the topic was absent (even if they tended to believe the topic was not in the collection), so the percentage of correct responses is low, but we can notice less uncertainty in the group of users who could control the attraction forces (FDPFC), with higher success rates (mean of 4 versus 2,33 correct responses out of 18 participants). In the case of topics that were not in the collection but were intentionally close to other topics in the collection (Buildings, Pedagogy and Software agents), the success rate is the lowest, as expected, with a slightly highest success rate in the FDPFC group (mean of 1 versus 0 correct responses). In both cases, the reduced amount of time left for exploration (5 minutes) as well as the fact that users only had access to a keyword and the abstract of the documents, can explain the uncertainty that prevented them from asserting the absence of the topics. The test on the equality between the mean number of

correct responses in the FDP and FDPFC conditions did not allow us to reject the null hypothesis ($\sigma=0.32$).

In the evaluation of search effectiveness, two tasks were demanded. The mean distance between the users' selected document and the optimal one was similar in both groups (0,29 for FDP vs 0,26 for FDPFC in the first task, and 0,09 for FDP vs 0,11 for FDPFC in the second one). No significant difference was found between FDP and FDPFC, with both groups being able to find approximately half of the best neighbors.

Regarding user satisfaction, if we analyze the mean response value for each of the Likert format questions, we can see that all means are higher than 2 (the neutral point in the scale), showing a generally favorable opinion. For the overall questionnaire, the mean response value for the FDP group is 3.12, while it is 3.29 for the FDPFC group, slightly better but still the difference is not significant ($\sigma=0.13$). The most remarkably positive opinions (above 3.5 points) were about the usefulness of keyword visualization, the possibility to jump the viewpoint from one group to another, and the possibility to select a group of documents and remove it. The most valued feature for both groups was keyword labeling. In both experimental conditions, users believed they could make more benefit from the system with more training and experience.

Regarding the features that were available for the FDPFC group but not for the FDP group it is interesting to see that the participants of the latter one did not value much the possibility to include them, despite not having any experience with them (with mean value of 2,31, close to 2-neutral), while the group that effectively had the opportunity to enjoy the features certainly valued much more their usefulness (with a mean of 3,32).

4. Conclusions and Future work

We presented some contributions towards a more intuitive and effective interaction with 3D visualizations of unstructured document collections. A prototype system has been developed fully automating the information visualization pipeline. A modification of the classical FDP algorithm has been proposed (the FDPFC variant) to determine the spatial position for each document's spherical representation in the 3D space. This modification allows the user to control the attraction forces to be applied among documents so that group separation or inter-document similarity can be visually enhanced, as required for the task at hand. A full set of interaction techniques have been implemented through a new gesture-based interaction paradigm based on tracking the position of two hand fingers (index and thumb) by using solitary markers. Navigation, selection and manipulation of both individual documents and document clusters, as well as adjustment of the visualization settings (number of clusters and attraction forces), can be achieved just with very basic hand movements and a set of reactive zones in space.

After experimentation FDPFC seemed to help decrease the uncertainty about the thematic structure of the collection, although no conclusive evidence was found that FDPFC allows the construction of a more precise mental model. Partly this could be explained by the fact that the exploration task was the first experience of the users with the system after a really very brief training period (just ten minutes), and the participants may have felt overwhelmed by the new way to interact with the system, the number of functionalities offered and the tasks required. In fact we observed very limited use of the force control options during exploration. We believe that a longer training period in

which users have the opportunity to really understand the effect and possibilities of each interaction mechanism could result into more effective use of the force control features and a more significant effect on the precision of the mental model.

Regarding the user satisfaction and subjective perception, keywords were the most valued option. This is probably due to the fact that it was the most evident way to look for the required documents in the search task. Some users applied an almost exhaustive search strategy, inspecting every document's keyword, or even explicitly consulting all documents' contents. This strategy could suffice in our limited-size experimental collection, but we believe that if faced with much more populated collections, users would start to realize the usefulness and higher efficiency of alternative manipulation techniques.

In our future work we plan to experiment with bigger and more complex document collections and with a longer and more explicit training period that really helps users to become proficient in the use of the interaction techniques, to face users to situations in which the application of the new interaction techniques really becomes a worthy alternative to more evident strategies, and to avoid results to be biased by the initial user's disorientation.

References

- [1] Fairchild K, Poltrock S, Furnas G. SemNet: Three-dimensional graphic representations of large knowledge bases. *Cognitive Science and its Applications for Human-Computer Interaction*. Lawrence Erlbaum Associates; 1988.
- [2] Robertson GG, Mackinlay JD, Card SK. Cone Trees: animated 3D visualizations of hierarchical information. In: *Proc. SIGCHI conference on Human factors in computing systems: Reaching through technology*. CHI '91. ACM; 1991. p. 189–194.
- [3] Risch JS, Rex DB, Dowson ST, Walters TB, May RA, Moon BD. The STARLIGHT information visualization system. In: *Proc. IEEE Conference on Information Visualisation*. IV '97. DC, USA: IEEE Comp. Soc.; 1997. p. 42–49.
- [4] Card SK, Mackinlay JD, Shneiderman B. *Readings in information visualization: using vision to think*. San Francisco: Morgan Kaufmann Pub.; 1999.
- [5] Mazza R. *Introduction to information visualization*. London: Springer-Verlag; 2009. Riccardo Mazza.
- [6] Shirley P, Marschner S. *Fundamentals of Computer Graphics*. 3rd ed. MA, USA: A. K. Peters, Ltd.; 2009.
- [7] Porter MF. An algorithm for suffix stripping. *Program*. 1980 Oct;3(14):130–137.
- [8] Abney S. In: Young S, Bloothoof G, editors. *Part-of-Speech Tagging and Partial Parsing*. Kluwer; 1996. p. 118–136.
- [9] Nadeau D, Sekine S. A survey of named entity recognition and classification. *Lingvisticae Investigationes*. 2007 Jan;30(1):3–26.
- [10] Salton G, Wong A, Yang CS. A vector space model for automatic indexing. *Commun ACM*. 1975 Nov;18(11):613–620.
- [11] Macqueen JB. Some Methods for classification and analysis of multivariate observations. In: *Proc. 5th Berkeley Symposium on Math, Statistics, and Probability*. vol. 1. UC Press; 1967. p. 281–297.
- [12] Davidson IS. Speeding up k-means clustering by bootstrap averaging. *Workshop on Clustering Large Data Sets*. 2003;.
- [13] Fruchterman TMJ, Reingold EM. Graph drawing by force directed placement. *Software: Practice and experience*. 1991;21(11):1129–1164.
- [14] Shneiderman B. The eyes have it: A task by data type taxonomy for information visualizations. In: *Proc. IEEE Symp. on Visual Languages*. IEEE; 1996. p. 336–343.
- [15] Furnas GW. Generalized fisheye views. In: *Proc. SIGCHI conference on Human factors in computing systems*. CHI '86. New York, NY, USA: ACM; 1986. p. 16–23.