From culture to text to interactive visualization of wine reviews

Kerren, Andreas; Kyusakova, Mimi; Paradis, Carita

Published in:
Knowledge visualization currents: from text to art to culture

2013

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Chapter 5

From Culture to Text to Interactive Visualization of Wine Reviews

Andreas Kerren,[1] Mimi Kyusakova,[1] and Carita Paradis[2]

[1]Linnaeus University, Sweden

[2]Lund University, Sweden

Abstract **On the basis of a large corpus of wine reviews, this chapter proposes a range of interactive visualization techniques that are useful for linguistic exploration and analysis of lexical, grammatical and discursive patterns in text. Our visualization tool allows linguists and others to make comparisons of visual, olfactory, gustatory and textual properties of different wines for example from different countries, from different grape varieties, or from different vintages. It also supports the visual exploration of sensory descriptions as well as confirmatory investigations of text and discourse. Besides a more technical discussion of our visualization approach, we also provide a more general overview of text and corpus visualizations and highlight linguistic challenges that we had to address during the development phase.**

5.1 Introduction

Documents and texts from the World Wide Web or from other digital resources present human beings with the challenge of managing and making sense of large amounts of complex textual information *within* cultures as well as *across* cultures. This challenge pervades our private and societal lives, trade and industry, research and innovation. Based on a relatively large corpus of wine reviews, or tasting notes as they are also called, from world famous Robert Parker's *Wine Advocate*, this chapter is a case study of how interactive visualization of textual data and related information might be made useful both in academia and in society at large. Through the lens of the wine reviews, within the broader culture of wine writing, wine production, wine consumption, wine trade and aesthetics, we show that suitable visualization techniques offer the tools to capture discursive usage patterns that represent interactions of different dimensions of language structure that may characterize text and discourse with the obvious implication that these tools can be used also within other socio-cultural practices. The great challenge for wine critics is of course to translate the sensory perceptions evoked during the tasting event into knowledge representations in writing, so that they can be understood by their readers.

In spite of the fact that viticultural knowledge and sensory pleasures related to wine consumption go back thousands of years, the high status of wine as a prestige consumable and an index of social status is a relatively recent phenomenon—"wine is becoming a cultural icon in an emerging hedonistic sub-culture accessible to an ever larger number of consumers" [3, 15, 24]. The emblematic impact of prestige consumables in contemporary society may be seen as a result of social and discourse practices associated with "you are what you say about what you eat" or, in the context of fine wine, "you are what you say about what you drink" [26]. Among wine lovers, it is common that the experience of wine is first and foremost seen as an aesthetic pleasure similar to the experience of art or music. The wine correspondent of The Guardian [11] makes use of a passage from Evelyn Waugh's Brideshead Revisited [48] which highlights the combination of pompous snobbery and paucity of sensory vocabulary. One of the many things that Charles Ryder, one of the main characters in the book, discovers through his relationship with Sebastian Flyte is a serious acquaintance with wine:

> "We warmed the glass slightly at a candle, filled it a third high, swirled the wine round, nursed it in our hands, held it to the light, breathed it, sipped it, filled our mouths with it, and rolled it over the tongue, ringing it on the palate like a coin on a counter, tilted our heads back and let it trickle down the throat [...] it is a little, shy wine like a gazelle. Like a leprechaun. Dappled, in a tapestry window. Like a flute by still water [...] and this is a wise old wine. A prophet in a cave [...] And this is a necklace of pearls on a white neck. Like a swan. Like the last unicorn." [48]

While there are many different types of wine writings, such as general reportages, editorials, advertisements and technical texts produced by oenologists and chemists, wine reviews represent a type of text written by wine journalists and connoisseurs aimed at both wine consumers and producers. Wine is a complex domain of aesthetic knowledge and tasting practice. It is complex in that it involves all of the sensory perceptions, *vision*, *smell*, *touch*, and *taste*, all of which engender emotional reactions and aesthetic responses. It is the task of the wine reviewer to communicate these experiences in a way so that they can be understood by their readers. This means that sensory perceptions have to be transformed into expressions that will have to pass through the readers' cognitive system in order to be interpreted.

As we will see later, wine reviewers' descriptions of the tasting event follow the journey of the wine from its appearance in the glass, through the nose and the mouth, and finally into the gullet. In contrast to the more synthetic or holistic descriptions of wine provided by the passage from Brideshead Revisited, the main body of Robert Parker's wine reviews are what Herdenstam [14] refers to as analytical descriptions. Such descriptions involve a decompositional approach to the tasting experience. The Aroma Wheel is a famous terminological attempt at a descriptor system, using descriptors from the vegetal, chemical and geological spheres. It was developed by oenologists at the University of California, Davis, for descriptions of smell [29]. The Aroma Wheel has been further developed for both smell and taste and for both whites and reds by the German Wine Institute [2]. The use of descriptors from such domains as fruit, minerals and spices is not wine lingo, but a necessity due to the relative lack of specific vocabularies for sensory domains [34, 37]. Aroma Wheel descriptors are primarily limited to objects. However, this is also true of color descriptions, as pointed out by Wittgenstein in *Remarks on Colour*:

> "When we're asked "What do 'red', 'blue', 'black', 'white', mean?" we can, of course, immediately point to things which have these colours, – but that's all we can do: our ability to explain their meaning goes no further." [50]

The aim of this chapter is to propose an interactive information visualization (InfoVis) tool to be used on text. Our contention is that visualization provides overviews and better understanding in equal measure. The tool has been developed on the basis of some 84,000 wine reviews from the *Wine Advocate*.[1] Thanks to the capacity of the tool to handle large quantities of data, it is useful for exploratory purposes as well as for confirmatory investigations of text and discourse, which in the case of the present data will increase our knowledge of the genre of reviewing in general and wine reviewing in particular [35, 20].

The remainder of this chapter is organized as follows. Section 5.2 gives a general overview of the challenges addressed. Next, in Section 5.3, we discuss related approaches within the more general field of information and text visualization as well as more specifically in representing sensory descriptions. In Section 5.4, we describe our database of wine reviews and corresponding metadata. Our own approaches to the visualization of wine tasting notes by using information visualization techniques are presented in Section 5.5. Initial results are briefly outlined in Section 5.6 including a small use case description. We conclude in Section 5.7 and suggest some investigatory paths for future work.

5.2 Challenges

As already indicated in the previous section, wine reviews are descriptions and evaluations of wines written by professional wine tasters. They have a strict rhetorical structure, consisting of three parts, starting with production facts and ending with an assessment and a recommendation of prime drinking time. The middle of the text, which is the most important part, is devoted to an iconic description of the wine tasting procedure from the taster's inspection of the wine's visual appearance through smelling, tasting and feeling its texture, i.e., from *vision* through *smell*, *taste*, and *mouthfeel* (*touch*) [33], cf. sample review (1).

---

[1] https://www.erobertparker.com/entrance.aspx

(1) “This great St.-Estephe estate has turned out a succession of brilliant wines. The 2005, a blend of 60% Cabernet Sauvignon and 40% Merlot, has put on weight over the last year. An opaque ruby/purple hue is accompanied by a sweet nose of earth, smoke, cassis, and cherries as well as a textured, full-bodied mouthfeel. While the tannin is high, there is beautifully sweet fruit underlying the wine’s structure. It will require 8-10 years of cellaring after release, and should drink well for three decades.” (Wine Advocate 170, April 2007)

The visual appearance of the wine in (1) is described in terms of its clarity and color using the descriptors ’opaque ruby/purple’. The olfactory perceptions are primarily described through concrete objects, e.g., ’earth, smoke, cassis, and cherries’, but also in terms of a gustatory property, ’sweet’, while taste and mouthfeel are described through various gustatory and tactile properties (’high’ (tannin), ’sweet’ (fruit), ’textured, full-bodied’). Because almost all wine reviews describe the wines in terms of four different perceptual modalities, i.e., visual appearance, smell, taste and texture, they are a gold mine for linguistic explorations of descriptions of human sensory perceptions in discourse. Of particular interest are the descriptions of olfactory perception. There is no specific olfactory vocabulary, neither in English nor in (most) other languages of the world. Olfactory descriptions have to be made using words from other domains. In wine reviews, words for taste or words for objects such as fruit, herbs or flowers of different color are used. In general, dark objects are used in descriptions of red wines and pale objects describe white wines. In other words, olfactory descriptions are primarily made on the basis of the smell of objects and also their color and taste. Exploring patterns for perceptual descriptors and the context of their use in wine reviews provides useful information not only about the relations between descriptors of odor and other modalities, but also about language, perception and cognition in general [28].

Advances in visualization offer important possibilities for organizing, presenting and analyzing linguistic data, in which case visualization techniques provide a way to view language in other formats than as linear stretches of letters. Visualization techniques offer the tools to capture lexico-semantic usage patterns and to represent interactions of different dimensions of language structure that characterize different texts and discourses. As demonstrated in the beginning of this section, descriptions of wines in wine magazines are short texts with a very strict rhetorical structure. The language of such texts is of interest to linguists at various different discursive levels. Linguists want to know what kinds of words are used to describe the wines’ visual properties, what kinds of descriptors are used for olfactory, gustatory and tactile perceptions. They are interested in what words and expressions are used where in the texts. For instance, what kind of temporal expressions are employed in different parts of a text, and what expressions of personal opinion, such as ’should’, ’drinkable’, ’recommend’ are used where in the texts and why. More generally, linguists take an interest in how all linguistic patterns combine into what might be our understanding of the discourse beyond the text itself. In other words, visual imagery provides a way to represent things that would otherwise go unnoticed. The added value of the visualization tool presented here is that it can be used interactively. The data can be easily explored, and because parameters and combinations of data and metadata can be changed, many questions regarding the potential of the data receive on-the-spot answers. As a result, new patterns emerge that can generate new research hypotheses about language use in different genres and text types.

Our tool supports the visual analysis of the corpus of wine reviews from the Wine Advocate. The wine reviews are available in the form of two databases that contain a large number of wines, metadata about the wines, and the actual reviews. In order for linguists to arrive at a better understanding of different text types, different discourses and their vocabularies, large corpora are of crucial importance. At the same time, it is also a challenge to identify linguistic patterns in large corpora, to organize the data, to make statistical calculations and to present the data to readers in intuitive and clear ways. Our contribution is to find solutions to some of these challenges. The first challenge is that we have to be able to represent large amounts of multivariate data. For that purpose, advanced interaction techniques are essential, because they ensure the opportunity for selecting a subset of tasting notes and for getting detailed information about the tasting notes in order to proceed with further analyses. Secondly, we have to find an efficient way to interactively visualize the text of individual wine reviews, which brings us to the field of interactive text visualization. Thirdly, a number of compatible visualization approaches have to be combined in order to efficiently explore the language used in the descriptions of the wines.

5.3 Related Work

The general design of our visualization tool is based on standard coordinated and multiple view visualization techniques that are presented in Section 5.5.1. An excellent starting point for related work of this kind of visualization techniques is the annual conference series on Coordinated & Multiple Views in Exploratory Visualization (CMV) or the work of Roberts [36]. In order to specify the layout of our tool and to define the functional requirements, we were inspired by the FilmFinder tool for exploring movie databases [1]. It was one of the first tools, which integrated the concept of a two dimensional scatter plot with color coding, filtering, and details provided on demand (dynamic queries). The developers realized different encoding and interaction techniques for the representation of multivariate data.

For the purpose of information visualization of complex textual data, we use different well-known techniques and interaction approaches as described in the next subsection. Here, we decided to provide a more general view on text and document visualization to embed our work in a broader context. Related work that concretely addresses the (visual) representation of wine attributes and/or descriptions is given in Section 5.3.2.

5.3.1 A Brief Overview on Text and Document Visualization

At present, we have access to texts in many different ways. We find them in printed books and documents, in online books, electronic documents like PDFs, the World Wide Web or other digital libraries, patient records, source code of programs, patents, emails, diagrams, etc. The availability of texts and documents is overwhelming, and people want to actively deal with them to solve specific problems. Typical questions are: what documents contain a text about a specific topic? Or, are there similar documents to those that I already have? Sometimes, we only want to search for a single word in a large text, or we look for interesting patterns. Such patterns might show how a text was written by an author, or if it was written by several authors even if not indicated in the header of the document. In this context, the visual analysis of specific metadata, e.g., comments, size, or number of words, is also important. The great interest in text and document visualization is also reflected by the increasing number of conferences and workshops that offer a place to discuss techniques and tools for visual text analysis, such as the Workshop on Interactive Visual Text Analytics for Decision Making held at VisWeek 2011 in Providence, USA, or the AVML (Advances in Visual Methods for Linguistics) conference in York, UK, 2012.

5.3.1.1 Text Representations

Information visualization is capable of supporting the aforementioned tasks in several ways. First, we focus on text visualization, i.e., on tools for and approaches to the visualization of a single document.

Tag Clouds provide information about the frequency of words contained in a text [17]. The approach uses different font sizes for each word in the text to indicate how often a certain word is used in comparison with the other words. Several extensions and related approaches exist, such as *Wordle* or *ManiWorlde* [46, 21]. *SparkClouds* extend the original tag cloud idea with a temporal variable by so-called “sparklines” [23]. Thus, trends can easily be identified and analyzed. In our tool, we use a simple tag cloud implementation to represent the word frequency in a group of wine reviews.

The research project *Many Eyes* provides alternative methods for data analyses using innovative visualization techniques [16]. One of the approaches for supporting text analysis is the representation of a given text as a *Word Tree* [47]. The purpose of this visualization method is to afford an insight into the different contexts in which a word is encountered in an unstructured text. We used this concept in one of the text visualizations of our tool to facilitate rapid exploration of the wine tasting notes (cf. Subsection 5.5.1).

An approach for visual literary analysis, called *Literature Fingerprinting*, was presented by Keim and Oelke [18]. It supports the visual comparison of texts by calculating features for different hierarchy levels and by creating characteristic fingerprints of the texts. Such features might be word/sentence length or measurement of vocabulary richness. A similar idea for the representation of wines is described in Subsection 5.3.2. However, no feature analysis was performed in this case; just wine attributes (color, production year ...) are used for the visual mapping.

Special Cases In the field of software visualization, there are several ideas about how to represent textual information, i.e., the source code in this case. Besides classic techniques such as line indention or code coloring used in order to represent code structure and keywords (depending on the programming language), the well-known *SeeSoft* approach [8] maps each source code line to one pixel row. The color coding of this pixel row represents statistical measurements or other metadata, for example, age of modifications. If the source file is too long to fit on the screen, the vertical arrangement of the rows representing the lines is folded and continues in the next column.

5.3.1.2 Corpora Representations

A collection of documents is usually called a *corpus*. Corpora can be structured to some extent (software packages, wikis …) or relatively unstructured (emails, patents …).

Early approaches, e.g., *Lifestreams* [9], simply arranged documents according to specific attribute values such as time tags. More recent works analyze the documents by metrics, such as similarity, and perform cluster analyses or compute so-called self-organizing maps (SOMs). A SOM (or *Kohonen Map*) applies techniques from the field of artificial neural networks to map n-dimensional data objects to simple geometrical structures, typically in a 2D grid. In our case, the data objects are the documents and the n-dimensional attributes are defined with the help of keywords from the title or the abstracts. In a typical SOM representation like WEBSOM [22], the number of documents that fall into a specific area (i.e., the density) is mapped to a color gradient, and categories are represented by simple labels.

Conceptually similar (by looking at the result) is *ThemeScapes* [49] that follows a natural landscape metaphor. Single documents are categorized and then applied to a document map as topic areas, whereas the documents themselves are shown as small dots. “Mountains” in the landscape represent document concentrations in a thematic environment (density), height lines connect concept domains, and interaction is possible through tool tips, zooming and panning, etc. There are many more recent approaches that make use of the same metaphor, such as [30].

Dengel et al. [7] regard document collections as information space and use a 3D metaphor for the visual representation. Depth is used to distinguish important documents from unimportant ones, i.e., a document is drawn as an icon and appears more in the foreground if it is important. It is also possible to visualize relationships and cluster analysis results using this approach.

In order to carry out comparisons of text documents using tag clouds, Collins et al. [6] introduced so-called *Parallel Tag Clouds*, where tags are arranged on vertical lines for each document. Identical words are highlighted by connection lines. Another way to compare documents was described by Strobelt et al. [44]. They developed so-called *Document Cards* which are compact visual representations that show documents' key semantics as a mixture of images and important key terms. The technique is also transferable to use on mobile devices.

Salton and Singhal analyze the relationships between text documents according to different topics. They developed a tool called *Text Theme* [38] to represent such correlations visually. Single topics can then be identified and be compared with the help of textures or color coding.

In contrast to most of the approaches discussed in this subsection, our tool operates more on the syntactic level, i.e., higher-level themes or entire tasting notes cannot be compared directly. Thus, our tool was originally not designed to focus on the significance of specific entities extracted from the tasting notes but rather on the exploration of their content and linguistic constructions.

Special Cases If the document collection is related to time-series, a usual task is to find thematic changes over time. There are several approaches that address temporal text analysis. One of them is *ThemeRiver* [12] which uses a river metaphor as a representation for the information flow over time. The themes themselves are stripes in the river that start at the first appearance of a theme and end if it is no longer contained in any documents like a pigment thrown into a river and distributed in the streaming. The thickness of a stripe represents the number of documents with that theme.

Stasko et al. developed a visualization tool for analyses of textual reports called *Jigsaw* [43]. The goal of their tool is to aid investigative analysts to faster understand the content of reports in order to predict possible threats and to prepare defensive plans accordingly. The main analysis units in their work are pre-defined entities in the texts and the goal of the implemented visualizations is to represent relations and connections between these entities.

5.3.2 Representing Wine Descriptions and Tastings

Using both corpus methodologies including visualization of the data and experimental psychophysical techniques, Morrot et al. [28] investigated the interaction between visual appearance and odor determination in wine description and wine tasting. Their work presents the results of a study carried out with the help of a tool called ALCESTE, which is based on statistics about the distribution of words in a corpus of text to determine groups of words that co-occur in the same context.

They found that the descriptors used to characterize white and red wines were different in terms of the colors of the objects used in the descriptions respectively (i.e., dark objects describe red wines and pale objects white wines). In addition to the corpus study, they also carried out a psychophysical experiment, which confirmed the corpus data, demonstrating the impact of vision on the human odor perception. In comparison to ALCESTE, our visualization tool gives users more possibilities to browse the text, to filter out uninteresting cases, and to interact with the visualizations. Thus, it does not afford pure statistical numbers only, but gives analysts an opportunity to explore the data set and to get a better understanding of the structure and content of the texts.

Another visualization approach, called *Wine Fingerprints*, has been discussed by Kerren [19]. In contrast to the tool presented in this paper, *Wine Fingerprints* focus on wine attributes, such as wine color, rating, grape type, price, or aroma, and not on the actual wine reviews. These data form a multivariate data set, part of which can be hierarchically structured into a so-called aroma hierarchy. Two different types of fingerprints based on a rectangular or radial drawing approach respectively are shown in Figure 5.1. The *Wine Fingerprints* approach has various applications for business and industry in that it can create visual patterns of combinations of wine attributes and support comparisons of visual, olfactory and gustatory properties of different wines from different parts of the worlds, from different grapes, from different vintages etc. Both customers and companies can make visual comparisons of wines and select wines on a pictorial basis instead of on the basis of a list of multimodal perceptual attributes.

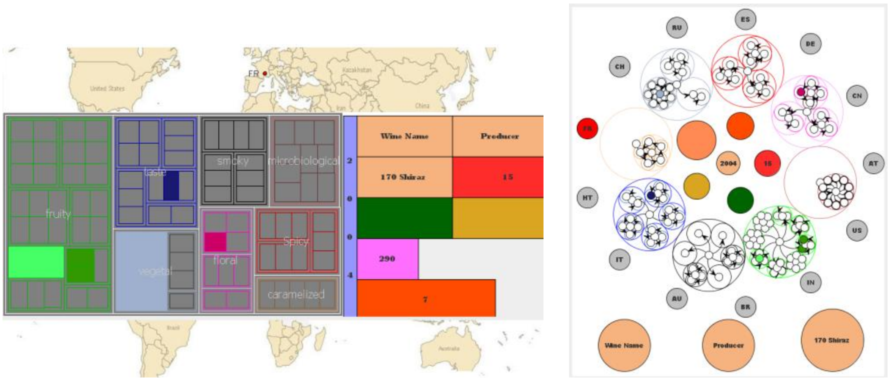


Fig. 5.1 On the left-hand side, a wine fingerprint is shown using a rectangular drawing approach as well as a *treemap* layout [40] for the aroma hierarchy. On the right-hand side, a wine fingerprint is displayed using a radial drawing approach as well as a balloon tree layout for the aroma hierarchy. (Taken from [19])

5.4 Notes on the Data Set

Text annotation by part of speech, so-called tagging, and sentences structure exploration of the wine reviews are indispensable for analyses of the sensory vocabulary and the context of perceptual descriptors use. Word classes (also called parts of speech) are linguistic categories of lexical items defined by their syntactic distribution in natural languages. Each word class has a corresponding abbreviation called word-class tag or word tag [39]. Typical examples of word classes are ”adjective” and ”noun” with abbreviations ”JJ” and ”NN” respectively. Word classes are essential for some of the approaches we apply in tasting notes analyses as they are a basic criterion we use for organizing the data in text visualizations.

The wine tasting notes are stored in two databases that contain information about different wines as well as the tasters’ comments about them. In each database, the tasting notes are represented in different ways. The first database contains descriptive information about the wines, their unique ID number, their origin, vintages, wine ratings, dryness, color and the complete original wine review. The second database contains the same tasting notes including ID numbers, but they are segmented into words with their respective word-class tags. The latter database was built from the former, the original database, by using the *WineConverter* tool, developed by Ekeklint and Nilsson from the language technology group at Linnaeus University in Sweden (formerly Växjö University). The result of this segmentation is a new structuring of the wine tasting notes where each word is described by additional information that accurately specifies its position in the text of the full tasting note. The location of each word in a tasting note is determined by the following information: ID number of the tasting note, number of the corresponding sentence in the tasting note, position of the word in this sentence, the word itself, and the word tag given to this word.

In order to get a better overview of appropriate visualization approaches for rep-resenting the tasting notes and their attributes, we had to take the great amount of analyzed data into consideration. Table 5.1 provides a list of substantial statistical numbers derived from the data set to give an idea about the sheer quantity of the data to be visualized.

Table 5.1 Statistical numbers derived from the wine databases.

Number of tasting notes	84,864
Total number of words used in the tasting notes	8,332,666
Number of different words used in the tasting notes	46,000
Maximum length of the tasting notes	496
Number of word classes	43
Number of vintages	104
Range of wine rating values	1 to 100

5.5 Visualization Framework

In order to provide an overall perspective of the analyzed wine reviews, we follow Ben Shneiderman's mantra of information visualization: "overview first, zoom and filter, details on demand" [41]. This gives users an initial overview of the data explored and the possibility to proceed with investigations of its subsets. For this, we combined several visualization approaches to achieve our goals: scatter plots, tag clouds, word trees, bar charts / histograms, and a world map. The scatter plot is used to be the main entry point for using our tool as described in the following sections.

5.5.1 Visual Representations

Scatter Plot The purpose of this visual representation is to give a first overview of the data. Because of the large number of tasting notes (cf. Table 5.1), we decided to use a scatter plot for their initial display, i.e., each single tasting note is represented by a blue circle. This approach also saves space and gives an idea about the distribution of the tasting notes on the basis of the values of two selected wine attributes, see Figure 5.2(a). Attributes currently supported by the scatter plot visualization are all possible pairs of "Wine Rating", "Wine Vintage", "Color Class", "Tasting Notes Length", and "Wine Country".

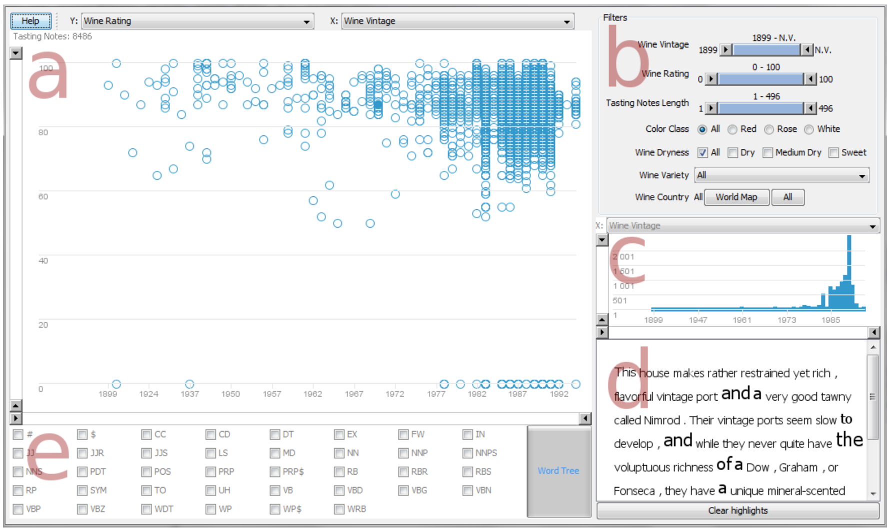


Fig.5.2 A snapshot of the main window of the application after starting. Note that one tasting note was selected in the scatter plot; its tag cloud is shown in the bottom right corner.

Filter Panel After getting an overall view of the data, users are given the opportunity to interact with the visualizations in order to find the best subset of elements to be further analyzed. *Dynamic Queries* are a widely used concept in information visualization for aiding the solution of this task [1]. This approach offers a simultaneous display of query and result. It is useful for dynamic exploration of the data. Following this concept, different types of filters were integrated for some of the attribute values to be used for finer delimitation of the visualized elements on the scatter plot, see Figure 5.2(b). More details on the effect of various filter possibilities are given in Subsection 5.5.2.

Bar Charts and Histograms Getting statistical information helps analysts to better understand the visualized data and to find the desired set of tasting notes. Bar chart diagrams are traditional approaches for statistical data visualization. In this work, they are supplied in order to show the number of tasting notes that correspond to the values of a specific wine attribute (Figure 5.2(c)).

Text Visualization The visualization approaches that we apply for the representation of tasting notes are word trees and tag clouds.

Word Tree The word tree visualization facilitates rapid querying and exploration of text bodies [47]. In our tool, a word tree describes the sequence of words and phrases used in a group of tasting notes. The structure of the word tree is organized into two main groups of nodes: word tags and words. There are three prerequisites for proceeding with the word tree visualization:

- users need to select a group of tasting notes for further analyses,
- a specific tasting note for deriving the initial data (from now on referred to as *root tasting note*), and
- the word classes of the words in this root tasting note that they would like to analyze (to be chosen with the help of the tag checkbox panel shown in Figure 5.2(e)).

The first three levels of the word tree contain data from the root tasting note. The other levels consist of data from the whole group. The color of the text of each node indicates whether it is contained by the root tasting note or not. Red color of the text means that it is part of the root tasting note, but it can be met in other tasting notes as well. Black color of the text implies that it is certainly not contained by the root tasting note. The root node of the tree is artificially added, and it contains the static text "Tags", which suggests that the following level is composed of word tags. The second level contains the selected word tags that correspond to words in the root tasting note. The third level consists of all the words from the root tasting note that belong to the word tags on the previous level. Figure 5.3 gives an example of the word tree and the organization of its nodes. The levels of the tree alternate with each other to represent either word tags or words that correspond to the tags on the previous level. The children of each node representing a word class are the words from the analyzed group of tasting notes that belong to this word class. For instance, the word tree in Figure 5.3 displays two (selected) word tags of the root tasting note, i.e., “JJ” (adjectives) and “NN” (nouns, singular common). By looking at the children of “JJ”, the user can see that the root tasting note has four adjectives, e.g., “coarse”. Then, by looking at the next two deeper levels, the user can see that “coarse” has two successors: one noun (plural common; “NNS” → “flavors”) in another note from the analyzed notes group (black) and one determiner (“DT”→ “, this”) in the root tasting note.

Our word tree visualization represents a large data set of words and word tags. It is restricted by the size of the display and people’s perceptive capabilities. To cope with these restrictions, our implementation applies the idea of Degree-Of-Interest (DOI) trees that provide a solution of these problems. They combine focus & context visualization techniques and degree-of-interest calculations to find a proper layout that fits within the bounds of the display. The technical idea is the use of a DOI function, which assigns a number value (DOI value) to each node indicating how interested the user is in this node. This value is then used as a criterion to determine, which of the nodes should be visible, which of them are in focus and how they should be displayed [13, 4]. The nodes in focus have the greatest DOI value and are slightly magnified. The size of all other nodes is directly proportional to their individual DOI value. An exception to this rule is the tree element that was selected last, which is the most magnified element, in spite of the fact that it has the same DOI value as the other focus nodes. Figure 3 demonstrates a degree-of-interest tree where the DOI values of the nodes are given in brackets on the right side of the node label.

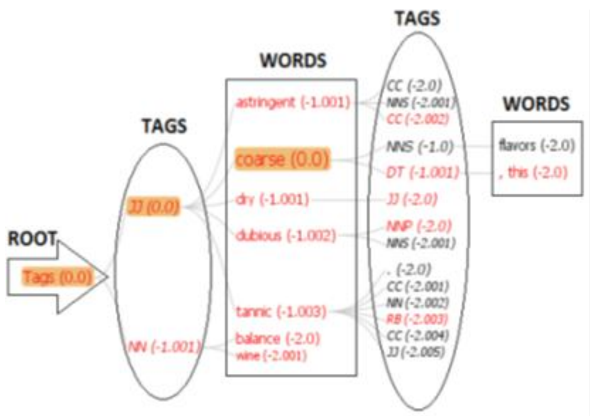


Fig. 5.3 A word tree that shows the node organization into two main groups: word tags and words. Furthermore, there is another partition of the tree nodes as well: nodes that contain data from the root tasting note (text colored in red) and nodes that contain data only from the other tasting notes (text colored in black). The DOI value of each node is given in brackets on its right side.

Tag Clouds The tag cloud visualization makes use of different font sizes for the words in a corpus of texts to give a hint about the frequency of their usage. There are two prerequisites for the application of tag cloud visualizations of a tasting note. A group of tasting notes needs to be defined for further analyses, and one of them has to be selected for its text visualization. The text of the selected tasting note is then visualized by using the tag cloud metaphor, where each word has a different font size depending on how often this word occurs in the whole group of tasting notes, cf. Figure 5.2(d).

World Map Visualization The origin of wines is important information and should be visualized in a way that gives the user a rough overview of wine-producing countries. A natural approach for visualizing it is an interactive world map indicating the density of wine production in different countries. Figure 5.6 shows a world map representing information about the wines produced in different parts of the world that have been tasted and described in the tasting notes. The color saturation is directly proportional to the density of wines produced in each country.

5.5.2 Interaction and Coordinated Views

We combined the visualization approaches described in Section 5.5.1 with appropriate techniques for user interaction to build an efficient tool for analyzing wine tasting notes. The subsequent subsections give a notion about the user interface and the overall layout of the application. More precisely, there are five particular views intended to build an efficient overview visualization as displayed in Figure 5.2:

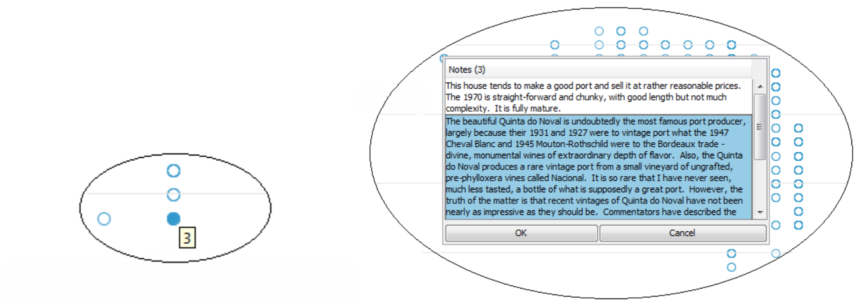
1. a scatter plot (showing the distribution of tasting notes),
2. filters (to reduce the complexity by filtering),
3. bar charts and histograms (to show statistical data),
4. tag clouds (for text visualization), and
5. tag checkbox panel (to select specific word classes).

The above views (1-5) are coordinated by standard highlighting and brushing techniques.

Distribution of Tasting Notes The scatter plot axes on the left and bottom sides of the display correspond to one wine attribute each. Range sliders [5] are added to the axes in order to make it possible for the users to change the range of the wine attribute values and therefore the scope of the tasting notes visualized in the scatter plot. The number of visible tasting notes can be observed at the upper left corner of the scatter plot (8,486 in our screenshot example of Figure 5.2(a)). Another possibility given to the user is to change the wine attributes plotted on the *x*- and *y*-axis by selecting other attributes from the combo boxes at the top of the display.

There is a drawback that appears as a consequence of the scatter plot concept and the data stored in the database: it might happen that more tasting notes share the same values for both of the wine attributes plotted on the axes. Such tasting notes overlap when they are visualized at the same spot in the scatter plot. This makes the selection of an element from the display more complicated. We added a tooltip to each element to give the user a hint about the number of overlapping tasting notes at the specific position (Figure 5.4(a)). Thus, an individual element can be selected from a popup list of the overlapping tasting notes, as shown in Figure 5.4(b). The selected tasting note differs from the others because of the filled blue circle in the scatter plot as well as the blue background color in the popup list. In case of overlaps, the selected element is visualized on top of the others, which guarantees that it is always visible.

Instantaneous Response of the System Information exploration is an interactive process between users and visualizations. Thus, it needs to be well supported and incorporated into the system's potentialities. This process consists of series of questions and answers, and furthermore, each succeeding question depends on the prior answer [42]. The means that our visualization tool supplies for asking questions are filters. To provide an adequate environment for data analyses, the system performs an instantaneous response to the users' actions. Changing the value of a filter has an immediate effect on the displayed visualizations and users do not need to perform any superfluous actions to request refreshing of the displays. This feature is essential for performing efficient data analyses.



(a) Tooltip help (b) Tasting note selection

Fig. 5.4 Overlapping tasting notes in the scatter plot view.

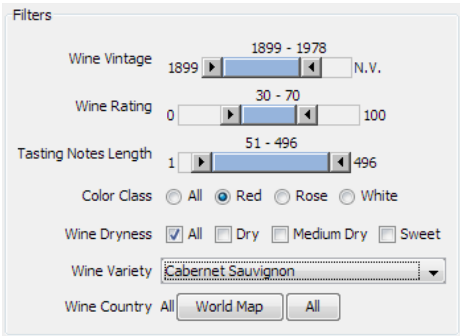


Fig. 5.5 Types of filters implemented in the application.

Filtering Filters are used to facilitate the task of the users to interact with the visualization and to find the best subset of elements to be further analyzed. Figure 5.5 shows a screenshot example of filters supported by our tool, which in this case selects only those tasting notes whose corresponding wines have a vintage between 1899 and 1978, a rating between 30 and 70 points, a length between 51 and 496 words, red color, no specific dryness, geographical information, and are made from Cabernet Sauvignon grapes.

The world map filter is a realization of the geographic visualization approach described in Section 5.5.1. It provides users with the opportunity to filter out tasting notes on the scatter plot depending on their origin (Figure 5.6). Different standard functionalities are supplied to assist working with the map like zooming in, zooming out, and panning to a specific region of interest. Users have the possibility to select a country on the map and our tool visualizes only those representations of tasting notes of wines produced in the specified region.

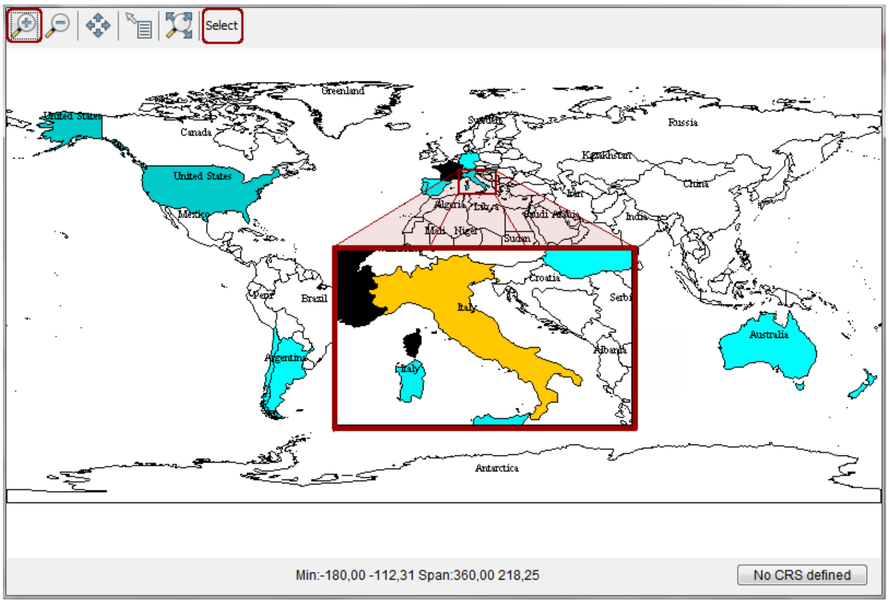


Fig. 5.6 World map providing information about the wines produced in different countries. It is also possible to use this view as an interactive filter for the specification of single countries.

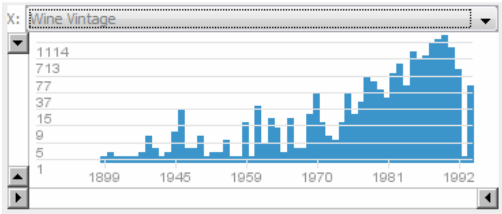
To provide better software maintenance, we use a specific property file that contains a list of wine attributes and their required filter types. In this way, filters are dynamically created on the basis of this information and can be easily added or removed.

Statistical Information The property file also contains a list of wine attributes that can be represented by bar chart diagrams. Figure 5.7 presents snapshots of histograms implemented in the application. An individual bar chart or histogram is created for each of the listed attributes showing the number of visible tasting notes correspond-ing to each of their values (Figure 5.7(a)). Only one of the diagrams is visualized at a time in order to save space. We added range sliders to the *x*- and *y*-axes to assist users in changing the range of visualized attribute values and to get a closer look at a specific section of the diagram, see Figures 5.7(b) and 5.7(c) where the vintage range was modified.

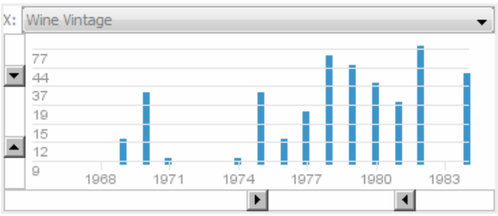
Word Frequency Analysis After the selection of a tasting note in the scatter plot view, its text is visualized using a tag cloud approach (cf. Section 5.5.1). The font size of each word is estimated according to the frequency of its occurrence in all elements visible at the same time, including the selected one. Figure 5.8(a) shows a tag cloud example generated by our application.

The tag checkbox panel contains all word tags available. A coordinated interaction exists between the tag cloud view and the checkbox panel. On the one hand, when the user selects a word from the tag cloud visualization, all words of the same class, i.e., with the same word tag, are highlighted in the tag cloud together with the tag itself in the tag checkbox panel. Figure 5.8(b) demonstrates this interaction after selecting the word “Last” in text of the tasting note. On the other hand, when a word tag is checked in the checkbox panel, such as “DT”, it is highlighted together with the words corresponding to this tag in the text.

(a) Attribute selection



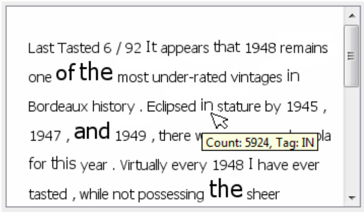
(b) “Wine Vintage” histogram



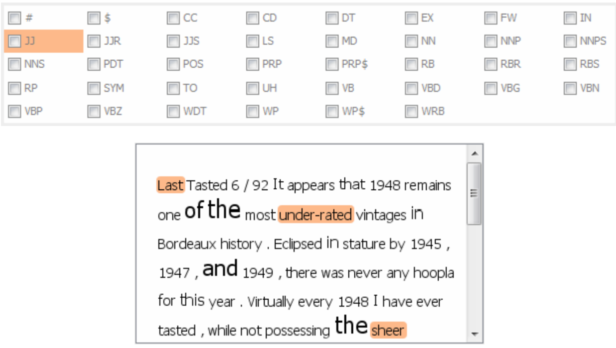
(c) “Wine Vintage” histogram with modified range

Fig. 5.7 Screenshots of an interactive histogram for attribute “Wine Vintage”.

Sentence Structure Analysis The basic concept and structure of the word tree visualization was described in Section 5.5.1. Figure 5.9 presents an additional example of a word tree generated by our system. The visualization consists of three basic components: (a) a display containing the word tree, (b) a text area presenting the text of the root tasting note, and (c) a text area presenting the currently constructed sequence of words. All nodes that build a path from the root node to the currently selected node are in focus. Selecting a node from the tree changes the focus to the nodes contained by the path from the root to this node. A smooth animation is used to change the state of the tree to the newly selected focus [13]. The nodes in focus are highlighted with another background color and slightly enlarged. In the example, the node selected is “raspberries”, and therefore, all nodes from the root to the node “raspberries” are in focus. These nodes constitute a sequence of words which forms part of one or more tasting notes in the current scatter plot. This sequence is displayed at the bottom of the word tree, and it is also highlighted in the root tasting note, if included there.



(a) Tag cloud visualization implemented in the application



(b) Tag cloud interaction together with the tag check box panel

Fig. 5.8 Word frequency analysis

In Figure 5.9, the actual sequence of words is “glass, offering aromas of ripe raspberries.”. The node labels are in red since they are contained in the root tasting note. Often, the tree depth and width exceed the display bounds. It is not possible for all the nodes to be visualized in the space available. In order to surmount such problems, different techniques are integrated into the visualization, e.g., zooming and panning controls [13].

There is a close relation between word tree visualization and the scatter plot. The word tree is constructed according to all combinations of words beginning with words from the root tasting note and followed by words from the whole group of tasting notes visualized in the scatter plot. This means that each sequence of words specified by the word tree exploration is contained in at least one tasting note of the current scatter plot selection. This relation is indicated by highlighting (using a black circle) those tasting notes in the scatter plot that contain the sequence of words construed by the word tree (Figure 5.10(a)). The tool makes sure that highlighted elements are always visible. To distinguish them, their texts are in blue in the popup list of overlapping tasting notes. In the example given, there is only one tasting note that contains the sequence “glass, offering aromas of ripe raspberries" and its text is in blue in the popup list, see Figure 5.10(b). It should be noticed that this is the same tasting note that has been selected for a root tasting note of the word tree since its background color is blue as well.

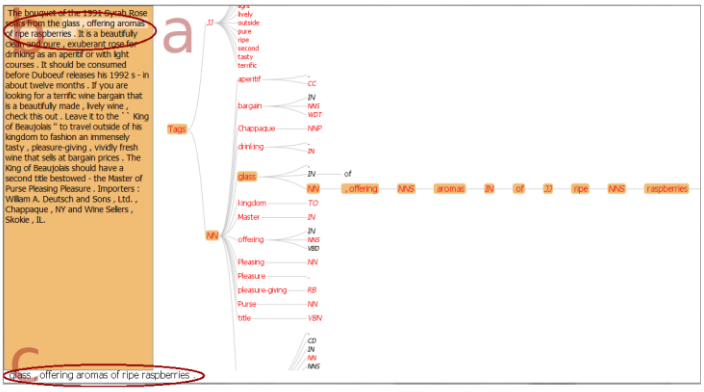
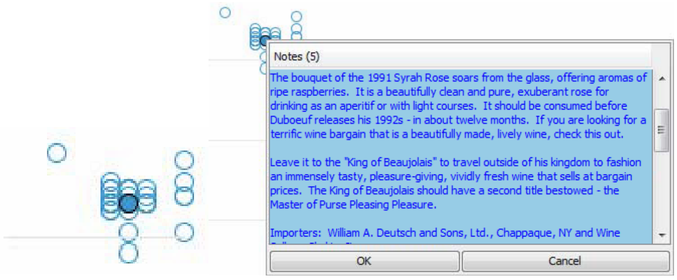


Fig. 5.9 Word tree visualization consisting of three basic components. The tree node labels corresponding to words that are contained by the root tasting note are colored in red.



(a) (b)

Fig. 5.10 Highlighting of tasting notes in the scatter plot that contain the selected sequence of words constructed by the word tree.

5.5.3 Implementation Aspects

The tool's software architecture can be represented by four logical layers as shown in Figure 5.11. Because the original database containing the wine tasting notes was created using Microsoft Access®, we decided to continue to use this database management system (DBMS). The programming language that we applied for the implementation of the application is JAVA. We used four open source JAVA libraries to implement the required functionalities. The JDBC library was employed for establishing connectivity between the JAVA programming language and the database [32]. The graphical user interface was created with the aid of the

JAVA Swing Toolkit [31]. We made use of the *Prefuse* Toolkit for the following interactive visualizations: the scatter plot, the bar chart diagrams and the word tree [45]. The world map visualization was created by the functionalities of the JAVA GIS Toolkit *GeoTools* [10].

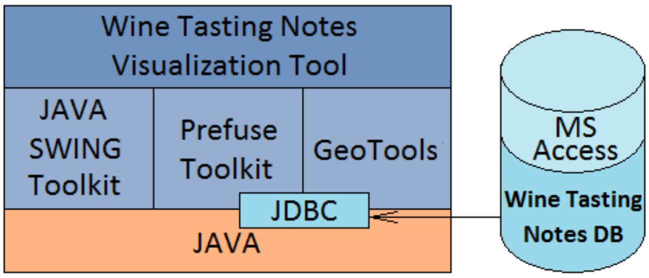


Fig. 5.11 Software architecture of the tool.

Scalability Issues The selected visualization approaches are appropriate for representing large amounts of data. From a theoretical point of view, there is almost no restriction placed upon the number of visualized tasting notes. The scatter plot and the bar chart diagrams provide different interaction techniques that give users the opportunity to focus on a subset of elements for further exploration. The tag cloud visualization represents the text of one tasting note using different font sizes for its words. With an increase in the number of tasting notes, the proportions between the font sizes of the words in the visualized text will be affected—not the number of the words. This feature makes the tag cloud visualization even more attractive. The word tree facilitates for users to explore the correlations beginning from a set of visible words and proceeding with other words which appear as a result of previous choices. This approach allows the exploration of unbounded sets of words. However, increasing the number of tree nodes makes it more complicated to browse through the tree and to preserve the mental map [27]. Because there is no upper limit for the number of wines tasted produced in the visualized countries, the world map is not restricted by the amount of visualized data either.

That said, our implementation currently imposes some restrictions on the functionality of the tool. The scatter plot together with the filters, the tag cloud visualization and the bar chart diagrams perform well for a number of up to 3,000 tasting notes. For more elements, the tool becomes slower and thus less interactive. One way of improving the application's performance is to migrate the database to another, more efficient DBMS. The current DBMS and the database schema are in fact the main bottleneck of our implementation. Because of the inappropriate design of the original wine database; the word tree visualization cannot be efficiently built for more than 60 tasting notes. In order to overcome this restriction, the design of the database should be modified in such a way that it represents the tree structure of the words in the tasting notes. The response time of the world map view for standard user interactions takes about six seconds, which is a relatively long time. Here, we have to find out whether the *GeoTools* Toolkit API may provide a solution to this performance problem or if we have to move to another library.

5.6 Results

Our tool offers possibilities for linguists to explore, analyze, and present large and complex data sets for investigations of the structure of texts and discourses and of the lexical resources that languages have for the expression of different meaning domains. Not only can visual images communicate concrete information, but they can also represent abstract information in the form of visual imagery, which is of particular significance in the case of wine descriptions of subjective sensory perceptions, which by nature are transient and volatile. Through these techniques, textual data can be visually represented and interactively explored at a glance at the same time, which is clearly innovative in linguistic research. The most essential part of wine descriptions is concerned with the description of transient sensory perceptions. They are captured by our visualization approaches in the form of scatter plots, tag clouds, word trees, and bar chart diagrams. Given the availability of tagged corpora, dynamic visualization techniques open up for linguistic advances through typological comparisons across different text types, different times, and different languages. For instance, with the aid of the various filters of metadata, we can explore linguistic patterns across subsets of tasting notes, subsets of ratings of wines, or subsets of grapes. And we can apply filters, such as ”only tasting notes containing more than 400 words”, ”only sweet wines” or ”only wines from Spain” in various different combinations. The tool provides direct feedback in the form of interactive visualizations and is immediately able to answer questions such as: Do tasting notes have the same format across time? Or how do wines pattern that are described with the attribute ’sweaty saddle’? The tool thus offers the possibility for linguists to explore different variables and get an immediate response to queries about, for instance, distributional differences across expressions in the texts or metadata in relation to the changes analysts make using the different filter settings. In addition to all the above functions, statistical information related to choices that we make in the form of bar chart diagrams and tag clouds can be obtained. This functionality is particularly important for various types of analytic improvements, more accurate parameter set-tings and for subsequent formulations of new hypothesis for corpus investigations of text.

5.6.1 Use-case

A possible use-case scenario that shows the functionality of our tool could be a case in which a linguist analyzes the words and phrases in the descriptions of odors of dry rosé wines with relatively high ratings. The focus of the user's interest could be descriptors that are most frequently used in a positive sense to represent olfactory descriptions of dry rosé wines. This process of analyses requires a course of successive actions illustrated in Figure 5.12. First, the linguist sets the desired values of the filters in order to select the definite set of tasting notes to be analyzed (e.g., wine ratings between 70 and 100 and wine dryness set to dry, see Figure 5.12(a)). Then, the user selects an appropriate tasting note to start his/her study from. A good starting point could be a long tasting note, since it contains more details and therefore more descriptors (Figure 5.12(b)). Referring to the tag cloud visualization of the tasting note and considering word frequencies and their lexical meaning in the specific context (Figure 5.12(c)), the user selects word tags corresponding to words that are related to olfactory perceptions (Figure 5.12(d)). In order to further analyze the words and phrases in the whole set of tasting notes used to describe dry, rosé wines, the linguist builds a word tree as shown in Figure 5.12(e). He/she then selects the path in the tree, observing the structured organization of the visualized text in relation to his/her own association of the phrases with olfactory perceptions. This process of interactive exploration of constructions in language combines the power of human cognitive processing with the speed and accuracy of computer processing which makes it very productive and effective. After constructing a relevant sequence of words, the linguist returns to the tasting notes on the scatter plot and observes the number of elements that contain the same phrase (i.e., the tasting notes highlighted in the scatter plot, see Figure 5.12(f)). This interaction between the visualizations facilitates fast perception and comparison of the numbers represented. The user can continue this process and repeat the same actions or select another tasting note in order to explore more words and phrases with comparatively high frequency of use. In this way, the linguist can identify and describe patterns of aroma descriptors that are most commonly used in positive assessments of dry, rosé wines.

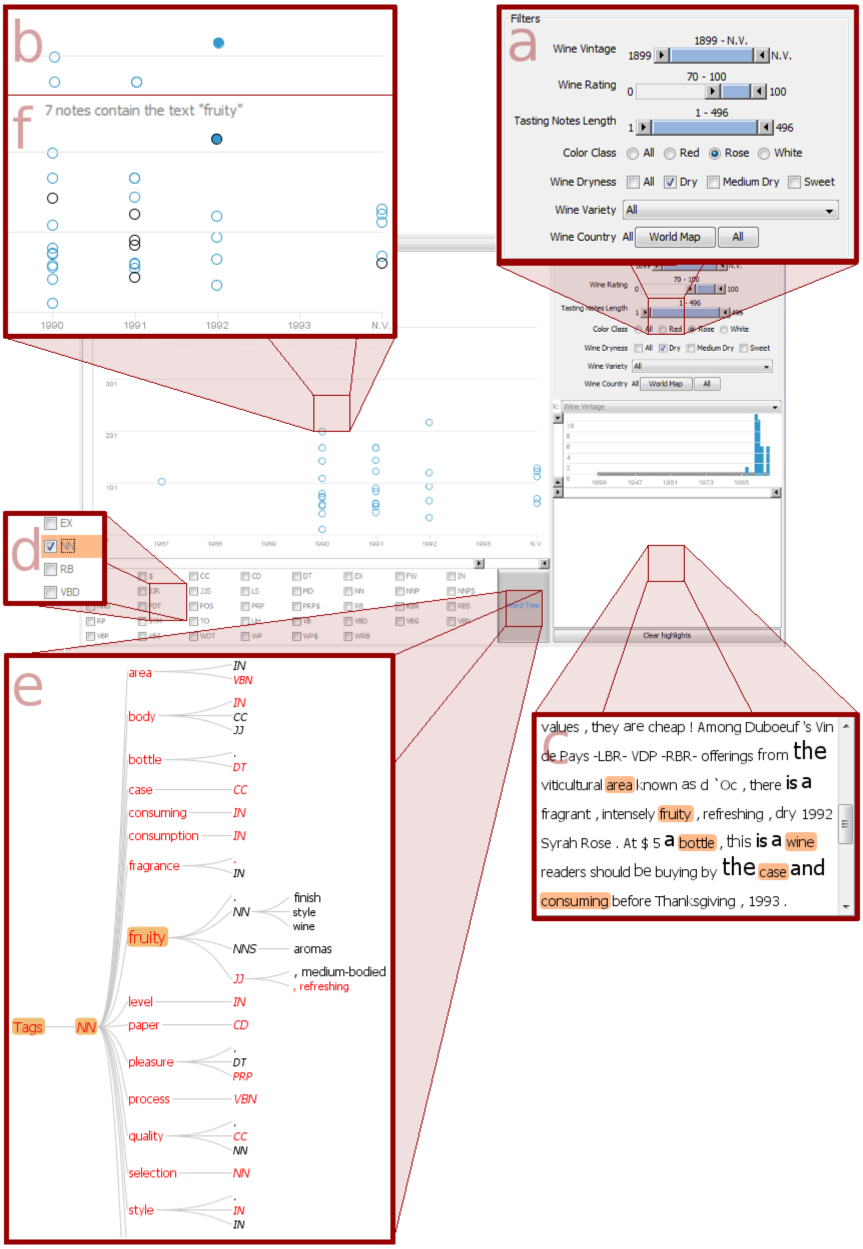


Fig. 5.12 Use-case scenario. Zoomed areas illustrate the process of interaction; the individual steps are marked with reddish letters

5.7 Conclusions and Future Work

This work is concerned with various approaches for visualizing wine tasting notes that can be used to support linguistic analyses. Our data sources are large databases containing tasting notes and metadata related to the wines tasted. Linguists are interested in the language of such texts and the possibilities offered by the language to describe sensory perceptions to better understand descriptions of them. The purpose of our tool was to visualize these data in a way that would help linguists to get a better picture of wine descriptions. All solutions presented in this paper were carefully discussed with linguists during their development.

There are several improvements that can be made to enhance the visualization tool for wine tasting notes. In Subsection 5.5.3, we discussed several problems with the current DBMS and the database schema. An improvement of this situation would be one of the first candidates for the next software revision.

Another issue would be the tag clouds that can be improved. There are function words that occur very frequently in general language like “a”, “the”, “of”, etc. They are visualized by the largest font sizes and therefore attract the attention of the users from other words that are more important and more interesting for linguistic analyses. An obvious solution to avoid this problem would be to create a user-defined black list of words that could be disregarded and excluded from the calculations. The tool could thereby avoid their overestimation. Additionally, it would be helpful to compare tasting notes with respect to their word usage. For this purpose, specific techniques, such as parallel tag clouds, could be added to the system.

The world map visualization and its performance could also be improved. It would be useful to add more interactive features to the map visualization. For example, the map could be extended by visualizing vintages, i.e., time-series data. Another idea would be to add an interactive control for tracing the wine production density in different countries on the basis of the wines’ vintages. Range sliders or other controls could be integrated to change the time period of the data visualized on the map.

Finally, we recognize that our visual analyses are also related to tasks in the field of Sentiment Analysis. It would be very interesting to develop our tool also in this direction, see for example the handbook chapter [25].

Acknowledgments

We are very grateful to Robert Parker for making his wine reviews database available to us.

References

1. Ahlberg, C., Shneiderman, B.: Visual information seeking: Tight coupling of dynamic query filters with starfield displays. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Celebrating Interdependence, CHI ’94, pp. 313–317. ACM, New York, NY, USA (1994). URL http://doi.acm.org/10.1145/191666.191775
2. AromaWheel: Web Site of the German Wine Institute (last accessed: 2012-01-23). URL http://www.deutscheweine.de
3. Caballero, R.: Cutting across the senses: Imagery in winespeak and audiovisual promotion. In: C. Forceville, E. Urios-Aparisi (eds.) Multimodal Metaphor, chap. 4, pp. 73–94. Berlin/New York: Mouton de Gruyter (2009)
4. Card, S.K., Nation, D.: Degree-of-interest trees: A component of an attention-reactive user interface. In: Proceedings of the working conference on Advanced Visual Interfaces (AVI ’02), pp. 231–245. ACM, New York, NY, USA (2002)
5. Carr, D.A., Jog, N., Prem Kumar, H., Teittinen, M., Ahlberg, C.: Using interaction object graphs to specify graphical widgets. Tech. Rep. CS-TR-3344, University of Maryland, De-partment of Computer Science, College Park, MD, USA (1994)
6. Collins, C., Viegas, F., Wattenberg, M.: Parallel tag clouds to explore and analyze faceted text corpora. In: IEEE Symposium on Visual Analytics Science and Technology (VAST ’09), pp. 91–98 (2009)
7. Dengel, A., Agne, S., Klein, B., Ebert, A., Deller, M.: Human-centered interaction with doc-uments. In: Proceedings of the 1st ACM International Workshop on Human-centered Multi-media (HCM ’06), p. 35. ACM Press (2006)
8. Eick, S.G., Steffen, J.L., Sumner Jr., E.E.: Seesoft – a tool for visualizing line oriented software statistics. IEEE Transactions on Software Engineering 18, 957–968 (1992)
9. Freeman, E., Fertig, S.: Lifestreams: Organizing your electronic life. In: AAAI Fall Sympo-sium on AI Applications in Knowledge Navigation and Retrieval, pp. 38–44. Association for the Advancement of Artificial Intelligence (1995)
10. GeoTools: The Open Source Java GIS Toolkit (last accessed: 2012-01-23). URL http://www.geotools.org/
11. Gluck, M.: Wine language: Useful idiom or idiot-speak? In: J. Aitchinson, D.M. Lewis (eds.) New Media Language, pp. 107–115. London: Routledge (2003)
12. Havre, S., Hetzler, E., Whitney, P., Nowell, L.: ThemeRiver: Visualizing thematic changes in large document collections. IEEE Transactions on Visualization and Computer Graphics 8, 9–20 (2002). DOI http://doi.ieeecomputersociety.org/10.1109/2945.981848
13. Heer, J., Card, S.K.: DOITrees revisited: Scalable, space-constrained visualization of hierar-chical data. In: Proceedings of the working conference on Advanced Visual Interfaces (AVI ’04), pp. 421–424. ACM, New York, NY, USA (2004)
14. Herdenstam, A.: Sinnesupplevelsens estetik: Vinprovaren, i gränslandet mellan konsten och vetenskapen. Stockholm dialoger (2004)
15. Hommerberg, C.: Persuasiveness in the Discourse of Wine: The Rhetoric of Robert Parker. Ph.D. thesis, Linnaeus University, Faculty of Humanities and Social Sciences, School of Language and Literature, Växjö, Sweden (2011). Linnaeus University Press
16. IBM Research: Many Eyes (last accessed: 2012-01-23). URL http://www-958.ibm.com/software/data/cognos/manyeyes/
17. Kaser, O., Lemire, D.: Tag-Cloud Drawing: Algorithms for Cloud Visualization. In: Proceedings of Tagging and Metadata for Social Information Organization (WWW ’07). Banff, Canada (2007)

18. Keim, D., Oelke, D.: Literature Fingerprinting: A New Method for Visual Literary Analysis. In: IEEE Symposium on Visual Analytics Science and Technology (VAST '07), pp. 115–122. Sacramento, CA, USA (2007)
19. Kerren, A.: Visualization of workaday data clarified by means of wine fingerprints. In: Proceedings of the INTERACT '09 Workshop on Human Aspects of Visualization, LNCS, vol. 6431, pp. 92–107. Springer (2011)
20. Kerren, A., Prangova, M., Paradis, C.: Visualization of sensory perception descriptions. In: Proceedings of the 15th International Conference on Information Visualisation (IV '11), pp. 135–144 (2011). DOI 10.1109/IV.2011.38
21. Koh, K., Lee, B., Kim, B., Seo, J.: Maniwordle: Providing flexible control over wordle. IEEE Transactions on Visualization and Computer Graphics 16, 1190–1197 (2010). URL http: //dx.doi.org/10.1109/TVCG.2010.175
22. Lagus, K., Kaski, S., Kohonen, T.: Mining massive document collections by the websom method. Inf. Sci. 163, 135–156 (2004)
23. Lee, B., Riche, N., Karlson, A., Carpendale, S.: Sparkclouds: Visualizing trends in tag clouds. IEEE Transactions on Visualization and Computer Graphics 16(6), 1182–1189 (2010)
24. Lehrer, A.: Wine and Conversation. Oxford: Oxford University Press (2010)
25. Liu, B.: Sentiment analysis and subjectivity. In: N. Indurkhya, F.J. Damerau (eds.) Handbook of Natural Language Processing, 2nd edn. Chapman and Hall/CRC (2010)
26. Michael, Silverstein: Indexical order and the dialectics of sociolinguistic life. Language and Communication 23(3-4), 193–229 (2003)
27. Misue, K., Eades, P., Lai, W., Sugiyama, K.: Layout adjustment and the mental map. Journal of Visual Languages and Computing 6, 183–210 (1995)
28. Morrot, G., Brochet, F., Dubourdieu, D.: The color of odors. Brain and Language 79(2), 309–320 (2001). DOI 10.1006/brln.2001.2493
29. Noble, A.C., Arnold, R.A., Masuda, B.M., Pecore, S.D., Schmidt, J.O., Stern, P.M.: Progress Towards a Standardized System of Wine Aroma Terminology. Am. J. Enol. Vitic. 35(2), 107–109 (1984)
30. Oesterling, P., Scheuermann, G., Teresniak, S., Heyer, G., Koch, S., Ertl, T., Weber, G.: Two-stage framework for a topology-based projection and visualization of classified document collections. In: IEEE Symposium on Visual Analytics Science and Technology (VAST '10), pp. 91 –98 (2010). DOI 10.1109/VAST.2010.5652940
31. Oracle: JAVA Swing Toolkit (last accessed: 2012-01-23). URL http://docs.oracle.com/javase/tutorial/ui/overview/intro.html
32. Oracle: The Java Database Connectivity (last accessed: 2012-01-23). URL http://www.oracle.com/technetwork/java/javase/jdbc/index.html
33. Paradis, C.: A sweet nose of earth, smoke, cassis and cherries: Descriptions of sensory perceptions in wine tasting notes. In: Proceedings of the 7th AELCO International Conference. Toledo, Spain (2010)
34. Paradis, C.: Touchdowns in Winespeak: Ontologies and Contruals in Use and Meaning-making. In: M.G. Rambaud, A.P. Luelmo (eds.) Proceedings for the 1st Congress on Linguistic Approaches to Food and Wine Descriptions, pp. 57–72. Madrid: UNED University Press, New York, NY, USA (2010)
35. Prangova, M.: Visualization of sensory perception descriptions. Master's thesis, Linnaeus University, School of Computer Science, Physics and Mathematics, Växjö, Sweden (2010)
36. Roberts, J.C.: Exploratory visualization with multiple linked views. In: A. MacEachren, M.J. Kraak, J. Dykes (eds.) Exploring Geovisualization. Amsterdam: Elseviers (2004). URL http://www.cs.kent.ac.uk/pubs/2004/1822
37. Rouby, C., Schaal, B., Dubois, D., Gervais, R., Holley, A.: Olfaction, Taste and Cognition. Cambridge: Cambridge University Press (2002)

38. Salton, G., Singhal, A.: Automatic text theme generation and the analysis of text structure. Tech. Rep. TR94-1438, Cornell University, Ithaca, NY, USA (1994)
39. Santorini, B.: Part-of-speech tagging guidelines for the penn treebank project. Tech. Rep. MS-CIS-90-47, University of Pennsylvania, Department of Computer and Information Science, Philadelphia, PA, USA (1990). (3rd Revision)
40. Shneiderman, B.: Tree Visualization with Treemaps: A 2-d Space-filling Approach. ACM Transactions on Graphics 11, 92–99 (1991)
41. Shneiderman, B.: The eyes have it: A task by data type taxonomy for information visualizations. In: Proceedings of the IEEE Symposium on Visual Languages (VL '96), pp. 336–343 (1996)
42. Spence, R.: Information Visualization: Design for Interaction, 2nd edn. Prentice Hall (2007)
43. Stasko, J., Görg, C., Liu, Z.: Jigsaw: supporting investigative analysis through interactive visualization. Information Visualization 7, 118–132 (2008). URL http://portal.acm.org/citation.cfm?id=1466620.1466622
44. Strobelt, H., Oelke, D., Rohrdantz, C., Stoffel, A., Keim, D., Deussen, O.: Document cards: A top trumps visualization for documents. IEEE Transactions on Visualization and Computer Graphics 15(6), 1145–1152 (2009). DOI 10.1109/TVCG.2009.139
45. The Berkeley Institute of Design: Prefuse Visualization Toolkit (last accessed: 2012-01-23). URL http://prefuse.org/
46. Viegas, F.B., Wattenberg, M., Feinberg, J.: Participatory visualization with wordle. IEEE Transactions on Visualization and Computer Graphics 15, 1137–1144 (2009). URL http: //dx.doi.org/10.1109/TVCG.2009.171
47. Wattenberg, M., Viègas, F.B.: The word tree, an interactive visual concordance. IEEE Transactions on Visualization and Computer Graphics 14, 1221–1228 (2008). URL http: //portal.acm.org/citation.cfm?id=1477066.1477418
48. Waugh, E.: Brideshead Revisited. Back Bay Books (1999)
49. Wise, J., Thomas, J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., Crow, V.: Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In: Proceedings of the IEEE Symposium on Information Visualization (InfoVis '95), pp. 51–58 (1995)
50. Wittgenstein, L.: Remarks on Colour. Oxford: Basil Blackwell (1977)