

Learning winespeak from mind map of wine blogs

Hirokawa, Sachio

Research Institute for Information Technology, Kyushu University

Suzuki, Takahiko

Research Institute for Information Technology, Kyushu University

Yin, Chengjiu

Research Institute for Information Technology, Kyushu University

Flanagan, Brendan

Graduate School of Information Science and Electrical Engineering, Kyushu University

<https://hdl.handle.net/2324/1457492>

出版情報 : Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 8522 LNCS (PART 2), pp.383-393, 2014-01-01.
Springer Verlag

バージョン :

権利関係 :

Learning Winespeak from Mind Map of Wine Blogs

Sachio Hirokawa¹, Brendan Flanagan², Takahiko Suzuki¹, Chengjiu Yin¹

¹ Research Institute for Information Technology, Kyushu University
{hirokawa,suzuki,yin}@cc.kyushu-u.ac.jp

² Graduate School of Information Science and Electrical Engineering,
Kyushu University, B.FLANAGAN.885@s.kyushu-u.ac.jp

Abstract. When faced with complex situations, it can often be hard to put into words and accurately express it appropriately. This becomes increasingly difficult when specialist expressions are required that are not used in everyday language. The problem is faced when trying to express in words to another person the wine that you just drank, or a wine that you want to drink to a waiter at a restaurant or shop assistant. It requires the expression in words of numerous senses including complex flavors, smells, colors, and personal emotion that is felt. These expressions are often subjective, with different people having using different expressions for the same wine. In this paper, we propose the use of wine related expressions collected from the internet and clustered to generate mind maps.

1 Introduction

It is the foundation of communication to express one's idea as language and to understand others' ideas from their message. Many years are required to master language. The course in which a child masters a native language is not necessarily the same as the course in which an adult masters a foreign language. Anyone who has lived in a foreign country for certain period would have experienced difficulties in expressing himself in the local language. At the same time, he would have experienced that he was able to make himself understood only by putting known words in order. The process of communication can be thought of as the transformation from those keywords in our brain into an actual sentence. The other side of the communication, i.e., those who listen to you talk, need to reproduce the feeling of the person who talked.

It would be natural to assume that we have something which represents words and relation of words in our brain. This idea is close to the notion of "interlanguage" by Selinker[16] who analysed how people learn and use foreign language. The present paper realises this "interlanguage" as a mind map to describe the relation of words.

There are mutually related words in a our mind. We express our thoughts using the words that we know and by constructing sentences that follow the grammatical pattern in our native language or in a foreign language we are

learning. It is necessary to build an association chart or graph of such words in our head when we learn a language. The noun representing an object can be learned by seeing the concrete object. However, experience is required to master a linguistic expression that describes feelings or emotion. We need to learn the situation and context where such a feeling takes place. We need to learn with what kind of words are used.

Expressing one's own impression is a hard task even for expert. Emotional phrases would be the most difficult subject to learn for foreign language learner. Wine would be one of the best and most difficult material to learn sensory expressions. Indeed, wine experts and wine lovers write their tasting notes with cryptic comments and mysterious acronyms. Those phrases are called as wine-speak. They are wonderful objects to tease [15]. Quandt [12] complains and enjoys exaggerated and too personal expressions to describe wines. He listed 123 wine vocabularies then invented an algorithm to choose 10 or 15 words randomly. He applied his software to "Chateau La Merde, 1995", "Chateau Grand-Cul Coteneuve, 1998" and "Chateau L'Ordure Pomerol, 2004" to create tasting notes. Then he claimed that he was not be able to find the difference between the real wine tasting notes and the automatically generated wine tasting notes.

Writing a tasting note is not easy task. This paper proposes a system which helps the acquisition of expression showing feeling using wine blogs. Given a keyword, the system retrieves the documents that contain the word, extracts characteristic words of the search result and then constructs a mind map. The input keyword is shown as the root of the map. Typical sentences are shown by specifying a keyword in the map. This visualization and interaction helps users to learn sensory expression of wine tasting.

2 Related Work

2.1 Analysis of Wine Tasting Notes

There are many magazines and Web pages which cover wine tasting notes. Analysing those expressions is one of the research fields of experts. Caballero [2] analysed manner-of-motion verbs which are frequently used in wine tasting notes such as "earthy flavors run though this firm-textured red". She collected 6,000 tasting notes from the Wine Enthusiast, Wine Spectator and Wine Advocate. The tasting notes are very short text of 30 to 100 words, and selected 50 verbs and analysed the occurrences of verbs from the view points of intensity and persistence. She classified them with '+/- force', and collected 56 typical sentences that contain such verbs.

Paradis [9] carefully analysed how prime drinking time is expressed in 200 texts from 80,000 wine tasting notes of the Wine Advocate. 38 sentences are explained in detail. Paradis and Eeg-Olofsson [10] focused on the words representing sensory experiences of vision, smell, taste and touch in analysing 84,864 wine reviews of [6]. They examined the basic statistics such as, the frequency of the words and the combination of adjective seed and nouns, and gave 39 typical phrases of those sensory expressions.

2.2 Visual and Text Mining

Visualization for text analysis has been attracting many researchers. Hearst [5] spent one chapter on information visualization for text analysis. Alencar et. al. [1] classified visualization techniques based on input (single document or collection of documents) and purpose (content, relation, temporal analysis or interaction) of the analysis. The method we use in the present paper is classified in relation extraction from texts. They explained two techniques, i.e., WordNet and Phrase net, as relation analysis of keywords. In WordNet visualization, the set of keywords displayed varies according to the content. But the structure of the net is fixed and determined by the hierarchy of WordNet. The two words linked in a phrase net represents an adjacent occurrence of them in a sentence. The phrase net is useful to extract typical sample sentences. However, it does not give an overall structure of the documents obtained as search results. The analysis of adjacent relation requires dependency parsing and is not practically applicable to the large number of documents. The relation of words in a mind map, which we use in the present paper, represents a co-occurrence of the two words. This process is implemented by a standard inverted index of target documents, which is the same as ordinary search engines.

Kerren et. al. [6] constructed a visualization system for 84,864 tasting notes of the Wine Advocate. They implemented visualization for statistic analysis and used "word tree" to represent the relation of words in a sentence. There are two kinds of nodes – tags and words. The tags are the parts of speech such as JJ(adjectives) and NN(nouns). An edge from one word to another word represents an adjacent co-occurrence of the two words. In other words, it represents a context of a word. The word tree is useful to analyse each sentence one by one. However, the overall structure of sentences cannot be grasped from one figure. On the other hand, the mind map proposed in the present paper visualizes the whole documents obtained as the search result of a query.

2.3 Concept Map and Mind Map

Concept maps and mind maps are widely used for representing contents of documents and users' knowledge. Each node of a map represents a word and an edge represents relationship between two words.

Perez-Marin et.al. [11] proposes the use of concept maps to automatically generate students conceptual models from their free texts. The difference of concept map and mind map is in the meaning of edges. An edge of a concept map has meaning among a fixed set of semantic. On the other hand, an edge of a mind map is determined by users' intuition or by some algorithmic evaluation of the relation of the two words.

Lau et.al. [7] used concept maps automatically generated from e-learning environment messages to help teachers quickly understand their students learning progress and provide appropriate guided responses. They represent a word by the context vector and used the similarity of the words to construct the map. Zouaq and Nkambou [17] created ontology from 10-36 documents and evaluated

them. The ontology they created is static and does not correspond to the context of the query. Egusa et. al. [3, 8, 13] visualized student knowledge as a mind map. Subjects are asked to draw their knowledge before and after exploratory search of particular topics. The mind maps are used to compare the effect of learning process.

The algorithm of map generation of the present paper is an extension of that in Flanagan et. al. [4], where they constructed mind maps automatically from user's twitter texts for helping communication between foreign language learners.

3 Mind Map Generation Algorithm

3.1 Wine Spectator Blogs

There are many magazines and Web sites which cover blogs and wine tasting notes. Bill Daley's "The Dalay Question"³ is one of such Web sites where he writes on wines and food. There are research articles that analyse the wine tasting notes of Robert Parker's "Wine Advocate"⁴. "Wine Spectator"⁵ is another famous magazine on wine which provides blogs as well as tasting notes. We chose blogs in "Wine Spectator" as an analysis corpus instead of tasting notes as other researches[2, 9, 10] did, because we thought that the vocabularies and phrase in tasting notes would be too specific compare to that in blog articles. Blog articles would be much better to learn words and phrases for ordinary people, particularly for English learners. Typical tasting notes are short sentences written in 30 to 100 words. Blogs are much longer than that. We collected 2,240 blogs from Wine Spectator. An average blog is written in 34 lines and contains 600 words. We constructed a search engine of the blogs using GETA⁶ and realized a visualization system.

3.2 Mind Map vs Co-occurrence Graph

We would like to visualize the contents of blogs that contain the query as a graph whose nodes are characteristic words and whose edges are relation of those words. Firstly, we have to determine how we choose the characteristic words from the search result. Secondly, we have to evaluate the importance of a relation of two words. Concerning to the choice of words, we adopt the Salton's measure [14] provided as a standard measure of GETA search engine. Naive choice of edges would be the co-occurring pairs of words. However, the graph obtained this way looks too complicated because of the numerous edges. Figure 1 compares the mind map obtained by our algorithm and the naive co-occurrence graph for the query "blue". Note that the red lines in the co-occurrence graph are the edges chosen in the mind map.

³ <http://www.chicagotribune.com/features/food/>

⁴ <http://www.erobertparker.com/info/wineadvocate.asp>

⁵ <http://www.winespectator.com/>

⁶ <http://geta.ex.nii.ac.jp/>

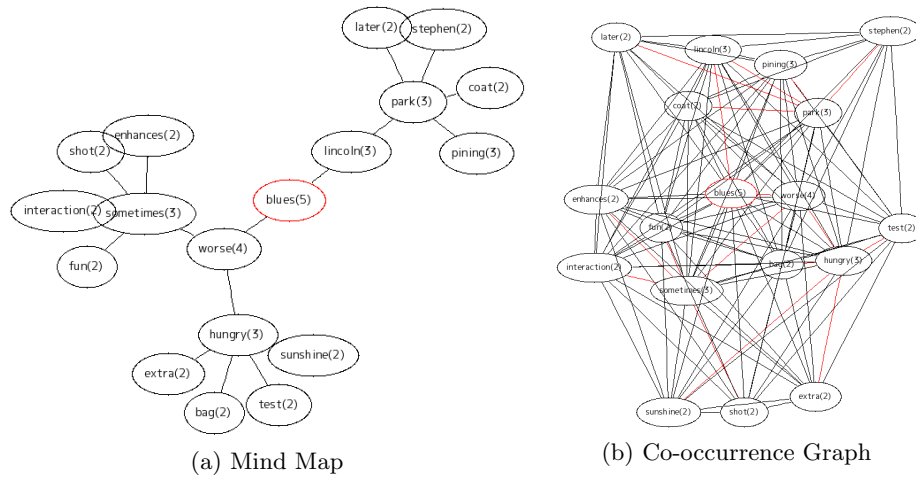


Fig. 1. Comparison of Mind Map and Co-occurrence Graph

3.3 Depth First Spanning Tree

We construct a mind map of a given query based on depth first search of related words. Each word appears only once in a map, so that the map forms a tree with the query word as the root of the tree. Given a query word, the system retrieves the documents that contain the word. The top B characteristic words of the search result are determined according to the SMART measure. The top words are linked from the focused word unless the word does not appear in the tree already constructed. The expansion proceeds until the given depth D is reached. Once a word is linked from a node, it never occurs as another node. Figure 2 shows the psuedo code of the algorithm.

```

input : w;
nextnodes = {w};
while(nextnodes is not empty){
  n = get-first-node(nextnodes);
  child = get-child(n);
  foreach c (child) {
    addedge(n,c);
    addnode(nextnode,c) unless appeared(c);
  }
  deletenode(nextnode,n);
}

```

Fig. 2. Mind Map Generation Algorithm

4 Interactive System of Mind Map

Figure 3 displays the screen shot for a query “blue”. Both of the breadth and the depth of tree expansion are set to 3. The input query is drawn in the red circle. Characteristic words are displayed as a tree as well as an indented list. A click on a word will generate another mind map of the word. The number following each word denotes the number of documents that contain the word. For example, the word “garrigue(4)” implies that there are four documents that contain all the keywords “blue”, “deliver”, “wooden” and “garrigue”. The four documents are shown if we click the number attached to the word. In this case, the four documents are shown in the right frame. Note that only the sentences that contain one of the characteristic words are displayed. Thus, we can grasp the whole picture of the search result with a mind map, an indented list and with actual instances of the feature words.

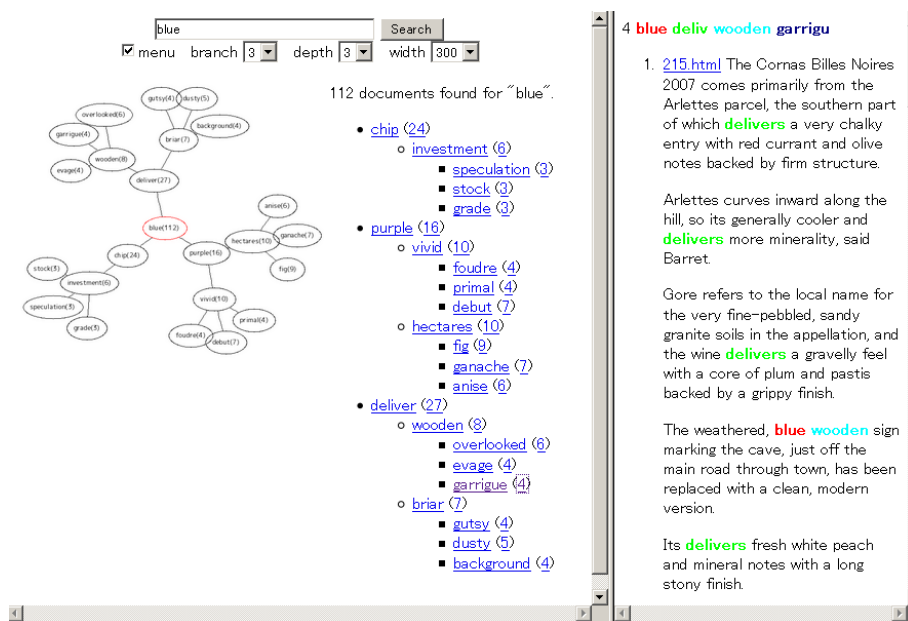


Fig. 3. Interactive System of Mind Map

5 Mind Maps of Sensory Expressions

Paradis and Eeg-Olofsson [10] analysed wine tasting notes based on sensory modalities (Table 1) and showed 39 typical sentences that contain sensory ex-

perience. In this section, we show 3 mind maps together with sample sentences that contain sensory expressions.

modality	example
VISION	purple, ruby, straw, gold, light, dark
SMELL	fruity, floral, spicy, smoky, weak
TASTE & TOUCH	flabby, soft, heavy, thin, long, crisp

Table 1. The sensory modalities [10]

5.1 Mind Map of “smoky”

Figure 4 shows the mind map for “smoky”. The names of fruits such as peach, citrus, apricot, melon and orange, looked strange at first sight. We noticed that the map consists of an upper part and a lower part. In the upper part, we see kinds of grape – Marsanne and Rousanne. The names of fruits in the upper part are peach, citrus, appricot and melon. By clicking those words, we confirmed that those fruits and the grapes co-occured in sentences such as “Pay Les Sauvages (an appellation that is rapidly improving) is made from Marsanne with 10 percent Roussanne, and it shows candied citrus peel, lemon verbena and melon notes”. Then we presumed some relationship between wines of Risling Kabinett and Auslese with orange mamalade and caramel. Indeed, it was confirmed by a sentence such as “The Kabinett 2007 (\$19) expressed this beautifully, with its racy structure defining the mineral, orange and smoke flavors”.

5.2 Mind Map of “weak”

Figure 5 is an example of the mind map for the keyword “weak”. The map displays three main branches off the root: dollar, hundred, and caressing. The related keyword “dollar” suggests an economic relation, which can be seen in the following example: “But the rapidly changing dynamics of the Australian wine industry and the weakness of the US dollar spell trouble”. Most of the nodes of this branch are related to the effect that economic change has on the global trade of wine, and in particular the impact of changes in the US dollar and consumption of wine from certain regions. On investigation we found that the child node “Tyrrell” refers to the name of a family owned Australian winery that has commented on the recent hardships faced by Australian wineries because of the prolonged drought and the narrowing disparity between the US and Australian dollar that is driving up export prices of Australian wines.

The next related keyword “hundred” could suggest the numerous varieties or quantities of wines available. This is demonstrated in the following example that is taken from a single article: “They [the wines] shine brightest in the great years, but also transcend the weak vintages ... Instead, a benchmark wine

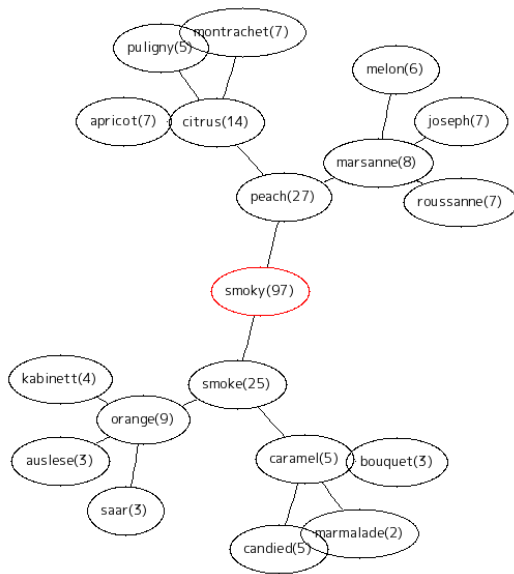


Fig. 4. Mind Map of "smoky"

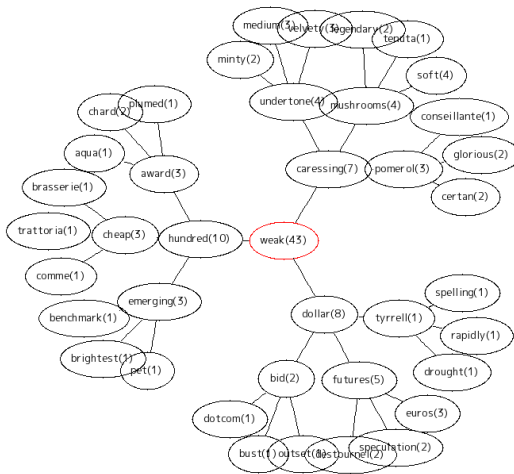


Fig. 5. Mind Map of "weak"

forme represents the emergence of an estate or terroir or even an entire region But there was just one potential classic among the several hundred that I tasted, and that was the 05 Columella from Eben Sadie”. In this example the word hundred is used to describe the numerous wines that are available, from which the writer selects a single wine as a “potential classic”. The words brightest and emergence are used to describe wines that stand out among the hundreds of wines that are available.

The next related keyword “caressing” is usually associated with sensory touch. However, as seen from the results of the mind map, when used in the context of describing a wine it can be used to express how the wine feels. This can be seen in the following example which is an excerpt from an article that also uses the child nodes mushroom and soft: “The wine in my glass was good quality with currant bush, mushroom and chocolate aromas It was round, full bodied with soft and caressing tannins (I know it was a weak vintage for California!) But it reminded me of the early days of my career when I drank a lot of bottles of Clos Du Val from 1974 and 1975”. In this context the word mushroom describes the smell, and soft and caressing describe the feel (in the mouth) of the tannins in the wine.

5.3 Mind Map of “long”

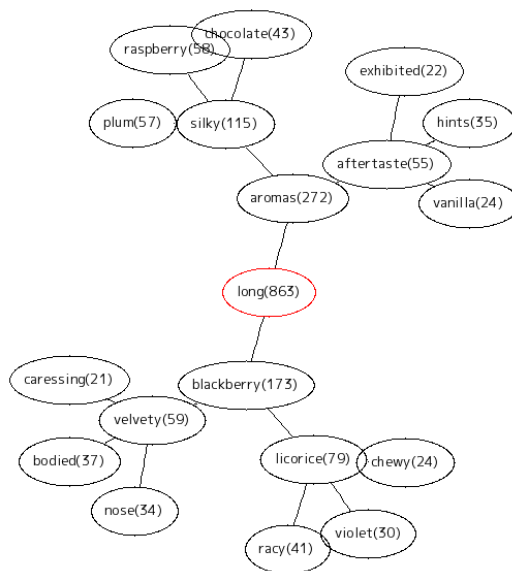


Fig. 6. Mind Map of “long”

Figure 6 is an example of the mind map for the keyword “long”. As shown in the figure, the related words are divided into 2 main groups: “aromas” and “blackberry”. The “aromas” group suggests a “taste and smell” relation, such as “after taste”, “vanilla”, “silky” and “chocolate”. These related words can be found in the following example:

- a) The Alsace Grand Cru Wintzenheim Hengst SGN, with about 240 grams RS, showed a strict nose of dried apricot and spice, followed by an intense, rich and elegant feel on the palate, with a long spicy “aftertaste” (95 points, non-blind).
- b) Moving on to Pinot Gris, the Alsace Turckheim Herrenweg had the allure of candy, offering aromas and flavors of peach, quince and “vanilla” on an opulent, yet juicy and mouthwatering frame (90 points, non-blind).
- c) It was drinking beautifully, with wonderful aromas of cherry, raspberry and milk “chocolate”, and a palate that was full yet refined, “silky”-textured and long and refreshing.

The “blackberry” group suggests a “feeling” relation, such as “velvety”, “caressing”, “nose”, “bodied” and “chewy”. These related words can be found in the following example:

- a) Full-bodied, with “velvety” and rich tannins and loads of ripe fruit. Sweet tobacco, with fresh cep and hints of blueberries and blackberries. Raspberries and blackberries.
- b) Hints of blackberry jam and fresh porcini. “Full-bodied”, with super soft tannins and a round and “caressing” finish. Meaty and long.
- c) The Roman-Conti was simply gorgeous from one barrel, exhibiting a “nose of rose” and violet, red fruits and spices.
- d) He also introduced a new Shiraz to Penfolds’ already long list of divergent styles when, in 1998, he rolled out RWT. The oldest wine in the tasting, Plexus 2003, showed rustic qualities, and was “chewy” in comparison to the other wines.

6 Conclusion and Further Work

Learning language, particularly learning foreign language, requires large amounts of experience and exposure to the real situation where words are used vividly. Sensory expression and emotional expression would be most difficult to master. The present paper proposed a visualization system of sensory words found in Wine blogs. Characteristic words are extracted and drawn as a mind map from the search result of the user’s query word. Sample sentences are shown only by clicking a word that appears in the map. Examples of mind maps and sentences are analysed as case studies. It is confirmed that meaningful groups of words and typical phrase were found quite easily.

The mind map is much easier to understand compared with the co-occurrence graph of words. Comparison and qualitative evaluation of the proposed method

will be a further work. Empirical analysis with wine vocabulary is necessary. Focusing on particular manner-of-motion verbs or positive/negative adjectives are an interesting application of the proposed method.

Acknowledgement

This work was partially supported by JSPS KAKENHI Grant Number 24500176.

References

1. Alencar, A.B., De Oliveira, M.C.F., Paulovich, F.V, Seeing beyond reading: A survey on visual text analytics, *Wiley Interdisciplinary Reviews, Data Mining and Knowledge Discovery* 2(6), pp.476-492, 2012
2. Caballero, R., Manner-of-motion verbs in wine description, *Journal of Pragmatics* 39(12), pp.2095-2114,2007
3. Egusa, Y., Saito, H., Takaku,M., Terai, H., Miwa M., Kando, N., Using a concept map to evaluate exploratory search, *Proceedings of 3rd Symposium on Information interaction in context*, pp.175-184, 2010
4. Flanagan, B., Yin, C., Inokuchi,Y., Hirokawa, S., Supporting Foreign Language Learning Using Mind-Maps, *Journal of Information and Systems in Education* 12(1), pp.13-18, 2013
5. Hearst, M.A., *Search User Interfaces*, Cambridge University Press, 2009
6. Kerren, A. , Prangova, M., Paradis, C, Visualization of sensory perception descriptions, *Proceedings of the International Conference on Information Visualization*, pp. 135-144, 2011
7. Lau, R. Y., Song, D., Li, Y., Cheung, T. C., Hao,J. X., Toward a fuzzy domain ontology extraction method for adaptive e-learning, *IEEE Transactions on Knowledge and Data Engineering* 21(6), pp.800-813, 2009
8. Miwa, M., Egusa, Y., Saito, H., Takaku, M., Terai, H., Kando, N., A method to capture information encountering embedded in exploratory Web searches, *Information Research*, 16(3), 2011
9. Paradis, C., This beauty should drink well for 10-12 years A note on recommendations as semantic middles, *Text and Talk* 29 (1) , pp.53-73, 2009
10. Paradis, C., Eeg-Olofsson, M., Describing Sensory Experience: The Genre of Wine Reviews, *Metaphor and Symbol*, 28(1), pp.22-40, 2013
11. Perez-Marin, D., Alfonsoeca, E., Rodri'guez, P., Pascual-Nieto, I., Automatic Generation of Students' Conceptual Models from Answers in Plain Text, *User Modeling*, pp.329-333, 2007
12. Quandt, R.E., On Wine Bullshit: Some New Software?, *Journal of Wine Economics* 2(2), pp.127-135, 2007
13. Saito,H., Egusa,Y., Takaku,M., Miwa, M., Kando, N., Using Concept Map to Evaluate Learning by Searching, *Proceedings of the 34th Annual Meeting of the Cognitive Science Society*, 2012
14. Salton, G., McGill, M.J., *Introduction to Modern Information Retrieval*, McGraw-Hill, 1983
15. Searle, R., *The Illustrated Winespeak: Ronald Searle's Wicked World of Winetasting*, Souvenir Pr Ltd, 1983

16. Selinker, L., Interlanguage, IRAL-International Review of Applied Linguistics in Language Teaching, Vol.10, No.1-4, pp.209-232, 1972
17. Zouaq, A., Nkambou, R., Evaluating the generation of domain ontologies in the knowledge puzzle project, IEEE Transactions on Knowledge and Data Engineering 21(11), pp.1559-1572, 2009