Modelling Climate Change Effects on Wine Quality Based on Expert Opinions Expressed in Free-Text Format: The WEBSOM Approach

Subana Shanmuganathan and Philip Sallis

Centre for Geoinformatics Research, School of Computing and Mathematical Sciences
Auckland University of Technology, New Zealand
subana.shanmuganathan@aut.ac.nz

Abstract. The motivation for modelling the effects of climate change on viticulture and wine quality using both quantitative and qualitative data within an integrated analytical framework is described. The constraints and solutions evident when taking such an approach are outlined. WEBSOM is a novel self-organising map (SOM) method for extracting relevant domain-dependent characteristics from web based texts and in this case, investigated for modelling wine quality resulting from climate variation, by web text mining published descriptions made by sommeliers about this phenomenon. This paper describes experiments using the WEBSOM method with their results.

1 Introduction

Climate change impacts on all forms of agriculture and vegetation. Current awareness of climate change and the phenomenon known as 'global warming' has increased scientific and commercial interest in it and predictions relating to it. The potential influence of climate variation on viticulture and *enology* is considered as *dramatic*. Historical viticulture records show that the ability of winemakers to produce premium quality wine is *highly prone* to climate change; both short and long term. Literature reveals that in the past viticulturists and winemakers adopted subtle changes to cultivation practices and winemaking processes to overcome short term or annual climate change effects. Similarly, major shifts on cultivation in whole wine-producing regions as well occurred to overcome the long term climate effects. These can be seen as occurring over centuries to turn the climate variation effects to producing finer wine by producing grapes with a higher percentage of sugar but without comprising the other aroma, flavour and colour protein compounds in the berry ripening process. Hence, these characteristics of wine quality are considered to be the principal factors relating to climate for viticulture regions throughout the world. Irrigation, frost, wind and solar influences are also major factors in grape production of course and therefore, are determinants of wine quality [1], [2]. The main objective of the overarching research project (called *Eno-Humanas*, see www.geo-informatics.org, also [3] and [4]) is to build models based on correlations of dependent variables in the combinatorial set of relationships that comes from data collected relating to climate, atmosphere,

M. Köppen et al. (Eds.): ICONIP 2008, Part I, LNCS 5506, pp. 917–925, 2009.

soil, terrain, moisture and plant response in association with sensory perception data relating to flavour, odour and fruit robustness. In essence, *Eno-Humanas* is about analysing precise ecological data and the less precise qualitative opinion data that comes from wine consumers, within an integrated framework.

This paper is about the second aspect of the data analysis and it analyses free text descriptions of wine quality coming from experts; Master Wine Sommeliers as explained in section 2. Section 3 outlines the results of Kohonen self-organising map (SOM)[1] based text mining approach to establishing implicit annual variations within expert opinions on 95 New Zeeland wines, expressed in free-text extracted from a web magazine called *Web Enthusiast* [5]. Section 4 elaborates upon a novel WEB-SOM approach [6] being investigated for this purpose with the same sample data set investigated with SOM. The final section of this paper proposes future research to model the effects of climate change in greater detail with larger data sets from more grape growing regions and to look at the climate change effects on the world's major wine regions. From this analysis we expect to be able to predict the wine style and appellations suitable for future climate change, short and long term, using climate change data from models already developed.

2 The Effects of Climate Change on Viticulture

Climate change is predicted to bring about significant modifications to all forms of agriculture and vegetation on earth [7]. However, its potential impact on Viticulture, the world's most expensive cultivated crop and *enology* as the science that underpins it, suggests the variability across the globe to be inconsistent (severe in the northern hemisphere and mild in the southern) and to have a variable effect on different grapevine varieties. Grapevine phenology, such as crop budburst, *floraison*, *veraison*, and harvest, greatly depends on weather and climate conditions in different regions, and this is a major factor in determining wine quality. For example, even a single degree centigrade change in temperature is predicted to impact on the production of the world famous Mediterranean wine appellations. Grape varieties thrive under significantly narrow/ niche climate and environmental conditions, and historical evidence as well supports this fact (see Introduction). Research findings with Australian grapevines and wines [8] suggest that a change of grapevine varieties could be a way to overcome the future climate variation effects in that country's major wine regions. This would of course, be an extremely expensive exercise [9], which is why objective scientific analysis for scenario building and prediction is of great significance at the moment. Here onwards the paper looks at finding implicit, relevant and meaningful taste quality descriptors, indicative of annual climate change effects, within sommeliers comments of 95 New Zealand wines.

[1] A SOM (self-organising map) is a single layered, feed forward artificial neural network (ANN), useful in displaying multi D data sets on low D maps while preserving the topology of the data. ANNs are collections of elements called 'neurons' with weighted connections between them. The neuronal structure, the connections, training and recall algorithms define the network architectures that mimic biological nerve/brain cell structures. ANNs became more popular since the 1960s due to their ability to resolve issues relating to conventional algorithmic computing methodologies.

3 Previous Research

This section of the paper discusses past efforts from literature and that of our own research with wine quality and sensory data analysis. It also describes the WEBSOM.

3.1 Wine Sensory Analysis

The wine quality literature relating to sensory and chemical data analysis can be broadly categorised into the following:

Wine characterisation and discrimination using chemical and sensory properties: Most of the papers looked at fall into this category. Wine of all main appellations, such as, *champagne*, *chardonnay*, and *pinot noir*, have been studied to find the best way for identifying the differences in support of the sub-appellations within that of main. A classification for a distinctive New Zealand wine *Marlborough Sauvignon Blanc* using sensory characterisation and chemical analysis of selected aroma compounds is discussed in [10]. An investigation into designating three sub-appellations of red Niagara Peninsula *Bordeaux* style based on differences in chemical and sensory analysis on forty-one commercially available wines is reported on [11]. Similarly, [12] and [13] looked at strategies a) to establishing control over *champagne* wine quality based on sensory and b) red table wine quality characterised by pleasing and complex mouth-feel sensations respectively for these regions.

Professional versus novice taster abilities: There are many studies in this area cited in the literature and another project within *Eno-Humanas* is considering this from an audio-mining perspective to elicit the degree of emphasis (passion) expressed about wine quality in recording of wine tasting by both professionals and novices.

Wine ratings, favourable climatic conditions and price fluctuations: Research on this subject described in [14] looked at climate and global wine quality factors and discussed a year-to-year comparison over a ten year period. It includes a description of wine quality factors in juxtaposition with prices and vintage ratings. Citing many earlier studies the authors of this work pointed out that the analysis of the relationships between climate variables and wine prices as based on an underlying hypothesis that beneficial climate conditions would improve the wine quality and that in the past, these had in turn led to short term price hikes. The study also reflected that the unavailability of consistent price data for multiple regions and with different styles over many years to be a shortcoming for any analysis on studying long term effects. The vintage ratings are seen as a strong determinant of the annual economic success of a wine region but not necessarily a predictor [15] [16].

Analysis of wine taster descriptions in free-text: There are not many studies of this kind cited in the literature. Of the studies reviewed, two major approaches are outlined herein. Taster comments analysis discussed in [17] investigated into the structure of the language used by different wine experts with a software product called ALCESTE. In this study, analysts grouped the word categories in different expert corpuses by calculating the chi square of co-occurrence of words and classified the categories into different classes, such as *idealistic, odour, colour, somesthesic, taste* and *hedonistic*. The study concluded that the language structure used by wine experts

as not organised along their sensory dimensions instead with wine prototypes, this is reflected in a recent study [18] as well.

In another interesting piece of work [19], researchers looked at calculating synthetic liking scores by studying the correlations between pairs of original scores and word groups/counts in free text comments and then comparing these synthetic scores with that of the real for a sample set of wines studied. The authors used multiple factor analysis to establish the correlations between wine comment-liking score pairs studied.

Research conducted at University of California Davis [20] found that only 25% of wine liking ratings to be linked to wine sensory descriptive data in a map created with statistical analysis results of the latter on *y* axis and ratings on *x* axis. The authors as well found some descriptors, such as "*leather*" and "*sour*", as having a negative effect and a few others as preferred, such as "*vanilla/oak*" and "*canned vegetable*". Furthermore, noted even though 75% of the variations in liking scores could not be explained, the results should be read with caution.

3.2 WEBSOM Applications

WEBSOM provides an efficient methodology for full-text information retrieval and exploration of large collections of documents. It uses SOMs to statistically analyse the relations between words based on their co-occurrence in documents, and then based on the relationships, creates document maps. As the word co-occurrence details are used as basic components for SOM clustering, similar documents get clustered close to each other on a map of documents. The WEBSOM approach first developed for creating maps of documents could be applied to search and exploration of documents [21] with unsupervised [22], or partially supervised in processing of newsgroups [23], browsing interface for web pages for the exploration of document collections [24] and as a method/ tool for data mining in textual databases [25].

In this research, WEBSOM is used to modelling any implicit year-to-year variations that may exist between groups of words (depending on their co-occurrences) found in sommelier comments, that best describe wine quality/ wine sub appellations in terms of appearance, aroma and mouth-feel; this is possible as wine experts tend to describe wine quality relating to prototypes of wine style and quality (see section 3.1 Wine sensory analysis).

3.3 Text Mining of Sommelier Comments Using SOM

As a first step in implementing the *Eno-Humanas* concept to model the correlations between climate change effects on grapevine phenology that in turn impact on wine quality with ecological data and sommelier/ wine taster comments, the latter is being investigated herein using sommelier comments extracted from a popular web based wine catalogue called *Wine Enthusiast* [3]. The initial results of SOM based text mining approach to analysing the comments of 95 New Zealand wine (in free text format) are discussed in this section.

After removing stop words, such as a, the, in and etc, and stemming the comments using perl scripts, a matrix of words (*lemmas*) was created from the wine taster comments (Fig. 1). Weights w_i of (1) for the selected words were calculated by applying

the well-known information retrieval system called Salton's vector space model, which is based on a) local information from individual documents and b) global information from the collection of documents.

$$w_i = tf_i * log\left(\frac{D}{df_i}\right) \quad (1)$$

Where,

tfi = term frequency (counts) or number of times a term i occurs in a document.
dfi = document frequency or number of documents containing term i
D = number of documents in the collection/database.

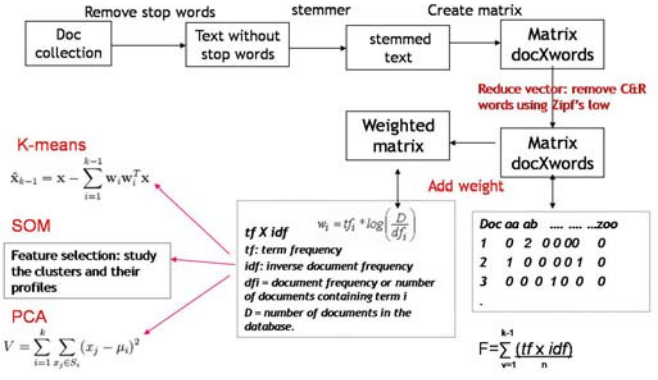


Fig. 1. Diagram on the steps used to create a word matrix of 95 New Zealand wines

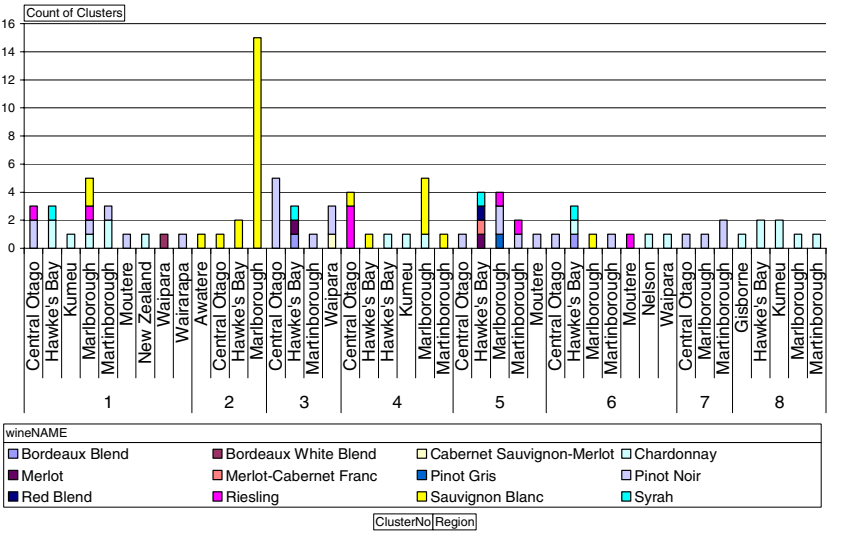


Fig. 2. Histogram showing the number of wines in different clusters (*y* axis) and regions (*x* axis) of a SOM created with 95 NZ wine word frequencies as per Fig 1 and formula (1)

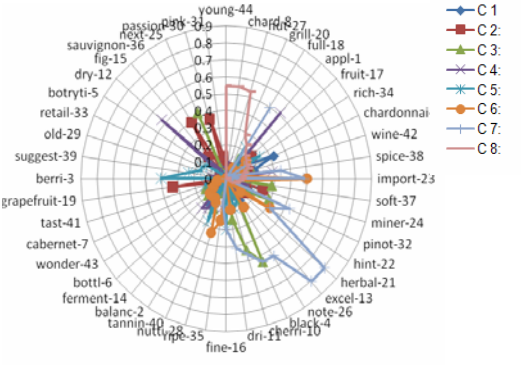


Fig. 3. SOM cluster profile radar (word average)

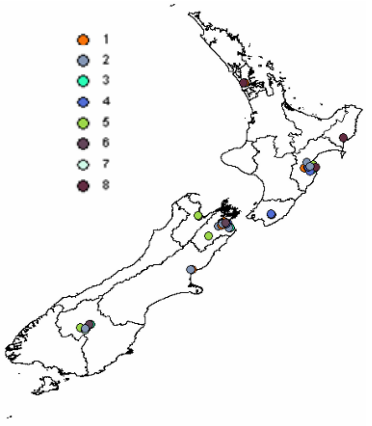
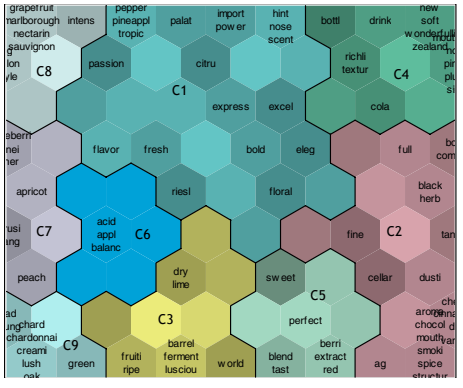


Fig. 4. 95 NZ wine clustering projected on DIVA map



C 1: pepper pineappl tropic, palat, import power, herbal hint nose scent, passion, citru, express, excel, flavor, fresh, bold, eleg, riesl, floral
C 2: full, bodi complex, black herb, fine, tannin, cellar, dusti, cherri cinnamon dri vanilla, ag, aroma chocol mouth smoki spice structur
C 3: dry lime, fruiti ripe, barrel ferment lusciou vineyard, world
C 4: bottl, drink, layer new soft wonderfulli zealand, richli textur, mouthfeel noir pinot plum silki, cola
C 5: sweet, perfect, blend tast, berri extract red rich
C 6: acid appl balanc
C 7: gooseberri honei miner, apricot, citrusi orang, peach **C 8**: blanc grapefruit marlborough nectarin sauvignon, intens, fig melon style
C 9: load young, chard chardonnai creami lush oak smoke toast, green

Fig. 5. SOM of word frequencies. SOM clusters C1-C9 show the grouping of words based on their co-occurrences in 95 New Zealand sommelier comments being analysed.

4 The WEBSOM Approach to Text Mining Sommelier Comments

The WEBSOM approach to wine expert comments analysis aims to establish any implicit annual variations within the comments and uses the same matrix of words used in the SOM analysis (fig 1). The WEBSOM results are discussed in this section.

no	Wno	C 1:	C 2:	C 3:	C 4:	C 5:	C 6:	C 7:	C 8:	C 9:	year	wtype	Regior	Cluste	price	year	rate	Comments
75	w75	0.19	0	0	0	0	0	0.31	0.13	0	2004	Sauvic	Awate	1	20	2004	91	bell pepper and
57	w57	0.04	0.05	0	0.02	0	0	0	0	0.04	2002	Pinot_	Centra	1	0	2002	91	the region this is the
60	w60	0.03	0.07	0	0.11	0	0	0	0	0	2002	Pinot_	Centra	2	43	2002	91	the bottle in a slightly
58	w58	0.03	0.1	0	0	0.09	0	0	0	0.06	2002	Pinot_	Centra	4	0	2002	91	this is a
62	w62	0.14	0.29	0	0.01	0.18	0	0	0	0	2003	Pinot_	Centra	5	62	2003	91	and no it-s not
61	w61	0.05	0.23	0	0.06	0	0	0	0	0	2004	Pinot_	Centra	2	60	2004	91	bacony edge to broad
63	w63	0.08	0.09	0.04	0.23	0	0	0	0	0.04	2005	Pinot_	Centra	2	63	2005	91	of reduction on the
25	w25	0.02	0.36	0	0.33	0.07	0.21	0	0	0	2005	Pinot_	Centra	4	35	2005	92	award winner back in
59	w59	0.02	0.22	0	0.46	0	0.2	0	0	0	2005	Pinot_	Centra	1	40	2005	91	wonderfully silky
26	w26	0.02	0.11	0	0.18	0	0	0	0	0.04	2006	Pinot_	Centra	4	42	2006	92	change from
64	w64	0	0.24	0	0.36	0	0	0	0	0	2006	Pinot_	Centra	5	63	2006	91	structured this is the
71	w71	0.07	0	0.31	0	0.06	0	0.03	0.06	0.01	2001	Rieslir	Centr	2	21	2001	91	peaches and
32	w32	0.42	0.07	0.12	0.01	0.12	0	0.13	0.06	0	2002	Rieslir	Centr	5	23	2002	92	expression of Riesling
72	w72	0.14	0.11	0.11	0	0	0.69	0.02	0	0.1	2002	Rieslir	Centr	2	23	2002	91	Riesling is loaded
73	w73	0.04	0.08	0	0.02	0.05	0.37	0.19	0	0	2006	Rieslir	Centr	1	27	2006	91	the sugars are amply
11	w11	0.09	0.02	0.28	0	0.19	0.49	0.07	0	0.23	2001	Sauvic	Centra	1	16	2001	93	s no oak aging or
76	w76	0.12	0	0	0.06	0	0	0	0.3	0	2004	Sauvic	Centra	1	20	2004	91	requisite pea pepper

Fig. 6. Table showing the different descriptors & their frequencies in wine taster comments that reflect regional, wine style and annual variations in which *Pinot noir* wines from Central Otago, do not consist of C7-8 words. Two of 2005 wines, consists of C6 words in addition to C1, C2, C4 & C5 descriptors.

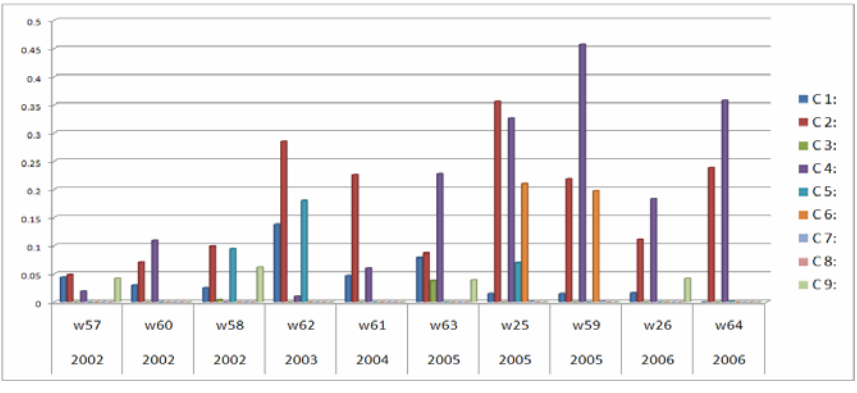


Fig. 7. Histogram showing the word frequencies of Central Otago's *Pinot noir* wines

4.1 WEBSOM Results

The different groups of wine words/descriptors from the WEBSOM approach that is based on the co-occurrences of words in the expert comments of 95 New Zealand wines, are given in fig. 5. Of these groups, C1 words/lemmas (*pepper pineappl tropic, palat, import power, herbal hint nose scent, passion, citru, express, excel, flavor, fresh, bold, eleg, riesl, floral*) are used in most of the wine comments hence, this cluster could be further divided or some discretion is needed in the selection of words to make the wine descriptor grouping more meaningful.

Based on the word frequencies of *Pinot noir* wines from Central Otago (fig. 6 & 7), 2005/6 wines have SOM cluster C4 descriptors (*bottl, drink, layer new soft wonderfulli zealand, richli textur, mouthfeel noir pinot plum silki, cola*) at higher frequencies. C 2 descriptors (*full, bodi complex, black herb, fine, tannin, cellar, dusti, cherri cinnamon dri vanilla, ag, aroma chocol mouth smoki spice structur*) as well show some variations.

5 Conclusion and Future Work

The initial results of the WEBSOM approach to discerning the descriptors that have the potential to transform the free-text wine comments into quantitative data on wine quality appear convincing. These could be used to analyse the correlations between wine quality and climate change effects. More experimental work is needed to select appropriate wine quality descriptors and confirm the correlations between these descriptors and other wine factors, such as short term climate change effects, wine liking scores/ ratings and price. We contend that continued work of a scientific nature such as reflected in this paper, will provide useful insights into the dynamics of the interaction between environment and plant growth such that better informed decisions can be made by the growers, especially at this time when global climate change is providing much angst among both wine producers and consumers alike.

References

1. Jones, G.V., White, M.A., Cooper, O.R., Storchmann, K.-H.: Climate and Wine: Quality Issues in a Warmer World. In: Proceedings of the Vineyard Data Quantification Society's 10th Eonometrics Meeting, Dijon, France (2004)
2. Hansen, A., Dale, V.: Biodiversity in US Forests under Global Climate Change. Ecosystems 4, 161–163 (2001)
3. Sallis, P.J., Shanmuganathan, S., Pavesi, L., Jarur, M.: A system architecture for collaborative environmental modelling research. In: Samari, W.W., McQuay, W. (eds.) The 2008 International Symposium on Collaborative Technologies and Systems (CTS 2008), Irvine, California, pp. 39–47. IEEE, New Jersey (2008)
4. Shanmuganathan, S., Sallis, P.J., Pavesi, L., Jarur, M.: Computational intelligence and geo-informatics in viticulture. In: Al-Dabass, D., Turner, S., Tan, G., Abraham, A. (eds.) Proceedings of the Second Asia International Conference on Modelling & Simulation, Kuala Lumpur, Malaysia, pp. 480–485. IEEE computer society, Los Alamitos (2008) (CD version)
5. Wine enthusiast magazine, `http://www.winemag.com/buyingguide/search.asp?db=`
6. Lagus, K., Honkela, T., Kaski, S., Kohonen, T.: WEBSOM - A Status Report. In: Alander, J., Honkela, T., Jakobsson, M. (eds.) Proceedings of STeP 1996, pp. 73–78. Publications of the Finnish Artificial Intelligence Society (1996)
7. Atkins, T.A., Morgan, E.R.: Modelling the effects of possible climate change scenarios on the phenology of New Zealand fruit crops ISHS Acta Horticulturae 276. In: II International Symposium on Computer Modelling in Fruit Research and Orchard Management,
8. Webb, L.B.: The impact of projected greenhouse gas-induced climate change on the Australian wine industry, PhD thesis, School of Agriculture and Food Systems, University of Melbourne, p. 277 (2006)
9. Gutierrez, A.P., Luigi, P., Ellis, C.K., d'Oultremont, T.: Analysis of climate effects on agricultural systems. Report published by California Climate Change Center CEC-500-2005-188-SD, pp. 28 + appendices A1-7 (2005)
10. Parr, W.V., Green, J.A., White, K.G., Sherlock, R.R.: The distinctive flavour of New Zealand Sauvignon blanc: Sensory characterisation by wine professionals. Food Quality and Preference 18, 849–861 (2007)

11. Kontkanen, D., Reynolds, A.G., Cliff, M.A., King, M.: Canadian terroir: sensory characterization of Bordeaux-style red wine varieties in the Niagara Peninsula. Food Research International 38, 417–425 (2005)
12. Vannier, A., Bruna, O.X., Feinberg, M.H.: Food Quality and Preference, vol. 10, pp. 101–107 (1999)
13. Gawel, R., Iland, P.G., Francis, I.L.: Characterizing the astringency of red wine: a case study. Food Quality and Preference 12, 83–94 (2001)
14. Jones, G.V., White, M.A., Cooper, O.R., Storchmann, K.: Climate and Global Wine Quality. Climatic Change by Springer 73, 319–343 (2005)
15. Nemani, R.R., White, M.A., Cayan, D.R., Jones, G.V., Running, S.W., Coughlan, J.C.: Asymmetric climatic warming improves California vintages. Clim. Res. 19, 25–34 (2001)
16. Ashenfelter, O., Jones, G.V.: The demand for expert opinion: BordeauxWine. In: VDQSAnnual Meeting, d'Ajaccio, Corsica, France, October 1998. Cahiers Scientifique from the Observatoire des Conjonctures Vinicoles Europeenes, Faculte des Sciences Economiques, Espace Richter, Ave. de La Mer, BP 9606, 34054 Montpellier Cedex 1, France (2000)
17. Brochet, F., Dubourdieu, D.: Wine Descriptive Language Supports Cognitive Specificity of Chemical Senses. Brain and Language 77, 187–196 (2001)
18. Valentin, D.: Wine language and expertise level: A cognitive point of view. In: Proceeding of Bacchus at Brock, the third international interdisciplinary wine conference, St. Catharine, June 7-9, 11 p. (2007)
19. Beće-Bertaut, M., AÍvarez-Esteban, R., Pageś, J.: Food Quality and Preference 19, 122–134 (2008)
20. Frøst, M.B., Noble, A.: Preliminary study of the effect of knowledge and sensory expertise on liking for red wines. American Journal of Enology and Viticulture 53(4), 275–284 (2002)
21. Lagus, K., Honkela, T., Kaski, S., Kohonen, T.: In: Alander, J., Honkela, T., Jakobsson, M. (eds.) Proceedings of STeP 1996, pp. 73–78. Publications of the Finnish Artificial Intelligence Society (1996)
22. Kaski, S., Honkela, T., Lagus, K., Kohonen, T.: Creating an order in digital libraries with self-organizing maps. In: Proceedings of World Congress on Neural Networks, WCNN-1996 (1996)
23. Honkela, T., Kaski, S., Lagus, K., Kohonen, T.: Newsgroup exploration with WEBSOM method and browsing interface. Technical Report A32, Helsinki University of Technology, Laboratory of Computer and Information Science, Espoo. WEBSOM home page (1996), http://websom.hut.fi/websom/
24. Lagus, K., Kaski, S., Honkela, T., Kohonen, T.: Browsing digital libraries with the aid of self-organizing maps. In: Hopgood, B. (ed.) Proc. of Fifth International World Wide Web Conference, Paris, vol. posters, pp. 71–79 (1996)
25. Lagus, K., Honkela, T., Kaski, S., Kohonen, T.: Self-organizing maps of document collections: A new approach to interactive exploration. In: Proceeding of Knowledge Discovery and Data Mining, KDD 1996 (1996)