Weinakademie Österreich
D7 Weinakademiker Thesis

Diego Berea Cabaleiro
diego.berea@ozonaconsulting.com

Artificial Intelligence-based WSET SAT© wine tasting note generator. A practical application to Sherry wines

I. Motivation for choosing this topic

Since 2016, I have been involved in research on artificial intelligence techniques related to natural language generation (NLG and Data2Text) both in my company and in collaboration with the Intelligent Systems Unit of the Research Center for Intelligent Technologies (CiTIUS) at the University of Santiago de Compostela (Spain).

Upon commencing my studies for the WSET Diploma, I became interested in analysing the objectivity/subjectivity of tasting notes based on the WSET Systematic Approach to Tasting (SAT©). As I am a member of two international ISO standardisation committees (ISO/IEC JTC1 SC40 Service Management and Governance and ISO TC37 WG 11 Plain Language) I initially sought to examine standardisation efforts undertaken by ISO (vocabulary, sensory analysis, recognition of odours, sensitivity of taste, etc.).

Subsequently, I began to question the extent to which a tasting note based on the SAT could be considered objective: both in terms of whether a tasting note objectively describes or allows for the identification of a specific wine within a set and in terms of exploring ways to generate tasting notes from parameters that can generally be considered objective and not subject to interpretation.

This study represents the initial step in a research line focused on the application of NLG to the creation of tasting notes, using multiple languages and several language styles. This approach is applied to dry types of Sherry wine (Fino, Manzanilla, Amontillado, Palo Cortado and Oloroso) and is based on a reduced number of oenological parameters:

- Ethanol
- Glycerol
- Volatile Acidity
- Acetaldehyde
- Ethyl Acetate
- Dry Extract
- Gallic Acid
- Etc.

II. Problem/Objective

The objective of this thesis is to explore the feasibility of generating tasting notes following the WSET SAT using artificial intelligence techniques based on objective quantitative information, such as the numerical values of certain standard parameters of a Sherry wine's chemical analysis.

As part of this study, the practical implementation of a tasting note generator for Sherry wines has been carried out. This can be accessed at https://wineabe.com/tasting-note.
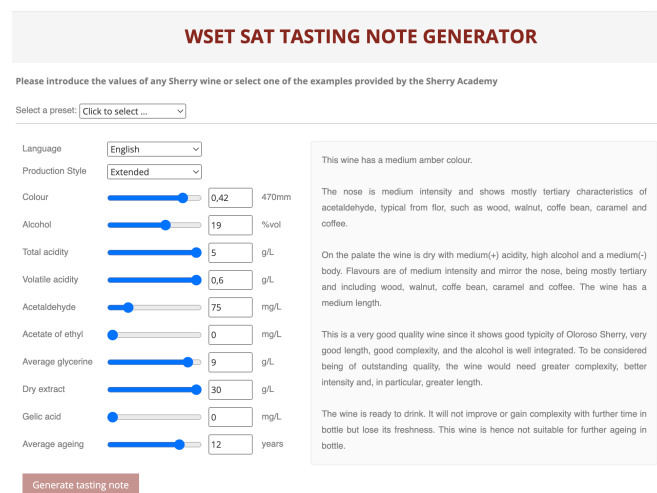

Figure 1. Prototype of the WSET SAT Tasting Note Generator

This thesis presents a model for the automatic generation of tasting notes which is compliant with the SAT and valid for three writing styles and several different languages (English, Spanish, etc.).

Different artificial intelligence techniques are employed for the acquisition and classification of knowledge, for the synthesis of the tasting note and for the evaluation and training of the model.

III. Methodology

Natural Language Processing (NLP) involves a series of artificial intelligence fields and computational linguistics, which, although they are related, use very different techniques:
- Speech Recognition
- Natural Language Understanding
- Natural Language Generation
- Speech Synthesis

Natural Language Generation (NLG) is the subfield of NLP concerned with the construction of computer systems that can produce understandable texts in English or other human languages from underlying non-linguistic representations of information.

An NLG system can be of use when it is more convenient and efficient to present information in a written way rather than presenting it via tables of data, diagrams or graphs. This is the case with tasting notes when the expected result is a written description of the phases of the sight, smell and taste of a wine and, eventually, the drawing of conclusions on its quality and ageing capacity.

One of the aims of this thesis is to evaluate to what extent it is viable, useful and efficient to be able to automatically generate tasting notes based on the oenological/chemical information of a wine sample.

In 2000, Reiter & Dale defined a sequential structure of six phases which have become standard in the design of NLG systems.

This thesis is organised into two main chapters which match these six standard phases:

- Intelligent Data Analysis
    - Content Determination
- Document synthesis
    - Discourse planning
    - Sentence aggregation
    - Lexicalisation
    - Referring expression generation
    - Linguistic realisation

The first phase is specific to the field (in this case, wine tasting) and requires specific knowledge. The second phase is common to any NLG system and essentially requires knowledge of linguistics.

To these two macro-phases can be added a third of validation and learning. Validation can be carried out by using experts to compare real tasting notes with those generated automatically.

With the results of the validation feedback and learning, mechanisms can be defined so that subsequent tasting notes can be generated with a greater degree of accuracy.

IV. Content

The intelligent data analysis phase addresses the determination of the contents and the different artificial intelligence techniques which can be employed.
- Initial corpus of output texts
- Computable data for an SAT wine-tasting note
- Elaboration of a linguistic map relating the input data with the output texts
- Management of impossible or anomalous combinations

The document synthesis phase comprises the five standard phases common to any NLG system and discusses the possible architectures for realisation.

V. Conclusion

This thesis concludes, via the creation of a functional prototype, that the development of a computational system based on artificial intelligence for the writing of tasting notes aligned with the requirements of the SAT is feasible and can have multiple applications:
- The effortless generation of standardised tasting notes based on a small set of standard chemical parameters for inclusion on a label or website, ensuring that they are understandable for anybody familiar with the systematic tasting technique.
- As a study tool for WSET students, enabling them to compare their hand-written tasting notes with those generated by the NLG system, just as they would compare their notes with those of an expert.
- Etc.

The following aspects could be of interest as future lines of research:
- The analysis of results and the validation of the degree of accuracy of the tasting notes generated using the criteria defined in ISO 8586:2012 for the selection, training and monitoring of expert sensory assessors, particularly as far as repeatability, reproducibility and discrimination ability are concerned.
- The completion and extension of the four techniques proposed to elaborate more complete parallel data-text corpora.
- The application of fuzzy logic to the labelling of quantitative parameters to deal with subjective labels such as medium(-), medium and medium(+).
- The incorporation of more powerful supervised and unsupervised learning mechanisms to improve the tasting notes generated.
- The incorporation of additional mechanisms of variability in the language generation in order to provide more natural and less repetitive texts.
- The extension of the model to other wines similar to Sherry with biological or oxidative ageing from other regions of the world (Sardinia, Jura, California, McLaren Vale, etc.).
- The extension of the model to any type of wine.