The Building Blocks of an Artificial Intelligence Model for quantifying Similarity among Plant Variety Denominations

Ángela H. Martínez López

Contents

Summary

The present research paper serves to devise the foundations of an Artificial Intelligence ("AI") powered model for testing the similarity among plant variety denominations (hereinafter, "denominations"). The imperious need for such a tool is explained by *the* so-called "saturation" phenomenon, that is, the fact that the number of denominations in registers is progressively increasing and in consequence the likelihood of conflicts between these and novel proposed denominations increases too. An algorithm can help much here because, in a matter of seconds, it is capable of scanning through a large database and compute relatively accurate similarity rankings.[1]

The proposed similarity testing model is aimed at supporting examiners from Plant Variety Protection ("PVP") offices across the globe in the process of assessing the suitability of proposals for denominations. Further, the model can be projected outwardly in the form of "pre-apply" services, so that applicants for denominations can also make use of it for evaluating if their designation of interest is likely to be deemed eligible for registration by the PVP authority. Accordingly, the overall architecture is conceived as a "bicephalous" system, which first "head" is a back-office for use by denomination examiners in a PVP office and which second "head" is a front-office rendered accessible to any external user.

The model proposed thus represents the convergence between two fascinating domains, namely, that of computer science and that of PVP and denominations. More specifically, the conceptualisation exercise conducted here is framed within the disciplines of similarity intelligence (i.e.: the process of discovering intelligence through similarity among objects)[2] and of "jurimetrics" (i.e.: the science of applying quantitative methods, especially probability and statistics, to law)[3]. From a methodological perspective, the study has been based to a large extent on literature concerning computer-automated models for use in the domain of trade mark protection. Therefore, despite the model has been specifically envisaged for its deployment in the environment of a PVP office, some of its building blocks can also be of interest for Intellectual Property ("IP") offices competent for the registration of trade mark rights.

It is important to warn that the ambitions for developing a well-oiled machinery for testing similarities does need to be tempered with the right dose of realism. Indeed, the subjectivity inherent to the examination of proposals for denomination implies that encoding into

[1] An algorithm can be defined as "a sequence of computational steps for solving a problem". Algorithms boast an advantage of scale *vis-à-vis* humans. They have the capacities of multitasking and of computing vast amounts of data at an incomparable speed (K. A. Chagal-Feferkorn, 'How Can I Tell if My Algorithm Was Reasonable?', 2021, Vol. 27, Issue 2, *Michigan Technology Law Review*, 213).

[2] Z. Sun, 'Similarity Intelligence: Similarity Based Reasoning, Computing, and Analytics', 2023, Vol. 5, Issue 3, *Journal of Computer Science Research*, 2.

[3] K. D. Ashley, 'Modelling Legal Argument: Reasoning with Cases and Hypotheticals', 1990, *Artificial Intelligence and Legal Reasoning Series*, MIT Press (Bradford), Cambridge (Massachusetts), 128.

algorithmic formulas legal notions and common-sense nuances proves extremely complex. When it comes to the use of the model by denomination examiners, the importance of keeping them on the helm of the procedure cannot be overemphasised. It is not machines but them, as agents competent for the overall assessment of the likelihood of confusion among designations, who have the final say on the suitability of denominations. The computer-generated output can serve them as supportive aid, but they should always analyse such with a critical spirit. Moreover, examiners must receive proper training on the use of the model and be aware of the fact that automated decisions are generally bestowed of a "halo of objectivity" that can prove deceiving at times. On the other hand, the model's external users must be informed about the potential errors in the model's outputs. The use of disclaimers on this matter is, for instance, an effective mechanism to convey this message.

In the present research paper, the above-referred aspects are scrutinised in detail. In terms of organisation, the work has been structured into five sections concerning different thematic areas. The study is opened with Section 1, consisting in a general introduction to the investigation. It is followed by Section 2, containing a necessary description of the nature and function of the denomination, as well as of its regulatory framework. The ins and outs of the AI-powered similarity testing model are then described in Section 3. Next, Section 4 contains some brief considerations about the perspective and potential expectations of the two main types of users targeted with the model, namely, denomination examiners and applicants for PVP rights. Lastly, Section 5 serves as a wrap-up of the main takeaways of this investigation.

1. Introduction

In a time of drastic change, it is the learners who inherit the future. The learned usually find themselves beautifully equipped to live in a world that no longer exists.[4]

Breakthrough AI technologies are seeping through all layers of industrial sectors across the globe. The public sector is no stranger to this phenomenon and is already marching towards quantitative prediction.[5] Long-established yet outdated methodologies in the public administration should be challenged as much as they should be, and managers be ready to coordinate the necessary organisational, structural, and procedural adjustments for the integration of AI models in administrative workflows.[6] This notwithstanding, the utopia that machines will take over procedures and officials be relegated to the land of *il dolce far niente* is simply not realistic. What can be expected, though, is for machines to take up more space in administrative procedures. When glancing into the crystal ball, the medium-term future of public administration entities features strong synergies between humans and machines, working together as one to deliver services more effectively that either could alone.[7]

Against this backdrop, the present research paper is aimed at exploring, from an empirical perspective, the digital enhancement of the administrative procedure for the examination of the suitability of proposals for plant variety denomination (i.e.: the name by which each plant variety must be designated, hereinafter referred to as "denomination") in Plant Variety Protection ("PVP") offices. To this end, the foundational bases of a multifaceted computer-automated model for the testing of similarity among denominations are here sketched out.

The content of the research paper has been compartmentalised into several thematic sections. Following the present introductory section, Section 2 delves into the nature and function of the denomination, as well as into its regulatory framework. Section 3 then describes the building blocks of the AI-powered similarity testing model, that is, a model for testing the similarity of a given designation against earlier registered denominations in the state of the art. Next, Section 4 contains some observations on the perspective and expectations of the two main types of users targeted by the model, namely, denomination examiners and applicants for PVP rights. Lastly, Section 5 serves as a wrap-up of the ten main lessons to take home from this investigation.

---

[4] E. Hoffer, *Reflections on the human condition*, Harper and Row, 1973.

[5] H. Surden, 'Machine Learning and Law' in R. Vogl, *Research Handbook on Big Data Law*, Edward Elgar Publishing, 2021, 171.

[6] OECD (J. Berryhill et al), 'Hello, World: Artificial intelligence and its use in the public sector', 2019, *Working Papers on Public Governance*, p. 123.

[7] A. Moerland and C. Freitas, "Artificial intelligence and trade mark assessment" in R. Hilty et al, *Artificial Intelligence and Intellectual Property*, Oxford University Press, 2021, 266 (hereinafter, "A. Moerland, AI and trade mark assessment"), p. 289.

2. A glimpse into the captivating realm of plant variety denominations

A denomination can be defined as the generic designation of a specific plant variety, thereby enabling its identification and differentiation from other plant varieties. The denomination plays an important role in trade by enabling purchasers to identify in the marketplace the specific variety which they are seeking for on account of its technical or botanical attributes.[8] Historically, the notion of "variety denomination" was legally formalised in the Convention of the *Union Internationale pour la Protection des Obtentions Végétales* ("UPOV"). UPOV is the intergovernmental organization competent for the harmonisation of the international legal framework for PVP systems (i.e.: the *sui generis* regimes for the industrial property protection of plant varieties).[9] PVP offices grant PVP rights, which are valid and enforceable in the territory of protection. The subject matter of protection of a PVP right is the "plant variety", defined in Article 1(vi) of the UPOV Convention as "a plant grouping within a single botanical taxon of the lowest known rank" (*see* Figure 1 below).[10]

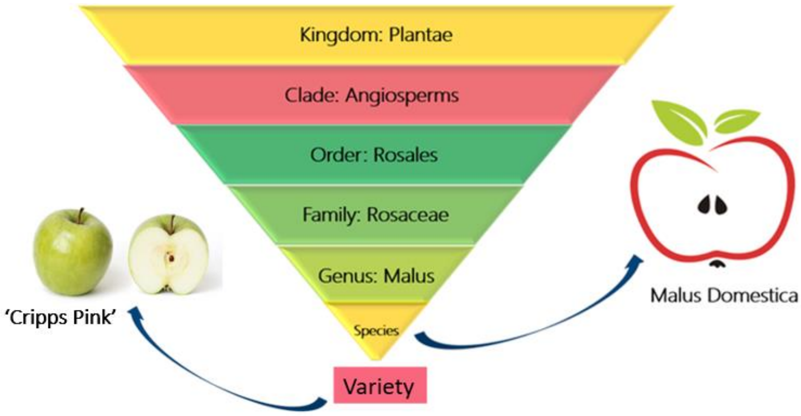

Figure 1: Taxonomic classification for apples © CPVO

For a plant variety to be eligible for PVP, it must be compliant with the technical requirements of distinctness, uniformity, and stability, as well as with the requirement of (commercial) novelty.[11] In addition, a plant variety must be designated by a denomination. It is the applicant for the PVP right that must file with the PVP office a proposed denomination, which suitability is then evaluated by denomination examiners. The suitability of denominations is determined based on the rules enshrined in Article 20 of the UPOV Convention. If the denomination examiner deems suitable a proposed denomination and no third-party objection is filed (and

---

[8] U. Löscher, 'Variety Denomination according to Plant Breeders' Rights', 1986, Vol. 182, *Acta Horticulturae*, p. 59. Upon a mere visual inspection of the plant material alone (e.g.: seeds), the identification of a plant variety may not prove possible.

[9] UPOV was set up in 1961 based on the UPOV Convention (adopted in Paris, and later revised). It has its seat in Geneva (Switzerland), legal personality, and 78 Contracting Parties, among which the EU as inter-governmental organisation.

[10] The taxonomic ranks in the classification of plants are, in descending order: Kingdom, Division, Class, Order, Family, Genus, Species and Varieties. The only tangible embodiment is the variety: higher rankings in plant taxonomy remain theoretical abstract classifications (S. J. R. Bostyn, 'Patentability of Plants: At the Crossroads between Monopolizing Nature and Protecting Technological Innovation?', 2013, Vol. 16, Issue 3–4, *Journal of World Intellectual Property*, 105, p. 109).

[11] Article 6 of the UPOV Convention.

upheld) in the relevant time limit, that denomination is registered. Once registered, the denomination must be mandatorily used by any person marketing material of the variety in the territory where the variety is protected, even after the expiration of the PVP right.

One of the rules governing the suitability of denominations is that prescribing that a denomination cannot be identical nor too similar to an earlier registered denomination designating a plant variety belonging to the same or to a closely related species.[12] Two species are regarded as "closely related" if they belong to different species within the same genus. This means that a denomination must differ not only from denominations for plant varieties of the same species, but also from those belonging to a closely related species.[13]

When evaluating proposed denominations, examiners make use of similarity search engines with a global outreach, namely, the UPOV's PLUTO and the Community (EU) Plant Variety Office's ("CPVO") Variety Finder.[14] Both tools are equipped with a similarity testing functionality (*see* Figure 2 below) operating based on the same rule-based knockout algorithmic formula. Such formula leverages exclusively on text-based retrieval technologies, so there are no features in place for addressing the analysis of the similarity of designations from a phonetic or semantic perspective. The present study posits that the mentioned features can bring a strong added value for examiners and applicants alike. The possibility of conceiving such type of features to overcome the shortcomings described lies precisely at the heart of the current proposal for a modern IT architecture for similarity testing.

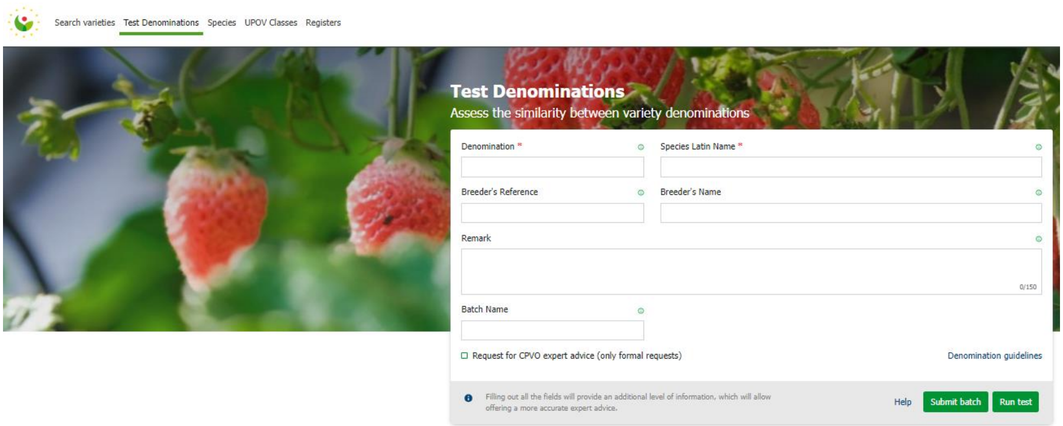

Figure 2: Functionality "Test Denominations" in Variety Finder © CPVO Variety Finder

---

[12] Article 20(2) of the UPOV Convention sets the grounds for the suitability of a denomination, reading as follows:

> *The denomination must enable the variety to be identified. […] It must not be liable to mislead or to cause confusion concerning the characteristics, value or identity of the variety […]. In particular, it must be different from every denomination which designates, in the territory of any Contracting Party, an existing variety of the same plant species or of a closely related species.*

[13] UPOV developed a system of classes in accordance with which botanical taxa within the same class are regarded as closely related. A denomination cannot be used more than once in the same "UPOV class", each class referring to a single genus. There are exceptions to this general rule, as laid down in Annex I of the UPOV Explanatory Notes on Variety Denominations.

[14] PLUTO and Variety Finder are respectively available at: https://www.upov.int/pluto/en/; and https://cpvo.europa.eu/en/applications-and-examinations/cpvo-variety-finder.

3. Foundations of an AI-powered model for testing similarity among plant variety denominations

In this section, the foundations of an AI-powered model for testing the similarity among denominations are presented. It is clarified at the outset that only the main contours of the model are delineated, since the "nitty-gritty" choices are left for each PVP office to define. The proposed model relies on a combination of AI technologies tailor-made for the multi-factorial assessment of the suitability of proposed denominations against designations in the state of the art. The model serves to retrieve all those earlier registered designations that are similar to the query designation inputted, to the extent that the similarity identified requires further individual (human) consideration before taking a decision on whether the designation tested effectively conflicts with an earlier right.

When devising the model's technicalities, a multidisciplinary team must be set up within the PVP office, comprising both data scientists and denomination examiners as domain-specific specialists. The latter are the best placed to bridge the gap between the model's theoretical foundations and its implementation in a live environment. Throughout the course of the design phase, they can act as interlocutors with data scientists, by providing guidance on the parameters to be encoded, evaluating the model's performance, and suggesting improvements. That said, the model does not necessarily have to be built from scratch: it is possible to capitalise on off-the-shelf commercial archetypes for trade mark similarity testing. The analogies between the examination of trade marks and that of denominations make them look like twins separated at birth, so it is worth taking advantage of them. An important facet of this convergence is, for instance, the fact that in both domains the examination of similarities is carried out at the *penumbra* of a certain degree of subjectivity.

In this connection, it is noted that the examination requires an input of (human) common sense, an "ingredient" that does not lend itself to easy algorithmic codification.[15] Attaining a not-so-perfect yet satisfactory output by means of "soft" analogical reasoning is, however, realistic. In any case, the success or failure in solving a complex problem by means of computer automation generally hinges upon the correct representation of the problem at the outset.[16] The "problem" in the case at stake should then be represented by the two facets of the "likelihood of confusion" test, namely, the identity ("same species") or similarity ("closely related species") of the confronted plant varieties, on the one hand, and the identity or similarity between the confronted designations, on the other hand[17]. Account taken of these benchmark factors, the machinery's pipeline should flow as follows:

---

[15] As M. Minsky put it, "easy things are hard" for machines (M. Minsky, 'Society of Mind', Simon and Schuster, 1987, p. 29).

[16] B. G. Buchanan and T. E. Headrick, 'Some Speculation About Artificial Intelligence and Legal Reasoning', 1970, Vol. 23, Issue 40, *Stanford Law Review*, 40, pp. 45–46.

[17] In cases of identity, the establishment of a conflict is straightforward, whereas in cases of similarity, an assessment of the likelihood of confusion must be carried out.

1. Demarcation of the relevant material scope of the search, based on the name of the species that is inputted into the system (along which the designation of interest);
2. Performance of the algorithmic operations for the multi-factorial assessment of similarity under the corresponding dedicated modules (i.e.: visual, phonetic, and conceptual automated comparisons); and
3. Return of the modules' outputs in a ranked order based on the degree of similarity.

The operations listed are described in detail in the upcoming sections below.

3.1. Demarcation of the relevant material scope of the search

The first operation in the pipeline concerns the identification of the species to which the plant variety in question belongs. Upon entering into the system the name of the species concerned a screening exercise is triggered, whereby the material scope of the search is limited to those earlier designations registered for varieties belonging to the same genus as that to which the concerned plant variety queried belongs. The internal workings of the functionality are relatively simple: each species has been manually attributed a code which the parameter is able to recognise each time. Upon the launch of a search, the program transmits the name of the designation to be tested along with the relevant species' code.

3.2. Multi-factorial assessment of the similarity among designations

The likelihood of confusion must be established based on the overall impression expected to be produced on the user when confronted to the designation in the tangible context of the marketplace.[18] To cover all facets of the user's polyhedric perception, the best-suited approach is that of combining various algorithms for addressing the visual, phonetic, and conceptual similarity comparisons. The algorithms should operate under the umbrella of the following search modules: i) string-based module; ii) shape module; iii) phonetic module; and iv) conceptual module (*see* Figure 3 below). All algorithms should serve to perform quantitative estimations of similarity between an entity and every other entity.[19] The criteria of "similarity" and "dissimilarity" should be the drivers here, based on a "binary choice" for the algorithms [i.e.: retrieving (+) or not retrieving (-) an earlier designation] coupled with heuristics to define the stopping conditions.[20] The referred four modules are envisaged to operate simultaneously yet independently, as gears of the same integrative machinery. Altogether, they lead to a "patchwork" of outputs, where each of these outputs are ranked based on a risk profile. Outputs can be served in a silver plate for the convenience of users in

[18] The relevant public targeted by plant varieties is made up of professionals in sectors concerned with the production or marketing of plant material (e.g.: breeders, seed propagators, growers, farmers, etc.). Their attention level towards denominations is high because they generally exhibit advanced biological knowledge of plants, place emphasis on technical specifications, and process information in a conscious manner.

[19] D. M. Katz, 'Quantitative Legal Prediction or How I Learned to Stop Worrying and Start Preparing for the Data-Driven Future of the Legal Services Industry', 2013, Vol. 62, Issue 4, *Emory Law Journal*, 909, p. 955.

[20] *Ibid*., p. 954.

one of the following forms: i) binary choice (flagged conflict); ii) statistical presentation (e.g.: 75% similarity match); or iii) traffic-light colour code (e.g.: the more intense the red colour, the higher the risk of confusion; while the "greener" the shade, the lower).[21] It is then for the (human) examiner or user to verify the reasonableness of the outputs as well as to perform the final operation of aggregating all outputs to complete the overall assessment.[22]

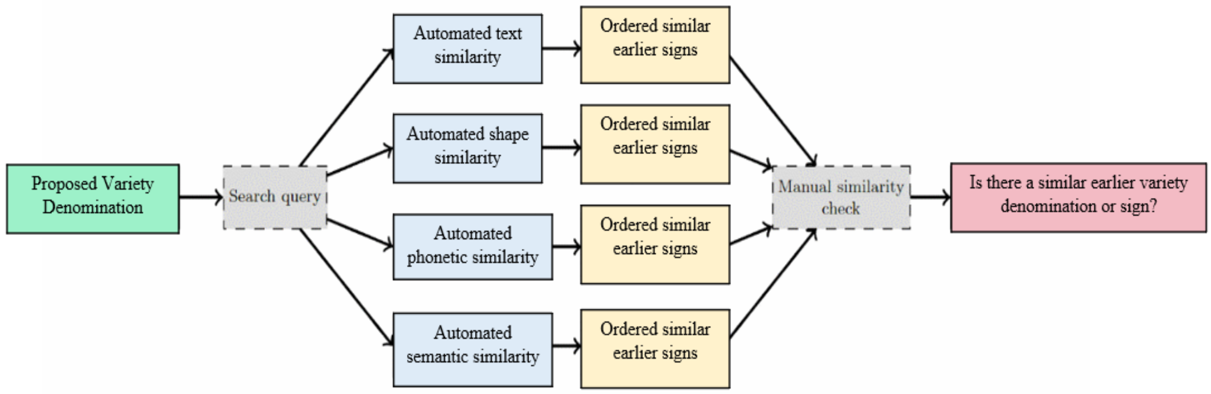

Figure 3: Multi-factorial Similarity Testing model © Author's creation based on I. Mosseri et al[23]

3.2.1. Visual similarity module

The visual similarity module is conceived to answer the question on whether each given pair of designations is visually similar and, if that is the case, to which degree. As a rule of thumb, a difference in only one character is deemed to be confusing, whereas differences in two or more letters are not deemed so. Oftentimes, similarities in denominations result from sharing the same start of end, or from swapping, adding, or deleting letters. The length of designations is also relevant, the effect in differences being stronger in shorter designations. A good algorithm should work well around all these circumstances.

For the visual comparison, two types of method can be relied on simultaneously, and their respective outputs be regarded as complementary: the text-based and the shape-based similarity methods. The text-based similarity method serves to measure the "distance" between two text strings by performing a string-matching check based on word length and on the number of similar and dissimilar characters.[24] For each pair of strings, the first string is

[21] D. S. Gangjee, 'A Quotidian Revolution: Artificial Intelligence and Trade Mark Law' in R. Abbott, *Research Handbook on Intellectual Property and Artificial Intelligence*, Edward Elgar Publishing, 2022, 325 (hereinafter, "D. S. Gangjee, A Quotidian Revolution"), p. 341.

[22] The idea of the multi-modular design has been taken from: I. Mosseri et al, 'TradeMarker: Artificial Intelligence based Trademarks Similarity Search Engine', 2019, *Conference: HCI International 2019* (Orlando, Florida) (hereinafter, "I. Mosseri, TradeMarker"), p. 3. It is fair to ask if a model covering the totality of the assessment could be envisaged. It is observed in this regard that compressing into a single metric the outcomes of separate assessments entails the risk that if an important circumstance is overlooked, the resulting final metric is "contaminated".

[23] I. Mosseri, TradeMarker, p. 3.

[24] The index is calculated based on the ratio between the number of common characters and the total number of characters of the query word (C. V. Trappey et al, 'Intelligent trademark similarity analysis of image, spelling, and phonetic features using machine learning methodologies', 2020, Vol. 45, *Advanced Eng. Informatics*, 2).

the text making up the designation queried, and the second string is the text making up the relevant earlier designation in the database. The size of the difference is computed with reference to a "similarity index" on a scale of value ranging, for example, from 0 to 1. Once the referred tests performed, a numerical "cut-off value" is applied onto them (e.g.: 0.7). Those designations that are found to be "too similar" to the designation queried are then captured by such threshold and returned as part of the output. The output consists in a list ranking the similarities in order of relevance. The higher the number retrieved is to the value 1, the higher the degree of similarity and thus the higher the likelihood of confusion.

Regarding the shape-based similarity testing method (*see* Figure 4 below), this method relies on an algorithm serving to compute the visual similarity between a pair of objects based on the use of "shape descriptors". The method can be applied to the analysis of denominations as based on the shape contours of the (verbal) characters in denominations. The algorithm of choice should then serve to identify similarities in the shapes of comparable characters, such as the number "1" and the capitalised letter "I" (*see* Figure 5 below).

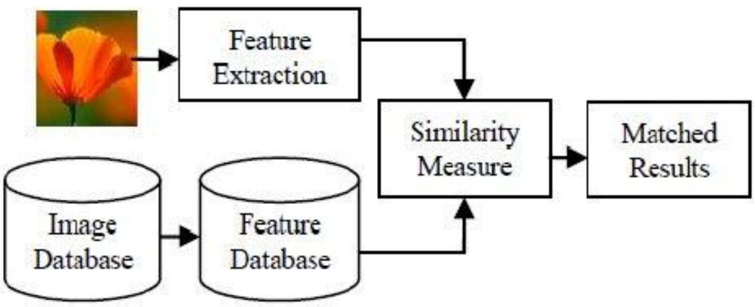

Figure 4: Content-based image retrieval method © K. Y. Ashwani et al[25]

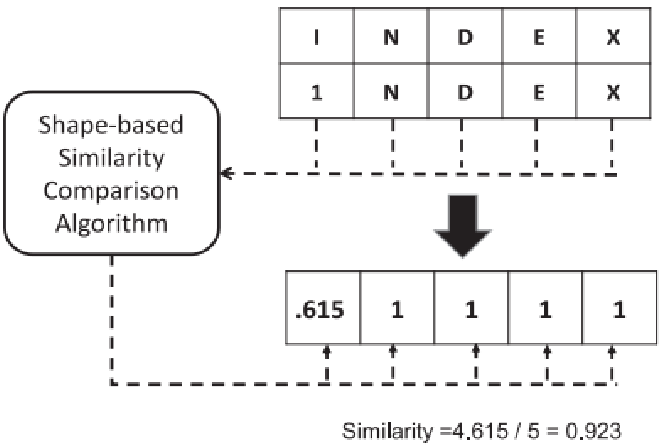

Figure 5: Shape-based similarity computation method © R. Setchi and F. M. Anuar[26]

[25] K. Y. Ashwani et al, 'Survey on Content-based Image Retrieval and Texture Analysis with Applications', 2014, Vol. 7, Issue 6, *International Journal of Signal Processing, Image Processing and Pattern Recognition*, 41, p. 43.

[26] R. Setchi and F. M. Anuar, 'Multi-Faceted Assessment of Trademark Similarity', 2016, Issue 65, *Expert Systems with Applications*, 16 (hereinafter, "R. Setchi, Multi-Faceted Assessment"), p. 20.

3.2.2. Phonetic similarity module

The phonetic similarity module begs the question on whether each given pair of designations is aurally similar and, if that is the case, to which degree. What matters is the overall phonetic impression produced by each designation, as influenced by the number and sequence of its syllables. The most obvious form of aural similarity between two terms is the use of letters in the same order, as resulting in a common rhythm and intonation. On the other hand, switching the position of letters can impact the pronunciation and distance the confronted designations on the phonetic front. A good phonetic algorithm should differentiate between similar phoneme pairs (e.g.: "m" and "n") and less similar ones (e.g.: "m" and "p"). In addition, it should be able to convert special characters or symbols to their corresponding semantic meaning (e.g.: "&" is converted to the word "and"), thereby mimicking human perception.[27]

Pronunciation-based algorithms usually involve a step of extraction of the phonological features of the phonemes present.[28] The algorithm divides the text into "sound clusters", based on which a phonetic "representation" for each designation in the pair is created. The two representations are then compared and, if deemed similar, it is inferred that the confronted designations are pronounced similarly. Examples of well-known phonetic algorithms are "Soundex" and "Double Metaphone".[29] The choice of the algorithm should in any case hinge upon the language(s) spoken in the territory of the legal order concerned.

3.2.3. Conceptual similarity module

The conceptual similarity module serves to answer the question on whether each given pair of designations is semantically similar and, if that is the case, to which degree. From the conceptual perspective, an algorithm should be capable of considering synonyms and antonyms, comparable words with similar meanings in the languages of interest, and lexical relations.[30] In addition, the algorithm should be able to convert special characters, symbols, or numbers fulfilling the role of letters to their corresponding semantic meaning,[31] as well as to dissect composite designations into their respective individual constituents.[32]

---

[27] C. J. Fall and C. Giraud-Carrier, 'Searching Trademark Databases for Verbal Similarities', 2005, Issue 27, *World Patent Information*, 135 (hereinafter, "C. J. Fall, Searching Trademark Databases"), p. 140.

[28] E. García Rodríguez et al, 'El Uso de Sistemas Inteligentes en el Registro de la Propiedad Industrial', 2020, Vol. 30, *La Propiedad Inmaterial*, 295, p. 308.

[29] "Soundex" was developed in 1918 by R. C. Russell and M. K Odell. Double Metaphone was developed in 1990 by L. Phillips.

[30] F. M. Anuar et al, 'A Conceptual Model of Trademark Retrieval', 2013, *Procedia Computer Science*, 450 (hereinafter, "F. M. Anuar et al, A Conceptual Model"), p. 453.

[31] Oftentimes, applicants propose designations containing semantic "word plays" (e.g.: "4ever", where the number "4" stands for the combination of letters "for", which together with the letters "ever" form the word "forever").

[32] Joining two terms with meaning can impact the overall meaning conveyed by the designation. The system should extract all possible meanings in a designation composed of several concatenated words (e.g.: in the designation 'Cooking', the meaning for the words "cooking", "cook" and "king" should be extracted).

It is important to remark that the conceptual comparison must be applied independently from textual and phonetic comparisons. For example: two designations looking similar may convey different semantic meanings (e.g.: "dessert" vs "desert"), while two looking different might converge in meaning (e.g.: "H2O" vs "water").[33] Another interesting situation is that of misspellings, that is, of words which are spelled incorrectly (e.g.: "Yelow" vs "Yellow"). If the semantic meaning of the well-spelled word can still be grasped in the misspelled word, then this means that it is the semantic content of the well-spelled word that must be considered for the semantic analysis. Account must also be taken of "homophones" [i.e.: words pronounced identically or similarly but having different semantic meanings (e.g.: the terms "sea" and "see")] and "homographs" [i.e.: words that are spelled identically but which meaning varies depending on the syntactic context in which they are placed (e.g.: the term "fair", which can mean "bazaar" or "just")]. Further, capitalisation alone can introduce notable differences [e.g.: "Polish" (national from Poland) vs "polish" (product to be used on furniture or shoes)].[34]

When scanning through the current state of the art of methods for automated semantic similarity testing,[35] it is observed that many computation models rely on natural language processing techniques coupled with external knowledge sources such as lexical ontologies[36]. These ontologies consist in "distributed word representations" or "word vectors" that have been trained on large amounts of linguistic data. Most current lexical ontologies follow the structure of the reference ontology "WordNet",[37] a freely available large electronic lexical database of the English language that was developed based on psycholinguistic theories. For multilinguistic demands, the word representation library "fastText" might do the job, as it works with word vectors pre-trained for more than 150 languages.[38]

3.3. Customisation of the parameters for a quality output

For any test performed, the algorithms should retrieve all those earlier designations in the database that are significantly similar. To achieve a good threshold of accuracy, algorithmic

---

[33] D. S. Gangjee, A Quotidian Revolution, p. 334.

[34] *Ibid*.

[35] *See, for instance*, J. Oliva et al, 'SyMSS: A Syntax-based Measure for Short-text Semantic Similarity', 2011, Vol. 70, Issue 4, *Data and Knowledge Engineering*, 390; and F. M. Anuar et al, A Conceptual Model, p. 454.

[36] A "lexical ontology" is a framework specifying the underlying structure and lexical relationships for knowledge representation and organization of lexical information (F. M. Anuar et al, 'Semantic retrieval of trademarks based on conceptual similarity', *IEEE Transaction of Systems, Man and Cybernetics: System*, 2016, Vol. 46, Issue 2, 220). Many methods for semantic similarity testing rely on two sets of features: the "token feature set" and the "lexicon feature set". The token feature set consists in a set of words included in a sign (e.g.: the token feature set for the designation "Red Bull" is "red, bull"), while the lexicon feature set includes specific sub-sets encompassing synonyms, direct hypernyms, and direct hyponyms. The similarity score is then computed using a contrast similarity model capable of considering the number of shared features and the edge-based word similarity score between the tokens as derived using a lexical ontology (A. Tversky, 'Features of similarity', 1977, Issue 84, *Psychology Review*, 327).

[37] R. Setchi, Multi-Faceted Assessment, p. 19. WordNet was developed in 1986 at Princeton University.

[38] fastText is available at: https://fasttext.cc/.

parameters must be customised. One must here accept the crude truth that there will always be similarities passing under the radar as well as dissimilarities that will be wrongly flagged. This phenomenon remits to the tug-of-war between the notions of "precision" and "recall". The notion of "precision" can be defined as the ratio between the number of correct hits and the total number of returned hits, while the notion of "recall" refers to the ratio between the number of returned correct hits and the total number of correct results that should have been returned. The two notions can be better understood under the light of the notions of "true positive", "false positive", and "false negatives", defined as follows:

- A "true positive" is the baseline upon which a search hit matches the reality which it is intended to reflect.
- A "false positive" is, by reference to the notion of "true positive", that search hit which is returned as if it were a "true positive" when it is actually not so.
- A "false negative" is that search hit which should have been returned but which the system failed to identify.

Search systems can be tuned to favour either recall or precision, by choosing the maximum distance between the search string and the target terms. If high precision is chosen, the system returns relevant results but fails to return other less relevant results (i.e.: low recall). Contrarily, if high recall is chosen, the system returns most or all the relevant items but at the expense of an excessively high number of irrelevant results (i.e.: low precision).[39] It is preferable to tip the scales towards high recall because it is logically better to retrieve some false positives than to miss relevant hits. However, high recall is no walk in the park. Research shows that the risk of human error is directly proportional to the size of the list of retrieve hits increases.[40] This risk can be minimised if certain mitigating measures are taken. For instance, highly similar hits could be flagged in an eye-catching manner. In this spirit, the algorithm could be configured in a way such that it can predict the degrees of similarity by means of a scoring system with the following values: very low (1), low (2), medium (3), high (4), and very high (5). The results could then be ranked in accordance with the said values in the form of a user-friendly "star-rating" system, such as a drop-down catalogue where users may filter results based on the rating of interest.

Metric tests are needed over time to establish the reliable base-truth set of data and cut-off value for optimising the model's performance. The prediction capability is compared with the base-truth data and then, to minimise the error between these two values, the model's parameters are adjusted.[41] Overall, the model must be regarded as a dynamic environment and as such be made subject to regular upgrades based on customisation exercises.

[39] C. J. Fall, Searching Trademark Databases, p. 142.

[40] *Ibid*., p. 141.

[41] D. Lim, 'Trademark Confusion Revealed: An Empirical Analysis', 2022, Issue 71, *American University Law Review*, 1285, p. 1358.

4. Considerations about the users of the model

The similarity testing model proposed is conceived as a "bicephalous" system serving to address the needs of the users targeted: one "head" being the back-office for exclusive use by denomination examiners, and the other being the front-office accessible to any external user (e.g.: applicants for PVP rights). While denomination examiners seek information to serve as basis in their decision-taking processes, applicants do so for anticipating if the denomination which they intend to file with the PVP office is available and thus likely to be deemed suitable.[42] Given that both informative needs are satisfied with the same datasets, the two "heads" of the *amphisbaena* system are regarded as two doors to the same reality (*see* Figure 6 below). However, the user difference in terms of motivation for using the system demands a differentiated treatment, to be embodied in the interface's formal presentation.

Figure 6: Bicephalous system © Simpsons Wiki[43] (modified by author)

Regarding denomination examiners, these are the ultimate gatekeepers for the approval of denominations. They must assess the reasonableness of the computer-generated outputs and aggregate these manually to take a decision on the overall assessment of similarity. It is important that they receive training for working alongside the model and for monitoring its functioning. Examiners should also be wary of the system's limitations as well as of their own biases as humans interacting with computers. Indeed, psychological research on human-computer interfaces demonstrates that automated devices can fundamentally change the way in which persons approach their work, what in turn leads to new kinds of errors.[44] In particular, the "automation bias", in accordance with which humans generally attribute a superior judgment to automated aid (without much reflection nor verification), can lead to

[42] The denomination examiner and the applicant usually act in anticipation of the other's actions or incentives, thereby adding new "noise" into the system on each occasion of interaction. This dynamic cycle entails that, over time, the substance of applied for and registered denominations keeps evolving (S. Katyal and A. Kesari, 'Trademark Search: Artificial Intelligence and the Role of the Private Sector', 2021, *Berkeley Technology Law Journal*, 501, pp. 576–577). The influential power of users over the development of the AI model can be leveraged upon if denomination registration is framed as an adversarial Machine Learning problem, that is, as a scenario where data distributions underlying applications change in response to exogenous stimuli.

[43] Simpsons Wiki, Two-headed dog, CC BY-SA, available at: https://simpsons.fandom.com/wiki/Two-Headed_Dog.

[44] R. Parasuraman and V. Riley, 'Humans and automation', 1997, Vol. 39, Issue 2, *Human Factors*, 230.

failures in the detection of errors in machine-generated outputs.[45] There is thus a risk that examiners become overconfident at the stage of the manual validation of the model's results, thereby mistakenly assuming that these must always necessarily be quality results, even when this is not so. A reminder is needed that the outputs of similarity testing models yield probabilities, not certainties.[46] As graciously expressed by Dr. Gangjee, "one should not get carried away by the seductive certainty of an enumerated risk".[47]

Regarding external users, the front-office interface should ideally be conceived as a "pre-apply" service platform to test the designation of interest against the state of the art of registrations. Denominations are granted on a "first-come, first-served basis", so ensuring *ex ante* that the designation of interest is available proves crucial. The pre-apply services could thus enable applicants to anticipate potential refusals by PVP offices as well as potential third-party objections based on earlier rights.

To augment user experience, it is advised that pre-apply services feature the following elements:

- Detailed explanatory notes about the structure and functioning of the model;
- A disclaimer warning that error-free test results cannot be guaranteed;
- A disclaimer warning that the test results cannot be regarded as definitive and that if the tested designation is filed for registration, it will be made subject to the evaluation and decision of examiners;
- For each test successfully completed, a drop-down tab can be rendered accessible for providing information on the "reasoning" paths followed by the algorithms;[48]
- A voluntary survey for users to rate their satisfaction with the services;
- A feedback space for users to report issues encountered when making use of the pre-apply services, including apparent output errors or biases.

[45] M. T. Dzindolet et al, 'The role of trust in automation reliance', 2003, Vol. 58, Issue 6, *International Journal of Human-Computer Studies*, 697, p. 635.

[46] A. Završnik, 'Algorithmic justice in criminal justice settings', 2021, Vol. 18, Issue 5, *European Journal of criminology*, 623, p. 632.

[47] D. S. Gangjee, A Quotidian Revolution, p. 338

[48] This action serves to build user trust and to adhere to the principle of "explainability" (i.e.: the action of "unearthing" in a human-intelligible manner the process followed by algorithms to get to a given output).

5. Final takeaways

"The best way to predict the future is to invent it",[49] and it seems that the time is ripe for this future to hit denominations' town. In this study, the conceptual foundations for an AI model to support with the similarity testing of denominations have been outlined. The main takeaways from the investigation can be wrapped-up in the following ten points:

1. A PVP office considering availing itself of a computer-automated model for testing the similarity among denominations is advised to take advantage of the methodological cross-fertilisation between the domains of denominations and trade marks.
2. The examination of proposed denominations is conducted under the Damocles' sword of the examination's inherent subjectivity. Further, common-sense nuances governing the assessment of the suitability of denominations represent a hurdle for machines.
3. A multi-factorial similarity system composed of four modules has been proposed, leveraging on similarity metrics applying from the visual, phonetic, and conceptual perspective.
4. The visual similarity testing can rely on a combination of text-based and a shape-based modules. For the phonetic similarity testing, algorithms such as Soundex or Double Metaphone can be used. As for the conceptual similarity testing, a functional connection with a lexical ontology that works with the language of interest is needed.
5. Once the concerned algorithmic operations are completed, it is for users to evaluate the reasonableness of the output returned and to take such output *cum grano salis*.
6. The archetype proposed bears two facets: an outward facet, consisting in a user-interface for external users and an inward facet for denomination examiners.
7. Regarding denomination examiners, relying on the system is expected to ease their daily endeavour of assessing the legal compliance of proposed denominations. This computer-aid can ultimately foster the adoption of agile and consistent decisions.
8. When it comes to external users, the notion of "user-friendliness" takes centre stage. The model's external façade can be conceived in the form of pre-apply services and it should convey information in a clear and effective manner.
9. When customising the parameters underlying the algorithms, the model's designers must strike the right balance between the notions of "precision" and "recall".
10. The similarity testing model is conceived as a dynamic and evolving environment, in line with the motto "the road to success is always under construction".[50]

[49] This quote is attributed to the computer scientist A. Kay.

[50] This quote is attributed to the actress M. J. "Lily" Tomlin.