Analysis of Trends in Search and Retrieval of Trademarks and Plant Variety Rights Information

7 October 2005

Clunies Ross St
Black Mountain ACT
www.cambia.org

Analysis of Trends in Search and Retrieval of Trademarks and Plant Variety Rights Information

Marie Connett Porceddu *PhD*
Douglas Ashton
Neil Bacon

CONFIDENTIAL

Analysis of Trends in Search and Retrieval of Trademarks and Plant Variety Rights Information

Introduction

There are aspects of search and retrieval of trademarks and plant variety rights information that present some specific challenges on which we have been asked to comment in this paper, supplemental to our related paper[97]. Here again, the perspective offered is that the interests of the internal customers, *i.e.* the examiners, and the external customers and stakeholders (such as business marketers, packaging industries, breeders, customs inspectors, and the public in general) may be different, but understanding of trends in search and retrieval and future plans for implementation of certain improvements relative to these trends may be beneficial to both.

Trademarks

General features of trademark searching

Although there are other aims of trademark searching, a common cause for both examiners and other stakeholders (*e.g.* businesses making applications, anti-infringement and counterfeit enforcement agents) is to uncover aspects that may be deceptively similar to existing marks covering similar products and services, *i.e.* within the same Nice classifications.

Distinguishing characteristics and similarities may be of both figurative and verbal types. Verbal similarities may include visual, phonological, and synonymic aspects, all of which comprise important language-specific factors. Thus, one section below relates to pattern recognition, with a focus on searching visual images, and another relates to particularities of foreign characters as marks, for example distinguishing characters of similar appearance or that have phonetical equivalence when the alphabet used is non-Roman.

Pattern recognition

Pattern recognition is a field within the area of machine learning with the aim to classify data (patterns) based either on *a priori* knowledge or on statistical information extracted from the patterns. More specifically, a pattern recognition system consists of a means of **sensing** something such as a pattern of high density marks within a low density background or a pattern of vibrations; a **feature extraction** mechanism that computes information from the observations gathered by the sensor; and **a classification or description scheme** that analyses the extracted information based either on probabilities extrapolated from the statistical regularities in it, or by comparing its structure to a set of patterns that have already been classified or described, known as the **training set**[98]. Mechanisms such as associative neural networks are commonly used for training.

Specific applications of pattern recognition include speech recognition, Optical Character

---

[97] Connett-Porceddu, Marie, Ashton DE, Bacon N, dos Remedios N, Nottenburg C, Okada S, Quinn G, Wei Y and Jefferson RA (2005) Analysis of Trends in Search and Retrieval of Intellectual Property-Related Information, IP Australia.

[98] http://en.wikipedia.org/wiki/Pattern_recognition

Recognition (OCR) of machine printed or handwritten text, text pattern classification (*e.g*. E-mail spam filters, or a search engine finding "similar" documents based on frequent co-occurrence of words), and image analysis, which ranges from bar-code reading to identification of human faces, fingerprints or iris patterns and identifying similar patterns in images (*e.g.* trademarks of various sorts including **shape marks**).

Our related paper[99] contains extensive discussion of Optical Character Recognition in the context of full-text and other text-based searching of patent documents, but the comments made in that context about the accuracy of OCR algorithms in general, and OCR of Chinese, Japanese, and Korean (CJK) characters in particular, are certainly relevant to searching trademarks[100]. Accuracy is affected by certain stylistic elements in the calligraphy such as brushy or indistinct edges, colours, and incorporated pictorial features, but based on the training set chosen, OCR engines using mechanisms such as associative memory neural networks can be "trained" to handle such elements.

Image analysis is the subtopic of pattern recognition concerned specifically with processing digital images. In the related paper design patent searching was addressed in connection with a survey of patent information search sites, and it was noted that visual image searching would be useful but is not much used in design patent searching because most searching is done using classifications. Image searching may however be similarly or more useful for normal trademarks, and shape marks. It was, however, noted that although this has been a field of informatics research since the 1950s, the software available still has significant technical limitations, and although there has been some valuable research work related to and tested on trademark recognition[101], it is still essentially at a prototype level, and the greatest technical progress has tended to focus on security-related images such as faces and fingerprints rather than IP industry uses[102].

Similarly, algorithms very much like those used in speech recognition can be used for processing **sound marks**, but again most of the emphasis has been on development of training sets for the human vocal range, and less into other types of sounds. In general, except for patterns with simple characteristics such as bar codes, the human eye and ear are still much more sophisticated pattern recognition tools than any software.

- However, as computing power at reasonable costs continues to expand, this is an area of endeavour that merits periodic re-examination.

---

[99] Connett-Porceddu, Marie, Ashton DE, Bacon N, dos Remedios N, Nottenburg C, Okada S, Quinn G, Wei Y and Jefferson RA (2005) Analysis of Trends in Search and Retrieval of Intellectual Property-Related Information, IP Australia.

[100] Outside the scope of this report, CAMBIA's informatics team had recently conducted extensive trials of the latest OCR technology with a focus on the recognition of Chinese and Japanese as well as European languages (evaluations of accuracy were done by native speakers currently employed by CAMBIA as IP research analysts), and was able to identify OCR engines that with our applications achieved accuracy levels of over 99% on patent documents provided in Chinese and Japanese.

[101] For example, Jain AK, Vailaya A (1998) Shape-based Retrieval: A Case Study with Trademark Image Databases. *Pattern Recognition* 31:1369-90.; Eakins, JP, Boardman JM and Graham ME (1998) Similarity Retrieval of Trademark Images. *Multimedia* 5:55-63, which also uses shape similarity, and Zhao T, Tang HL, Ip HHS, Qi F (2002) Content-based trademark recognition and retrieval based on discrete synergetic neural network. *Distributed Multimedia Databases: Techniques and Applications*, pp. 58 – 72. A useful survey of the trademark visual matching literature may be found in Leung WH and T. Chen (2002) Trademark retrieval using contour-skeleton stroke classification, *IEEE International Conference on Multimedia and Expo,* http://amp.ece.cmu.edu/Publication/Howard/icme2002_howard.pdf. Part of a thesis submitted to Carnegie Mellon University in 2003, it also describes a method to retrieve trademarks using query by sketches. Trademark images are filtered to remove noise, then segmented based on pixel connectivity. Either thinning or edge extraction is applied to each region to produce a stroke sketch.

[102] Fall CJ, Giraud-Carrier C (2005) Searching trademark databases for verbal similarities. *World Patent Information* 27: 135-143, faculty.cs.byu.edu/~cgc/Research/Publications/WPI2005.pdf

There is at least one significant development on the horizon of image recognition technologies specifically for trademark and design searching, the European eMAGE project, which will use an associative memory neural network for pattern characterisation.[103]

This project, first reported in 1998, has already undergone delivery delays and has yet to materialise[104]. However, there is a commitment by the French INPI as lead agency to use it and market it. Development in the current phase has been at a cost of about 2 million Euros[105], approximately half funded by the European Commission under the *eContent* programme[106], largely in a contract to the French company LTU technologies[107]. This company, already reputed for image processing software for law enforcement and intelligence, was acquired in March of this year by JASTEC International Inc., a US subsidiary of JASTEC Co., Ltd., a Japanese software development and systems integration company.[108] The business model of the eMAGE consortium is to out-license this software to other national IP offices, to marketing, packaging and IP divisions in private companies, and to anti-counterfeiting authorities such as those associated with ports and border controls.[109]

The searches will be against European registered logos and industrial designs databases; development has used sample data provided by the French and Portuguese INPI offices and the Œsterreichisches Patentamt, and search capacities are being set up for French, Portuguese, German and English. The project is not currently envisioned to encompass other languages[110] such as languages rendered in non-Roman characters.

- If IP Australia were to contemplate licensing this system, it may be desirable to implement measures that would assist searching languages more common in Australasia (see below).
- Often better results may arise from combining several methods rather than using a single method such as statistical analysis of Fourier descriptors, invariant moments or Zernike moments[111]. *Ad hoc* weighting functions may be used in the combination[112], and image analysis methods may even be used in combination with full-text searching methods, which if used as an initial filter for the images may decrease computational intensivity[113]. This could contribute to an interim solution for IP Australia.

Special features related to foreign word marks, particularly CJK characters

In English and many other languages in which it is possible to find similarities based on letter order, or similar length with a number of letters in common, “fuzzy search” algorithms have

[103] http://www.eu-projects.com/emage/Details.htm
[104] http://www.eu-projects.com/emage/results.htm
[105] http://www.eu-projects.com/emage/Details.htm
[106] http://www.cordis.lu/econtent/
[107] http://www.ltutech.com/en/
[108] http://www.ltutech.com/en/news.press-release-2005-04-18-00.html
[109] http://www.eu-projects.com/emage/objectives.htm
[110] http://www.eu-projects.com/emage/Details.htm
[111] Fall CJ, Giraud-Carrier C (2005), *ibid*.
[112] Chan DYM, King KCI (1999) Genetic algorithm for weights assignment in dissimilarity function for trademark retrieval. *Third International Conference on Visual Information and Information Systems, Lecture Notes in Computer Science* 1614: 557—565, http://www.cs.cuhk.hk/%7Eking/PUB/chan99a.pdf, uses an algorithm to determine weightings consistent with human judgements in order to classify 1360 monochromatic trademarks.
[113] Ravela S, Manmatha R (1999) Multi-modal retrieval of trademark images using global similarity. *U. Massachusetts Computer Science Technical Report* TR99-32, http://ciir.cs.umass.edu/pubfiles/mm-26.pdf describes initially narrowing the set of possibly similar trademark images by using full text search on the text associated with trademarks. The images are then filtered with Gaussian derivatives used to calculate geometric curvature and phase, and the similarity in the distribution of these statistics is then used as a measure of the similarity of the images.

been designed to focus on starting and ending letters with letters added, transposed, or deleted, similar to applications in molecular biology for aligning similar DNA sequences[114] as discussed in our related paper. However, with regard to the visual similarities of words, the common algorithms used to determine the similarity of words written in Roman letters are not applicable to languages written in ideographic scripts, such as Japanese or Chinese.

In many languages a given character can be written in different ways. For example Chinese characters can be written in a handwritten style, simplified characters are now used as a standard in mainland China, and more traditional complex characters are used in Hong Kong and overseas Chinese communities. In addition, calligraphy is often a form of artistic expression with diverse styles. Consequently, two characters that look quite different may be recognized as synonymous by a native speaker. While staff having specialized cultural and linguistic skills is one solution to this challenge for examiners, access to native speakers of many different languages and dialects would be required to address it fully.

- A recommendation that would enable any government to at least partially address this complexity for searching visual similarities using informatics is that applicants should be required to submit non-Roman letter-based word marks, in addition to the actual planned logo, using a standard character encoding system such as Unicode.[115]
- If it is not preferable to address this with a regulatory change, it could still be facilitated through the use of Optical Character Recognition (OCR) techniques and or skilled staff to extract text from the images submitted according to current requirements. The use of a standard character encoding would allow machine translation and other computational linguistic techniques to be applied to more common languages such as Chinese and Japanese.

Skilled trademark searchers often use a variety of techniques such as part word searches to identify phonetically similar word marks. These rely on understanding of the peculiarities of spelling and pronunciation to predict the likely pronunciation of “made up” words such as “h@ppy” or “STOCKX” or “OZZEE”. Numerous algorithms have been developed to aid in the automatic identification of similar words, using phonetic and or letter sequence-based approaches. This approach is obviously also language dependent, but considerable advances have been made particularly in countries with multilingual cultures, such as Switzerland; some of the better known approaches are edit distance (also known as the Levenshtein algorithm), N-grams, and Soundex[116]. There are open source software packages such as ht://Dig incorporating these algorithms, that provide rules for handling accented characters and capabilities for fuzzy searching once words are reduced to phonetic codes. However, these approaches are still based on a requirement that the pronunciation of words in the mark is known for comparison via these methods to the pronunciations of possibly similar words in a reference database.[117]

Currently a typical requirement is the applications country official language and a translation. For example, the requirement of the IP Australia Trade Marks Office Manual of Practice & Procedure[118] is that if the representation of a trade mark in an application for registration of the trademark includes words of a language other than English, the applicant must file in support of the application a translation of the words into English. Further, if the representation of a trademark in an application for registration of the trademark includes characters

[114] Fischer I Zell A (2000) String averages and self-organizing maps for strings. *Proceedings of the Second ICSC symposium on Neural Computation*, pp. 208-15.
[115] http://www.unicode.org/
[116] For more details see Fall CJ, Giraud-Carrier C (2005), *ibid.*
[117] Fall CJ, Giraud-Carrier C (2005), *ibid.*
[118] http://www.ipaustralia.gov.au/pdfs/tmmanual/part_10_details_of_formality_requirements/4._translation_transliteration_of_foreign_workds_and_non-roman_characters.htm

constituting words in characters that are not Roman letters, the applicant must file both the translation and a transliteration of the characters into Roman letters, using the recognised system of Romanisation of the characters.

Unfortunately, transliteration of non-Roman characters, even if performed to a high standard using recognized systems such as Pinyin, is problematic as a guide to pronunciation, particularly if characters are pronounced differently in different regions and dialects, so the current IP Australia requirement is not suited to assist the searcher to determine whether characters are phonetically equivalent. Furthermore, speakers from different Chinese-speaking regions may legitimately transliterate or Romanise the same Chinese name differently, particularly because of the variety of ways in which a single combination of characters can be pronounced. This can create difficulties in searching all variations. Fortunately, as dicussed in our related paper[119], each region uses a relatively small and consistent set of characters when transliterating.

- Thus, it may be useful for IP Australia to consider making the requirement for transliteration to include the common transliterative variants, as well as a phonetic rendering of the words with any common variants. This would be important to detect an element that appears prominently in a trademark but which is pronounced similarly to a generic term or previously trademarked name (for example, 可口可乐, which can legitimately be transliterated “kekoukele", refers to "Coca Cola"[120]).

Plant variety rights

General features of plant variety rights information searching

- While within a particular country plant variety rights are often readily searchable by examiners and the public, it may be of benefit to breeders and producers to be able to search the rights issued by other countries. This may be most easily accomplished with use of the data available from the International Union for the Protection of New Varieties of Plants (UPOV)[121], which was established by the International Convention for the Protection of New Varieties of Plants adopted in Paris in 1961 and revised in 1972, 1978 and 1991.

The main reasons this enlargement of the data available would be desirable are similar to those outlined for patent information in our related paper. Plant-derived products from one country that are subject to plant variety protection in other countries may be subject to impoundment when imported into these countries. With the standardization afforded by TRIPS, many multinationals are becoming increasingly active about protecting their PVR rights by this mechanism[122].

DUS test records

Another reason for facilitating searches of the plant protection information of other jurisdictions is that it can provide other information of value to Australian breeders and primary producers. For example, many European countries carry out VCU (Value for

---

[119] Connett-Porceddu, Marie, Ashton DE, Bacon N, dos Remedios N, Nottenburg C, Okada S, Quinn G, Wei Y and Jefferson RA (2005) Analysis of Trends in Search and Retrieval of Intellectual Property-Related Information, IP Australia.

[120] See http://www.daochinasite.com/eng/study/brand.shtml for some examples of Chinese versions of well known brand names.

[121] http://www.upov.int/

[122] van Wijk AJP, Louwaars NP and Eaton DJF (2004) Framework for the Introduction of Plant Variety Protection in Developing Countries. Centre for Genetic Resources, Wageningen.

Cultivation and Use) indexing for disease resistance and other performance characteristics concurrently with the DUS (Distinctiveness, Uniformity, Stability) testing required for determination of novelty, at least for certain food crops (particularly vegetables and cereals)[123]. Some countries, notably the Netherlands, are also experimenting with provision of visual documentation of the distinctive characteristics, for example in ornamental species by a requirement for standardized digital photographs of particular views at set distances from buds and flowers at various developmental stages[124].

It may be of value for IP Australia to provide more information on the website about the Technical Guidelines that are used for DUS testing, particularly as these are actively evolving in international collaboration in the Technical Working Parties of UPOV. For example, the definition of agreed standards on *Eucalyptus* is being led by Brazil in the Technical Working Party for Ornamental Plants and Forest Trees[125].

Plant patents and other *sui generis* systems

A complexity or advantage for searching plant variety rights in different countries is that they may be found in several different systems. UPOV allows the development of *sui generis* systems, for example, which are in active development in some of Australia's markets, and in countries from which plant material, particularly ornamentals and medicinal plants, arise, *e.g.* Thailand, Indonesia[126] and India[127].

In the USA, the use of plant varieties may obtain government-approved monopoly rights under one of three different systems: plant variety protection certificates (PVP) under the U.S. Plant Variety Protection Act (1970) cover plants that are sexually reproduced[128], with the intent to conform in UPOV 1978 and 1991; plant patents according to the U.S. Plant Patent Act (1930) cover plants that are asexually reproduced (*e.g.* by tissue culture, cuttings etc.) except tubers; and plants and plant parts may be covered by utility patents, for example in product-by-process claims.

- Searches of both can be provided in the USPTO patent databases[129], though they do differ in that the former may be infringed only by asexual propagation from the actual plant protected by the patent, and as there are no annuities, there is not the possibility as with utility patents that a failure to pay annuities may cause the subject matter to enter the public domain.

Australia's system has also incorporated various protection schemes and the search site could usefully provide more clarification on this. For example, a limited comparison is provided between Plant Variety Rights under the PVR Act of 1987 and Plant Breeding Rights under the PBR Act of 1994[130], but there is only a single paragraph on the interface between

[123] Dr AJP van Wijk, Centre for Genetic Resources, Wageningen, the Netherlands, personal communication.
[124] Dr Ir Liesbeth Hof, Centre for Genetic Resources, Wageningen, the Netherlands, personal communication.
[125] Dr Daniela de Moraes, National Plant Varieties Protection Service, Brasilia, personal communication.
[126] Dr. Ir Sugiono Moeljopawiro, Department of Agriculture, Center for Plant Variety Protection, Jakarta, Indonesia, personal communication.
[127] Prof. Dharam Pal Deswal, Dept. of Seed Science and Technology, CCS Haryana Agricultural Faculty, HISAR, India, personal communication.
[128] This Act was amended in 1996 to include potatoes and Jerusalem artichokes.
[129] For example, CAMBIA's Patent Lens currently contains not only the all the US utility patents in the life sciences, but also the plant patents, on which we've recently contributed to statistical studies with the International Science and Technology Practice and Policy Center.
[130] Amendments were introduced to the Act of 1994 in 2002 aimed at enhancing the access of plant breeders to the PBR scheme and clarification as to what constitutes "breeding" for the purpose of the Act and the rights of breeders with regard to "essentially derived varieties". Amendments as of January 2003 have been compiled at http://www.upov.int/en/publications/npvlaws/australia/Australia_2003.pdf, but on the IP Australia website, apart from a link to the Expert Panel that evaluated the Act and

plant patents and plant breeding or variety rights.

- It would be desirable for IP Australia to provide a straightforward way to check what varieties may have double protection in Australia[131] through coordination of search facilities between the PBR section and the Patent section.

Biological Diversity and Access and Benefit Sharing

The area of what information should be provided on plant variety rights and how it should be used in intellectual property protection is quite dynamic. In connection with worldwide consideration of the requirements of the Convention on Biological Diversity[132] and the Bonn Guidelines[133], the International Treaty on Plant Genetic Resources[134], and the potential WIPO development agenda, some calls are being made for plant accession information to be made available in connection with patent applications that claim plants or plant-derived material, or even in applications that use plant genetic resources in examples.

- Each country may consider what information requirements it may make of applicants that would be added to its searchable databases, in advance of any international requirements to provide this information.

Other recommendations

For the benefit of the Australian public and particularly plant breeders, it would be desirable to have more readily locatable links to case law, and to scholarly work such as journal articles and other types of examining the trends in this area as they emerge.

- It would also be valuable for the individual listing of articles under "Resources" to contain direct links to the websites of the journals from which the references are cited, which in some cases allow the download of the full article under the emerging trend of open access to scholarly information, particularly that produced with public funding.
- Some case law on PVRs is currently provided at the UPOV website[135].
- Along the same lines, it would also be desirable to show statistics on such aspects of plant variety rights such as the number of applications per year, local vs. foreign applicants, etc., and perhaps even information on usage, as it may be available; a government could certainly request, perhaps optionally, more information on intended usage at the time of application and in other interactions with the applicant or rights-holder.

---

proposed the changes finally adopted, it is not straightforward to find explanation of the amendments and how they address these two issues.

[131] The 1978 UPOV Convention did not accept double protection but the 1991 Act lifted this ban. In the USA, 1994 legislation clarified double protection, and in Europe there has also been clarification of rights in the event of a conflict of interest between PBR holder and a patent holder arisen on the same variety. The situation is less clear in Australia and some guidance on the website may be useful to public stakeholders searching for these rights.

[132] Concluded at the 1992 UN Conference on Environment and Development in Rio de Janeiro and in force since 1993, it requires signatories, which comprise over 180 countries including Australia, to facilitate access to genetic resources in ways that promote their sustainable use.

[133] For implementation of the Convention on Biological Diversity, requiring national competent authorities to provide national focal points for information.

[134] Entered into force 29.7.04, setting up a multilateral system with requirements for Access and Benefits Sharing

[135] *e.g.* Grain Pool of WA vs. The Commonwealth of Australia & Anor (P34/98) is provided at http://www.upov.int/en/about/legal_resources/case_laws/pdf/au_gazette_89_12-2000.pdf