



UPOV/WG-DST/2/6

ORIGINAL: English

DATE: September 29, 2015

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS

Geneva

**WORKING GROUP FOR THE DEVELOPMENT OF A UPOV DENOMINATION
SIMILARITY SEARCH TOOL****Second Meeting
Geneva, June 9, 2015**

REPORT

*adopted by the Working Group for the Development of a UPOV Denomination Similarity Search Tool**Disclaimer: this document does not represent UPOV policies or guidance*Opening of the meeting

1. The Working Group for the Development of a UPOV Denomination Similarity Search Tool (WG-DST) held its second meeting in Geneva, Switzerland, on June 9, 2015, under the Chairmanship of the Vice Secretary-General of UPOV. The list of participants is reproduced in the Annex to this report.
2. The meeting was opened by the Chair, who welcomed the participants in Geneva and those who participated in the meeting by means of electronic conference.

Adoption of the agenda

3. The WG-DST adopted the draft agenda as reproduced in document UPOV/WG-DST/2/1.

Report of the WG-DST Test study and possible use of a UPOV denomination similarity search tool within UPOV*Test Study*

4. The WG-DST received a presentation by Mr. Glenn Mac Stravic, Head, Brand Database Section, Global Databases Service, on the intermediate results of the Test Study. A copy of the presentation is contained in document UPOV/WG-DST/2/4 "Intermediate Report of the WG-DST Test Study".
5. The Test Study identified the 11 best algorithms out of 15 which, after initial testing, performed better than all existing tools as measured by $F3$ ¹. The $F3$ measure had been chosen because recall was assumed to be of higher importance than precision.

¹ $F3 = (1 + 3 \times \text{precision}) \times (\text{precision} \times \text{recall}) / (3 \times \text{precision} + \text{recall})$

6. The selected algorithms had been used to form 2,047 different combinations, and each of the combinations had been tested with 100 different “cutoff values”² (total of 204,700 settings). The top 10 settings that showed the best performance among all the tested settings measured by F3 were shown in the presentation.
7. The WG-DST agreed on using the F3 measure as a suitable basis for assessing performance because of the importance of recall for denomination similarity search purposes.
8. The WG-DST agreed that the second step of the Test Study should be to consider whether to add an additional layer of phonetic criteria to the potential algorithms, by assessing if it would improve performance. In particular, it was agreed that it would be necessary to assess if precision would be sacrificed.
9. The WG-DST agreed to prepare two or three algorithms with a phonetic layer and to compare the results to the algorithms without a phonetic layer. It was further agreed that the second step of the Test Study should be done on the PLUTO database using a dedicated test site. The WG-DST proposed to invite CAJ members and observers to participate in the second step of the Test Study, subject to consultation with the Chair of the CAJ.
10. The WG-DST agreed that the set of algorithms for the test should be available by September 1, 2015 at the latest, and the result to be reported to the WG-DST, at its third meeting, to be held in Geneva, on October 2, 2015.
11. In response to a request from the Community Plant Variety Office of the European Union (CPVO) to test the set of algorithms on its own system, the WG-DST agreed that the source code of the set of algorithms written in JAVA would be made available, on request, to WG-DST members.

Non-acceptable terms for variety denomination

12. The WG-DST agreed to invite the CAJ to consider whether to develop a list of non-acceptable terms for variety denominations as an additional feature for the UPOV denomination search tool.
13. The WG-DST agreed that the list of non-acceptable terms could include, for example, botanical names. With regard to the inclusion of offensive terms, it agreed that it could be problematic to develop such a list.

Phonetic issues

14. The WG-DST considered document UPOV/WG-DST/2/3 “Phonetic issues”.
15. The WG-DST agreed that the consideration of phonetic elements in the UPOV denomination search tool, as set out in paragraphs 8 to 11 above, would be the most effective approach and further measures would not be appropriate.

Linguistic issues

16. The WG-DST considered the document UPOV/WG-DST/2/5 “Linguistic issues”.
17. The WG-DST agreed that it would be useful to provide the possibility to accept accents and special characters in denominations in the PLUTO database.

² The size of difference between a denomination and the test denomination calculated by an algorithm is called the similarity index. Cutoff value is the threshold of similarity index to determine whether the denomination is regarded as similar to the test denomination or not. For example, in the “similarity factor” algorithm at its normal setting, it only selects the denominations whose similarity index is lower than 1.2 as similar denominations. Thus the cutoff value is 1.2. For the similarity index of the other algorithms, a larger similarity index mean denomination is the more similar: denominations with a similarity index larger than the cutoff value are counted.

18. The WG-DST noted that, although the PLUTO database did not currently contain accents and special characters, it was possible for those elements to be included (see document CAJ/69/6 "Information Databases", Annex I "Program for Improvements to the Plant Variety Database", Section 3.1.3³).

19. The WG-DST agreed to propose to the CAJ to accept accents and special characters contained in denominations provided in the PLUTO database. However, it agreed that the algorithms should continue to use only the character set ASCII [American Standard Code for Information Interchange] representation, as defined in ISO [International Standards Organization] Standard 646, for denomination similarity search purposes, as currently used in PLUTO.

Revision of UPOV/INF/12 "Explanatory notes on variety denominations under the UPOV Convention"

20. The WG-DST considered document UPOV/WG-DST/2/2 "Revision of UPOV/INF/12 "Explanatory notes on variety denominations under the UPOV Convention".

(a) *Section 2.2.2 (b)*

21. The WG-DST agreed to amend the proposed text as follows (underlining indicates insertion to the text):

"(b) accepted market practices for particular variety types (e.g. hybrids) and particular genera/species (e.g. Medicago, Helianthus)."

(b) *Section 2.3.1 (c)*

22. The WG-DST agreed to propose new examples of denominations that:

(i) might erroneously imply relation to, or derivation from, other varieties;

Example: "alpha" and "alpha-HT" (in a case where "alpha-HT" differs from "alpha" in characteristics other than herbicide tolerance (HT) and is not closely related to Alpha), and

(ii) would not imply relation to, or derivation from, other varieties;

Examples: "Prince Albert" and "Prince Alexander", and "Ivory Bells" and "Lilac Bells" where the words "Prince" and "Bells" have been used for unrelated varieties and for varieties bred by different breeders.

(c) *Section 2.3.1 (d)*

23. The WG-DST agreed with the addition of new Section 2.3.1 (d) to document UPOV/INF/12.

24. The WG-DST agreed to add an explanation in Section 2.3.1 (d) that *Carex* is the botanical name of the genus, for which the common name is sedge.

25. The WG-DST agreed that the use of the botanical or common name of a genus to which a variety does not belong should be avoided, unless the botanical name or common name had a wider meaning, e.g. "Rose", "Cosmos", "Lilac", "Veronica" and "Bianca".

(d) *Section 2.3.3 (a)*

26. The WG-DST agreed the need to differentiate between letters in the form of words and other cases, and agreed on the following:

³ "3.1.3 Subject to Section 3.1.4, the character set for data shall be the ASCII [American Standard Code for Information Interchange] representation, as defined in ISO [International Standards Organization] Standard 646. Special characters, symbols or accents (', ^, ", °, etc.) are not accepted. Only characters of the English alphabet may be used."

(i) in the case of denominations consisting of letters not in the form of words, a difference of a single letter should be regarded as a clear difference, with examples to be provided;

(ii) in the case of denominations consisting of figures, a difference of a single figure should be regarded as a clear difference, with examples to be provided; and

(iii) in the case of denominations consisting of figures and letters, not in the form of words, one letter or figure difference should be regarded as clear difference, with examples to be provided.

27. The WG-DST agreed to consider Section 2.3.3 (a) further at its fourth meeting, on the basis of the conclusions of the CAJ at its seventy-second session, on October 26 and 27, 2015, on the WG-DST proposals.

(e) *Section 2.3.3 (b)*

28. The WG-DST agreed to add the example of “Helena” and “Elena” as an example of where a clear visual difference may not provide a clear phonetic difference in a language other than English (in Spanish).

29. The WG-DST recalled that acceptability of denominations would be considered in their original alphabet for the territory concerned and transcriptions and/or transliterations would not be considered for similarity purposes. On that basis the WG-DST agreed to propose the deletion of the example “Raion” and “Lion”, because they would not be confused in the Roman alphabet.

(f) *Section 2.3.3 (d)*

30. The WG-DST agreed to recommend that consideration be given to avoiding re-use of denominations in all cases. However, it clarified that it would be important to consider only denominations of varieties (i.e. plant groupings that meet the definition of variety in Article 1 (vi) of the 1991 Act) and, in addition, to expand the content of PLUTO database to include all recognized varieties, including these that had not been, or were no longer, registered/protected.

Date, place and program of the next meeting

31. The WG-DST agreed to hold its third meeting in Geneva, on October 2, 2015.

32. The WG-DST planned to discuss the following items during the third meeting:

1. Report of the WG-DST Test study and possible use of a UPOV denomination similarity search tool within UPOV (document to be prepared by the Office of the Union)
2. Non-acceptable terms (document to be prepared by the Office of the Union)
3. Content of the PLUTO database (document to be prepared by the Office of the Union)
4. Revision of UPOV/INF/12 “Explanatory notes on Variety Denomination under the UPOV Convention” (document to be prepared by the Office of the Union)
5. Date, place and program of the next meeting

33. *This report was adopted by correspondence.*

[Annex follows]

ANNEX

LIST OF PARTICIPANTS

I. MEMBERS

EUROPEAN UNION

Jean MAISON, Deputy Head, Technical Unit, Community Plant Variety Office (CPVO), CS10121, 49101 Angers CEDEX 02
(tel.: +33 2 4125 6435, fax: +33 2 41 256410, e-mail: maison@cpvo.europa.eu)
[via Webex]

Bénédicte LEGRAND (Ms.), Denomination expert, Community Plant Variety Office (CPVO), 3, bd Maréchal Foch, B.P. 10121, 49101 Angers
(tel.: + 33 (0)2.41.25.64.50, fax: + 33 (0)2.41.25.64.10, e-mail: legrand@cpvo.europa.eu)
[via Webex]

Carole BONNEAU (Ms.), Database Manager, Community Plant Variety Office (CPVO), 3 boulevard Maréchal Foch, BP 10121, 49101 Angers CEDEX 02
(tel: +33 2 41 25 64 50, fax: +33 2 41 256410, e-mail: bonneau@cpvo.europa.eu)
[via Webex]

Rudi CAES, Denomination expert, Community Plant Variety Office (CPVO), 3, bd Maréchal Foch, B.P. 10121, 49101 Angers
(tel.: + 33 (0)2.41.25.64.50, fax: + 33 (0)2.41.25.64.10, e-mail: caes@cpvo.europa.eu)
[via Webex]

FRANCE

Catherine MALATIER (Madame), Régisseur, Comité de la protection des obtentions végétales, Paris
(e-mail: catherine.malatier@geves.fr) [via WebEx]

JAPAN

Kenji NUMAGUCHI, Senior Examiner, Plant Variety Protection Office, New Business and Intellectual Property Division,, Food Industry Affairs Bureau, Ministry of Agriculture, Forestry and Fisheries (MAFF), Tokyo (e-mail: kenji_numaguchi@nm.maff.go.jp) [via WebEx]

II. OFFICE OF WIPO

Glenn MAC STRAVIC, Head, Brand Database Section, Global Databases Service

Lili CHEN (Ms.), Software Developer, Brand Database Section, Global Databases Service

III BUREAU

Peter BUTTON, Chair

IV. OFFICE OF UPOV

Peter BUTTON, Vice Secretary-General

Yolanda HUERTA (Ms.), Legal Counsel

Jun KOIDE, Technical/Regional Officer (Asia)

Leontino TAVEIRA, Technical/Regional Officer (Latin America, Caribbean countries)

Ariane BESSE (Ms.), Administrative Assistant

[End of Annex and of document]