E

UPOV

INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS
Geneva

WORKING GROUP FOR THE DEVELOPMENT OF A UPOV DENOMINATION SIMILARITY SEARCH TOOL

First Session
Geneva, September 3, 2014

THE CPVO SIMILARITY SEARCH TOOL

Document prepared by the European Union

Disclaimer: this document does not represent UPOV policies or guidance

The Annex to this document contains a copy of a presentation made by Mr. Jean Maison, Community Plant Variety Office of the European Union (CPVO), at the Working Group on Variety Denomination Search Tool (WG-DST), on the CPVO similarity search tool.

[Annex follows]

The CPVO similarity search tool

Prepared by experts from the European Union

UPOV WORKING GROUP FOR VARIETY DENOMINATION
SEARCH TOOL

Geneva, Switzerland, September 3, 2014

The CPVO similarity search tool

- Background
- Main features of the tool
- Conclusions

The CPVO similarity search tool

Background

- In 2004, development of centralized database on variety denominations
- Necessity for a similarity search tool and an automation of the search
- Development of the CPVO tool on the basis of a tool developed by GEVES

3

The CPVO similarity search tool

- Background
- Main features of the tool
- Conclusions

4

The CPVO similarity search tool

1. Processing the denomination tested

- Accent on letters removed
- Double letters reduced to a single letter (Hello → Helo)
- Every word of the denomination compared to the list of words excluded for testing (Red, Purple, etc ...) and removed accordingly
- In case of denominations consisting of several words, test of word combinations.

  ➢ Tänau Tari Yellow

  Words tested: Tanau, Tari, Tanautari, Tariyelow, Tanautariyelow

  ⇨ 5 words tested for this denomination

5

The CPVO similarity search tool

2. Building up a list of reference

- Basis: denominations of all varieties of the UPOV class
- All words with more than 3 letters or less than 3 letters compared to the words tested excluded
- Same processing for all words selected as for the denomination proposal
- ⇨ Building up a list of words of reference

6

The CPVO similarity search tool

3. The principles of the test

- Tests of all the words processed as described above
- Calculation of a similarity index for each pair of words tested – word in the list of reference
- Results ranked by similarity index displaying the denomination the word tested or the word in the list of reference belongs to

7

The CPVO similarity search tool

4. The similarity test itself

- Combination of simple statistical tests taking into account
- The frequency of letters in common
- Calculation of the percentage of common letters
- Calculation of the percentage of NON common letters
- Calculation of the percentage of difference of length.
- Calculation of rank Kendall correlation
- Thresholds eliminating denominations from the results at some stage
- Calculation of the similarity index adding up the results of each step

8

The CPVO similarity search tool

- Background
- Main features of the tool
- Conclusions

9

The CPVO similarity search tool

Conclusion

- Empirical test, build up by GEVES some 15 years ago
- Exclusively based on written appearance
- Need at some stage to introduce a simple phonetical feature: I = Y
  E.g. Bikini / Bykyny
- Some shortcomings

E.g. proposal 'GRANriver', GRAN company identifier.

⇒ 'River' not found, more than 3 letters difference

⇒ Experienced examiner will provoke an additional test for 'River'

10

Thank you for your attention

[End of Annex and of document]