|  |  |  |
| --- | --- | --- |
|  |  | E  UPOV/WG-DST/4/2  **ORIGINAL:**  English  DATE:  February 2, 2016 |
| INTERNATIONAL UNION FOR THE PROTECTION OF NEW VARIETIES OF PLANTS | | |
| Geneva | | |

Working group FOR THE DEVELOPMENT OF A UPOV denomination SIMILARITY search tool

Fourth Meeting  
Geneva, February 4, 2016

report of the wg-test study and possible use of a UPOV denomination  
similarity search tool Within UPOV

*prepared by the Office of the Union*

*Disclaimer: this document does not represent UPOV policies or guidance*

# Executive summary

The purpose of this document is to report the results of the second step of the Test Study in order to improve the algorithm for the UPOV denomination similarity search tool and to propose the next steps in the development of the UPOV denomination similarity search tool.

The WG-DST is invited to consider the proposed next steps for the development of the UPOV denomination similarity search tool, as set out in paragraph 21 to this document.

The structure of this document is as follows:

[Executive summary 1](#_Toc442114716)

[BACKGROUND 1](#_Toc442114717)

[DEVELOPMENTS SINCE THE THIRD MEETING of the wg-dst 2](#_Toc442114718)

[Proposed next steps 3](#_Toc442114719)

ANNEX I: Responses to the second step of the Test Study

ANNEX II: Examples for filters to remove duplicated types of similar denominations

ANNEX III: List of denominations that two or more contributors selected as possibly similar

ANNEX IV: List of denominations that none of one contributors selected as possibly similar

The following abbreviations are used in this document:

CAJ: Administrative and Legal Committee

CAJ-AG: Administrative and Legal Committee Advisory Group

TC: Technical Committee

WG-DST: Working Group for the Development of a UPOV Denomination Similarity Search Tool

# BACKGROUND

The background to this matter is provided in document UPOV/WG-DST/3/2 “Report of the WG-DST Test Study and possible use of a UPOV denomination similarity search tool within UPOV”.

The WG-DST, at its third meeting, held in Geneva on October 2, 2015, noted that the algorithms selected from the first step of the Test Study had produced too many denominations that were above the cut‑off value for similarity (low precision). The reason for the low precision was that the number of non‑similar denominations identified in the first step of the test study was too small and the types of non‑similar denominations too narrow in diversity.

The WG-DST, at its third meeting, agreed as a revised second step, in order to improve the precision of the selected algorithm(s), to run the selected algorithms on the PLUTO database with the 20 test denominations used in the first test study and to produce lists of the denominations above the cut-off value. Members of the Union would then be invited to identify denominations above the cut-off values that were not so similar that they would require further, individual consideration before deciding if they were (sufficiently) different from existing denominations (see Article 20(2) 0 -of the 1991 Act and Article 13(2) of the 1978 Act).

The WG-DST, at its third meeting, also received information from Mr. Glenn Mac Stravic that the best algorithms selected by recombination of standard search algorithms could be significantly improved by customization by a suitable expert. Given the long-term benefits of developing the most effective search tool possible, the WG‑DST agreed that the Office of the Union should seek to develop a customized algorithm as a further step. It noted that the customization would not be effective unless there was sufficient analysis by denomination experts in the revised second step of the Test Study.

The WG-DST, at its third meeting, agreed on the following timetable for the above exercise as follows:

(a) Issue a circular to invite CAJ designated persons of the members of the Union by October 6 inviting them to participate in the second step of the Test Study with a deadline of November 4;

(b) Produce a refined algorithm by December 2; and

(c) Produce a customized algorithm by December 31, 2015.

The WG-DST agreed that the feed-back by users on the customized algorithm should be considered at the fourth meeting of the WG-DST. In that regard, the Office of the Union would invite the members to test the customized algorithm when it became available.

# DEVELOPMENTS SINCE THE THIRD MEETING of the wg-dst

In accordance with the program agreed by the WG-DST, at its third meeting, the Office of the Union issued Circular E-15/237, on October 21, 2015, inviting members of the Union to participate in the second step of the Test Study for the development of an effective denomination similarity search tool. A web page for the Test Study, containing 20 Test Denominations and denominations for each of the Test Denominations selected by the best algorithm proposed at the first step of the Test Study was prepared for the work.

A summary of the responses to the second step of the Test Study is attached as Annex I to this document.  There was considerable diversity in the number of denominations that were selected as similar and further investigations revealed that there was not a very high coincidence in the denominations selected as similar.

In order to seek to develop a list of denominations that could be accepted as similar by the participating experts, a further exercise was proposed by Circular E-15/291, on December 21, 2015, to the WG‑DST and the respondents to the exercise of the second step of the Test Study.

For each test denomination, a list of denominations selected as similar by at least two experts in the second step of the Test Study was prepared.  Alongside that list, a further list of denominations that were only selected by one expert was presented.  The denominations that were not selected as similar by any expert in the second step were excluded.

In order to minimize the work required in this additional exercise, the lists of denominations was filtered to remove duplicated types of similar denominations. The method for removing duplicated types of similar denomination is explained in Annex II.

Participants were requested to indicate if they considered if any of the denominations that were previously selected by only one expert should be considered as similar by a UPOV denomination similarity search tool.

Annex III to this document presents the list of denominations that were selected by more than one expert as similar, and are, therefore proposed to be the basis for similar denominations in refining the algorithm.

Annex IV to this document contains the denominations that were not endorsed as similar by any experts in the additional exercise and, therefore, are proposed to be considered as non-similar denominations for the purpose of refining the algorithm.

All the contributors to the second step of the Test Study contributed to the additional exercise. In addition a further three experts contributed to the additional exercise.

The preliminary view of Mr. Glenn Mac Stravic on the data on Annex III and IV was that it would provide a suitable basis for refining the algorithm. However, it is not possible to predict how much improvement in performance of the algorithm would be achieved.

# Proposed next steps

The following steps are proposed for the development of a UPOV denomination similarity search tool;

(a) Refine the algorithm obtained from the first step of the Test Study using the data in Annexes III and IV of this document;

(b) If the performance of the refined algorithm is sufficient, an expert would be invited to customize the algorithm for further improvement. The customized algorithm would be presented for testing by members of the Union with data in the PLUTO database later in 2016. Specific dates would be proposed at the WG‑DEN, to be held in Geneva, on March 18, 2016; and

(c) If the performance of the refined algorithm is not sufficient, the deficiencies of the refined algorithm will be analyzed. A further exercise might be required according to the deficiencies of the refined algorithm. A proposal will be prepared for consideration at the WG-DEN meeting, to be held in Geneva, on March 18, 2016.

The WG-DST is invited to consider the proposed next steps for the development of the UPOV denomination similarity search tool, as set out in paragraph 21 to this document.

[Four Annexes follows

[see pdf or Excel version]

[Annex II follows]

Example 1: structure

*The test denomination is “FLAVORGIO”.*

The following denominations are in the same structure;

-       FLAVOR + XOXX (X=any of letters not used in the test denomination);

FLAVORHOME

FLAVORBOMB

FLAVORBOND

FLAVORZONE.. etc.

-       XOX + FLAVOR

JOYFLAVOR

TOPFLAVOR…etc.

Denominations with the same structure are regarded as the same group of denominations, and one denomination representing the group will be in the list of denominations:

FLAVORHOME (representing “FLAVOR + XOXX” structure)

JOYFLAVOR (representing “XOX + FLAVOR” structure)

Example 2: location and use of letters

*The test denomination is “FLAVORGIO”.*

The following denominations have letter “A” in their last three letters (“A” is included in the test denomination (FL**A**VORGIO)).

FLAVOR**A**SE

FLAVORC**A**N

FLAVORDN**A**

These denominations could be different if the algorithm considers the positions of letters; therefore, these denominations are treated as different groups.  These denominations will be in the list of denominations.

In the case of “FLAVOR**SET**”, none of the letters in the last three letters exists in the test denomination: this is regarded as a different group from the denomination “FLAVORA**SE**”, “FLAVOR**C**A**N**”and “FLAVOR**DN**A”.

The following denominations have “FLAVOR + Z + AVOR” structure (Z=any of letters): “L” is included in the test denomination (F**L**AVORGIO) but “S” is not; therefore, these are different groups.

FLAVOR**L**AVOR

FLAVOR**S**AVOR

*The test denomination is “KADET”.*

In the case of following denominations, “BANET” and “BODET” will be in the list.

BANET

BODET

BARET

CARET

“BANET” has “A” which is included in the test denomination (K**A**DET) but not “D”, and “BODET” has “D” but not “A”, and the location of these letters (“A” for “B**A**NET”: 2nd letter and “D” for BO**D**ET”: 3rd letter) is different.  Therefore, these denominations are regarded as different groups.

“BARET” is regarded as the same group as “BANET” because neither of the different letters (“R” for “BA**R**ET” and “N” for “BA**N**ET”) are included in the test denomination.

“CARET” is regarded as the same group as “BANET” because none of the different letters (“C” & “R” for “**C**A**R**ET” and “B” & “N” for “**B**A**N**ET”) are included in the test denomination.

[Annex III follows]

[see pdf or Excel version]

[Annex IV follows]

[see pdf or Excel version]

[End of Annex IV and of document]