



UPOV/WG-DST/2/3
 ORIGINAL: English
 DATE: May 26, 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 Session
Geneva, June 9, 2015

PHONETIC ISSUES

prepared by the Office of the Union

Disclaimer: this document does not represent UPOV policies or guidance

EXECUTIVE SUMMARY

1. The purpose of this document is to consider phonetic issues in relation to the development of a UPOV denomination similarity search tool.
2. The WG-DST is invited to consider whether additional phonetic aspects need to be addressed, beyond those included in the Test Study.
3. The structure of this document is as follows:

EXECUTIVE SUMMARY	1
PURPOSE	1
BACKGROUND	1
TEST STUDY	2
PROPOSAL ON THE APPROACH TO CONSIDER PHONETIC ISSUES	3

4. 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 Variety Denomination Search Tool

PURPOSE

5. The purpose of this document is to consider phonetic issues in relation to the development of a UPOV denomination similarity search tool.

BACKGROUND

6. At its first meeting, held in Geneva, Switzerland, on September 3, 2014, the WG-DST recalled that the CAJ had anticipated that the first step of the WG-DST would be to review the search types currently

available in the denomination search tab of the PLUTO database and to review search types in use in other situations that might provide an alternative basis for a UPOV similarity search tool.

7. The WG-DST, at its first meeting, agreed that the function of a similarity search tool should be to identify those denominations that were similar to existing denominations to the extent that they would require further, individual consideration before deciding if the denomination was (sufficiently) different from existing denominations (see Article 20(2) of the 1991 Act and Article 13(2) of the 1978 Act).

8. The WG-DST, at its first meeting, agreed that linguistic and phonetic issues should be considered further at the next meeting (see document WG-DST/1/4, paragraph 18).

TEST STUDY

9. The test study for the development of an effective denomination similarity search tool is explained in UPOV Circular E-15/045.

10. In relation to the phonetic issues, the following aspects of the test study are particularly relevant:

(a) Test denominations

A total of 20 test denominations was selected for the test study. Six of those (Aladin, Baptysta, Estel, Josefine, Kadét and Rugby) had phonetic aspects indicated as a reason for selection (see document WG-DST/2/STUDY/1 "Test denominations for the study, selected from proposed test denominations", reproduced as Annex to this document).

(b) The search tools that were used as a basis for the development of a UPOV denomination similarity search tool include:

(i) Phonetic

This searches for denominations that contain words that sound similar to the terms entered. The following quotes from Wikipedia¹ explain the concept of the algorithm "Metaphone" and "Double Metaphone", and the latter is used in Phonetic:

"Original Metaphone codes use the 16 consonant symbols 0BFHJKLMNPRSTWXY. The '0' represents "th" (as an ASCII approximation of Θ), 'X' represents "sh" or "ch", and the others represent their usual English pronunciations. The vowels AEIOU are also used, but only at the beginning of the code. This table summarizes most of the rules in the original implementation:

1. Drop duplicate adjacent letters, except for C.
2. If the word begins with 'KN', 'GN', 'PN', 'AE', 'WR', drop the first letter.
3. Drop 'B' if after 'M' at the end of the word.
4. 'C' transforms to 'X' if followed by 'IA' or 'H' (unless in latter case, it is part of '-SCH-', in which case it transforms to 'K'). 'C' transforms to 'S' if followed by 'I', 'E', or 'Y'. Otherwise, 'C' transforms to 'K'.
5. 'D' transforms to 'J' if followed by 'GE', 'GY', or 'GI'. Otherwise, 'D' transforms to 'T'.
6. Drop 'G' if followed by 'H' and 'H' is not at the end or before a vowel. Drop 'G' if followed by 'N' or 'NED' and is at the end.
7. 'G' transforms to 'J' if before 'I', 'E', or 'Y', and it is not in 'GG'. Otherwise, 'G' transforms to 'K'.
8. Drop 'H' if after vowel and not before a vowel.
9. 'CK' transforms to 'K'.
10. 'PH' transforms to 'F'.
11. 'Q' transforms to 'K'.
12. 'S' transforms to 'X' if followed by 'H', 'IO', or 'IA'.
13. 'T' transforms to 'X' if followed by 'IA' or 'IO'. 'TH' transforms to '0'. Drop 'T' if followed by 'CH'.
14. 'V' transforms to 'F'.
15. 'WH' transforms to 'W' if at the beginning. Drop 'W' if not followed by a vowel.

¹ Wikipedia contributors, "Metaphone", *Wikipedia, The Free Encyclopedia*, 11 May 2015, 15:59 UTC, <<http://en.wikipedia.org/w/index.php?title=Metaphone&oldid=661858533>> [accessed 26 May 2015]

16. 'X' transforms to 'S' if at the beginning. Otherwise, 'X' transforms to 'KS'.
17. Drop 'Y' if not followed by a vowel.
18. 'Z' transforms to 'S'.
19. Drop all vowels unless it is the beginning.”

“The Double Metaphone phonetic encoding algorithm is the second generation of this algorithm.”

“Double Metaphone tries to account for myriad irregularities in English of Slavic, Germanic, Celtic, Greek, French, Italian, Spanish, Chinese, and other origin. Thus it uses a much more complex ruleset for coding than its predecessor; for example, it tests for approximately 100 different contexts of the use of the letter C alone.”

(ii) Similarity factor

This performs an analysis of the denomination on the basis of a combination of factors including letters in common, relative lengths of the words and positions of the common letters

(iii) Fuzzy

This will search for denominations that contain words spelled one or two characters differently from the terms entered.

(c) List of similar denominations

Participants of the study were requested to submit a list of similar denominations which may include phonetic aspects (e.g. “Lugby” was submitted as a similar denomination of “Rugby”).

MATTERS FOR CONSIDERATION

11. The members of the WG-DST may wish to consider whether further consideration of phonetic issues would be required in the development of a UPOV denomination similarity search tool, beyond those included in the Test Study.

12. The WG-DST is invited to consider whether additional phonetic aspects need to be addressed beyond those included in the Test Study.

[Annex follows]

ANNEX

TEST DENOMINATIONS FOR THE STUDY, SELECTED FROM PROPOSED TEST DENOMINATIONS

The following list of denominations are proposed as test denominations for the study. They were chosen to provide a range of different types and patterns of denominations (e.g. with/without figures, concatenated words, consist of multiple words).

Proposer	Proposed test denominations	Explanation provided by proposer
QZ	Aladin	has several possible similar denominations (e.g. 'Alladin', 'Aladdin', 'Alladdin', 'Yellow Alladin', etc.)
QZ	Baptysta	similar to 'Batista' (phonetic)
AU	Chara	similar to "Clara", "Jhara". Existing tool cannot detect similar denomination "Jura".
QZ	Codisolar	similar to "Solar", "Solal", "Sonar", "Solara". Company identifier attached
AU	Divva	similar to "Diva"
NL	Estel	not enough phonetical difference with the existing denomination "Estelle".
NL	Flavorgio	it causes confusion with the denomination "Flavorglo".
NL	Josefine	confusion with the denomination "Josephine". Although two characters difference, there is not enough phonetical difference.
QZ	Kadét	similar to 'Cadete' (phonetic)
AU	M67	similar to "M76", "M676", "M767"
AU	NN-9812AE	similar to "NN-9812AA"
AU	Pirate	similar to 'Partie'
QZ	Rugby	similar to "Ragbi", "Rugy", "Rubis", "Ruby red", "Rouby Rubino" (visual and phonetic)
AU	Sienna	similar to "Vienna"
QZ	Snowplanet	concatinated words, similar to "Red Planet"
AU	Sweet Caroline Sweet Heart Purple	Acceptable long denomination, possible similar denominations: Sweet Caroline Sweetheart Purple, Sweet Caroline Sweet Heart Red, Sweet Caroline Bewitched Purple
NL	Topaz	confusion with the denomination "Topas".
NL	Victoire	confusion with the denomination "Victory".
QZ	Waverider	concatinated words, similar to "New Wave"
QZ	Younique	similar to "Unique", and has company identifier

[End of document]