



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

1. 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.
2. 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.
3. The structure of this document is as follows:

EXECUTIVE SUMMARY	1
BACKGROUND	1
DEVELOPMENTS SINCE THE THIRD MEETING OF THE WG-DST.....	2
PROPOSED NEXT STEPS	3

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	

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 the Development of a UPOV Denomination Similarity Search Tool

BACKGROUND

5. 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".
6. 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.

7. 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) of the 1991 Act and Article 13(2) of the 1978 Act).

8. 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.

9. 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.

10. 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

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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

21. 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.

22. 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

UPOV/WG-DST/4/2

ANNEX I

SUMMARY OF RESPONSES TO THE SECOND STEP OF THE TEST STUDY

Name	ALADIN	Baptyst	Chara	Codisol	Divva	Estel	Flavorg	Josefine	Kadet	M67	NN-981	Pirate	Rugby	Sienna	Snowpl	Sweet C	Topaz	Victoire	Waverid	Youniqu
Contributor 1	8																			
Contributor 2	62	189	154	271	206	199	417	151	232	67	4	58	698	45	689	37	241	334	328	135
Contributor 3	32	28	45	38	49	76												53		36
Contributor 4	9	6	17	21	35	56	9	82	30		2	15	51	31	13	6	49	94	8	65
Contributor 5	45	11	92	160	210	116	237	39	75	60	1	51	579	33	522	10	179	22	250	35
Contributor 6	44	10	54	25	76	107	10	77	28	9	2	30	265	36	4	7	137	18	16	15
Contributor 7	3	4	2	3	6	9	3	7	4		1	4	3	3	2	2	5	2	2	4
Contributor 8	39	13	112	28	77	61	65	32	63	2	6									
Contributor 9	8	5	7	5	6	6	1	5	1		1	2	2	3	2	2	3	2	7	2
Contributor 10	28	51	41	44	69	3													1	
Contributor 11	8	5	3	4	7	5	3	6	2			8	5	5	3	3	5	5	6	4
Contributor 12	7	2	1	8	9	13	4	6	6		2	2	5	7	2	2	8	3	9	18
Contributor 13	2	3	3	6	6	7	2	20	7	48	2	12	8	11	9	26	22	39	25	16
Contributor 14	15	81	6	7	10	17	3	23	17	1	3	10	10							
Contributor 15	27	14		19	28	10	5	12	16	3	1	6	11	10	4	7	20	7	8	27
Contributor 16	2																		1	
Contributor 17	32				49								18						58	
Contributor 18	7	7	12	11	30	28	10	22	23		1									1
Contributor 19	18																			
Contributor 20	31	24	68	34	38															
Contributor 21	29						12	28	30		1	26								
Contributor 22								16						6				30	25	
Contributor 23											1									
Contributor 24	34												760	30	6	91	199	28	27	104
Contributor 25												3	4	4	3	2	3	4	3	5
Contributor 26																				1
Contributor 27																			17	8
No. of Contributors	22	16	15	16	17	15	14	15	14	7	14	13	14	13	12	12	17	14	14	15
"Possibly similar" denominations per contribution	22	28	41	43	54	48	56	35	38	27	2	17	173	17	105	16	58	45	51	32

Total contributors 27
 Contributions per Test Denomination 14.5

[Annex II follows]

ANNEX II

EXAMPLES FOR FILTERS TO REMOVE DUPLICATED TYPES OF SIMILAR DENOMINATIONS

Example 1: structure

The test denomination is "FLAVORGIO".

The following denominations are in the same structure;

- FLAVOR + XOOX (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 + XOOX" 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 (FLAVORGIO)).

FLAVORASE
 FLAVORCAN
 FLAVORDA

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 "FLAVORSET", none of the letters in the last three letters exists in the test denomination: this is regarded as a different group from the denomination "FLAVORASE", "FLAVORCAN" and "FLAVORDA".

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

FLAVORLAVOR
 FLAVORSAVOR

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 (KADET) but not "D", and "BODET" has "D" but not "A", and the location of these letters ("A" for "BANET": 2nd letter and "D" for BODET": 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 "BARET" and "N" for "BANET") are included in the test denomination.

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

[Annex III follows]

Cells with yellow highlights were selected by contributors in the additional exercise.

ALADIN	Baocvia	Chara	Codoclar	Divva	Estal	Favonab	Jokelina	Kadet	M67	NI/9612AE	Prata	Rusdy	Sienna	Snowberry L	Touzaz	Victoire	Waveridge	Younissa	Sweet Caroline Sweet Heart Purple
						FLAVORJO FLAVORJO								SNOWBELT 1 SNOW PEARL SNOW PEOPLE			WAVETREND WAVETRADER WAVETRAN		
						SUPERPRETZEL FLAVOR DUOS THE FLAVOR GUY WE FLAVOR THE WORLD WORLDS OF FLAVOR ASIA								SNOW PHANTOM SNOW PHONE SNOW PILOT Snow Plus SNOW PLATE SNOW PLOWERS Snow Plus SNOW PRINCE SNOW PRINCESS SNOW PULSE Snow Purple SNOW RIFPLE Snow River Snow Rose SNOW SANGO SNOW SATIN SNOW SPINNER SNOW STATION Snow Sunny Peak Snow Supreme SNOW SVIAT SNOW VALLEY SNOW VIEW SNOW VILLAGE SNOW WALKERS SNOWAPPLE SNOWBALL PRESS Snowball SNOWBEL 124 SNOWBLAZE SNOWBLAST SNOWBLASTER 2000 SNOWBLAZER SNOWBLIND Snowboard Snowboard Snowbird Snowbite Snowbalm Snowblossom SNOWCLEAN SNOWCLEAN E MIX SNOWCLEANES Snowcone Snowcone White Snowcrist SNOWCROCK Snowden Snowdon Snowdrift Snowdrift SNOWFLAME Snowfall Snowfla Snowflake SNOWFLAKES SNOWFLANE SNOWFLITES SNOWFLITE O Snowflora SNOWFLUENT Snowfly Snowfury Snowfury SNOWFLURRY Snowflake SNOWFLANDER Snowflake SNOWFLUTE DESIGN SNOWFLUTE BRAND SNOWMAPLE Snowmass SNOWMASTER Snowmass SNOWPACT SNOWPARADE SNOWPARROT Snowpeak SNOWPEAK SNOW PEAK SNOWPEN AIR SNOWPEN AIR SNOWPETAL SNOWPETALS SNOWPLANET SNOWPLAY NOBY BEAR SNOWPOINT SNOWPOLE SNOWPRANET SNOWPRICTS SNOWPROTEC SNOWQUEEN SNOWRENTAL SNOWSCENTS SNOWSHAKE DERBY SNOWSHAKER Snowsora Susame SNOWSPACE Snowsora SNOWSPRITE Snowstar Snowstar SNOWY FRONT 10th PLANET INTERNATIONAL SOUND PLANET SNOWPLANNER SNOWPLANET SODD PLANET SP SONIC PLANET SPARKLE SNOWLAND SPINPLANET STONE PLANET SUNNY PLANET EYEWEAR SUNPLANET SOLARJAMKESKUS SWIFPLANET SWISS SNOW LEAGUE TENTH PLANET PRODUCTIONS THE SNOW CONNECT THERE SNOW PLACE LIKE TOON PLANET PARK Tubus Snowsora TY SNOW FREEZE TYEPLANET VIRGIN SNOW APPAREL VIRGIN PLANET WINDOWPLANET WINDPLANET WOM PLANET WOOD PLANET WRONG PLANET VILETIDE SNOW HAULET		WAVETRENDS WAVETRICE WAVEVIEW WAVETRICE WAVEWISPER WAVEWISPER WEB RIDERZ WEVERAKER WEVERAKERE WEVERAKER WEVERAKER WEVERIDER WEVERIDER WHEEL RIDER WHEEL RIDER			

LIST OF DENOMINATIONS THAT NONE OR ONE CONTRIBUTORS SELECTED AS POSSIBLY SIMILAR

ALADIN	Bahaysta	Chara	Codisolar	Divina	Estel	Flavorito	Josefina	Kadet	M67	NN-2612AE	Pirata	Rugby	Sienna	Snowplanet	Topaz	Victoire	Waverider	Younique	Sweet Caroline	Sweet Heart Purple
Arlekin	BAPTISCARE	SUPER CHARRO	NO COST SOLAR SYSTEM	Dvina	SEETEL	DO ME A FLAVOR	JODEFINA INTA	KAFETTE	Mk 267 S	TANROSILB (NAAS BOTH)	E-Z ASPHATE	EARUGBY EUCLIDE AVRIDOR	Iena	NEW PLANET AGRICULTURE STAPLES	TO AZ PILOT SHOP	VILLA VICTORIA	Alder	COMUNIQUE	Sweet Jumbo	
Adair	Baptistine	IHARA MILBEX	CODISMARK	Codival	EYESTIL	FLAVORED TO THE BONE	SEFINE		M-4906		Pyrit			RINOPLANETT	Tabiz	VICTORIA'S GARDEN	ADVERRIDERS	BELUNIQUE	SWEETE	
Avdlin 2002	Bailista	Chari	CODISLOR	Vivia	Etmor	FLAVORS YOU CRAVE	HGC Josef Lember		M 178-213		PARITA			FASHION BODY SOUL PLANET	DPAZI'6'	VICTRIO	Waverie	BICUNIQUE	SWEETS	
		Shara	CODICOSA	Valva	Apotiel	FLAVO DROP	Josie		51 M					SNOWS LAKE WINERY	Topazio RR	KH Viktor	WABERO	IMMUNIQUE	SWEET	
		SCHIRA	CODIDOXAL	Aviva	Fastel	FLAVOCORM	Yoshine		MP3					MADE IN OUR PLANET	Voraz	Victoria Red Gabty	WAVE DRIVEN MARKETING SERVICES	SOIE-ET-TOI BEIYOUINIQUE	SWEETA	
		SHARAI	CODISL	Kviva	ES Talea				M 115-153					APPROCHE DE ANNA PLANET	LOPAZO	Victoria Emma	WAVERSAVER	EUNIQUE'S PIZZA EXPRESS 'EAT TO A HEALTHIER BEAT'	SWEET Y	
		Cherie	CODIDOL	Delo	Ves Yel				90M02					RESCUE FROM PLANET ZERO	Torcac	Wavetop		Yoshine	Sweeto	
		CH-ARRAYAN	CODISIL	Dufy	TESVEL				M-38					SWATCH SNOWPASS SNOW PASS		VIKTOR S	Wave King Akabono	Yoannika	T-SWEET	
		T-CHAKRA	CODICOL	Tuva	PASTELLE				M 48-42					SNOW CLOSE TO HOME		VICTORY-J	Wavetopus	Winica	SWEED	
		CHELA	ALMAN CHOUR I SOLAR		Pensel									THIS IS YOUR PLANET		VIKUTOARU	COOL WAVE SLIDE		Swi	
		Tschaka	CODILAC											GENUINE CANADIAN SNOW SNAKES		VICTORINO	DIGITAL WAVEGUIDE TOUCH		Q SWEET	
		Jouhara	CODIDANLO											DOWN TO PLANET EARTH		VINKOR	BLUEWAVE WATER IDEAS		A SWEET	
		Thira (OUD)	CODIPLAY											S3 SNOW SKATE SURF		VIKUTOWARU	C.E.O OVERRIDERS		Sweet 'n Pink	
			MEDISOLARE											SNOWPLICITY SNOWFLOWING MADE SIMPLE		Victoriana	Lowly River		SWEETN LOW	
			CODISCO											ODYSSEY SNOW AND SEA		Victorious	Blue Waves		AU-Sweet Scarlet	
			CODILSA											SLED DOGS SNOWSKATES		VICTORINJUS	HEAT WAVE SUPER QUIET		Hearty	
			CODIPRAL											THE ULTIMATE SNOWSKATE		VIKTORINA	RACERIDE		CAROLENA TSUBAKIHIME	
			CODICOSOL											ONCE UPON A PLANET		NKVICTORY	WAVEBIRD WAVE BIRD		Sweet-Creek	
			CODISCOPE											DJ BOBO PLANET COLORS		Tar-Victoria	WAVEGLIDER.NET		DAR 1-21-27 PURPLE	
			CODISCOPE											SEAMASTER HOUR PLANET		Barvictor	W WAVE WEED CONTROL		SKREET	
			CODIPLASH											PNO PLANETOUT		JAVTRADER	WATERWEAR		Sweet-N-Sterile	
			CODIDIXAL											PUREST ON THE PLANET			WATERWEAR		SWEET-EEZ	
			CODIFLORUM CL											MISS NUDE PLANET EARTH			HEAT WAVE PADDLE BALL		DEAR PURPLE	
			CODIDANLO											SNOW TRAIN TO JASPER		NWA 255 RIDERS	NWA 255 RIDERS		SUITO KYARORIN SUITO HATO PAPURU	
			CODIVISUAL											SNOWBALL THE DANCING COCKATOO		Summer Wave Violet	Summer Wave Violet		JACK OFFHEARTS	
			CODIAGRO											THE JELLY BEAN PLANET			COOL WAVE GOLDEN YELLOW		Karolina	
			CODISAG											SONGPLANET.COM			EXTREME WAVE FRUIT ROLLS		Sweetcorn Honey Bantam 83 Gun	
			Codisp											Plunus persica Snow Lady Rose			SUMMER WAVE PALE LILAC		Hurt	
			Codisp											SNOW LAND TALL BEAUTY			DOUBLE WAVE MISTY LILAC		Holland Purple	
			Wander											SNOW DELIGHT (OZWHITEDELIGHTTHREE)			WAVERS		Charm Purple	
			Codibook											SNOSWEET			Waverly		Calodine	
			Codnan											SNOWBALL PALL DI NEEVE			DOUBLE WAVE BLUE VEIN		Cacina	
			Codnan											SNOWCOURT SHRASAGI			COOL WAVE PURPLE		Impact Purple	
			Codiciap											SNOW SHOWER WHITE GEM			COOL WAVE FROST		Heartbeat	
			Codisp											SNOW CAP LOVE SONG			WONDERITER		Heartthrob	
			Codisab											ENERGY STAR GREEN PLANET			COOL WAVE YELLOW		Heartlet	
			Codifu											Super SnowBall UG			SUMMER WAVE BOUQUET BLU		Bee Purple	
			Codwer											Deknow Flake					BICOLOR NA PURPLE MOMOCHAN	
			Codlum											DEKSNOWPEAK					DW AG PURPLE 2 LILAC BEAUTY)	
			Codimel																SENHEART SUN COLOR	
			Codlux																Melato Purple	
			Codibal																CHHAYA PURPLE	
			Codipad																SECRET PURPLE	
			Codipus																IT PURPLE	
			Codipum																REDHEART	
			Codimono																LADY PURPLE	
			Codibim																FISTANEON (TANGO NEON PURPLE)	
			Codisea																Jewel Purple	
			Codiflex																GARDEN VENUS PURPLE	
			Codiboum																DAICHI BLUETTE FRESH PURPLE RY	
			Codimica																TEMARI PURPLE	
			KORISULAS																Hana Purple	
			Codese																PETINA PURPLE	
			CODIAPF																Butbanks Late Heartthrob	
			Codirush																Dekmacaron Purple	
			Codix																Supertunie Dark Purple	
			Codisank																Fort Purple Ven	
			Codibird																SPLASH PURPLE	
			Codigreen																MOHAWK (PURPLE PIXIE)	
			Codirect																PLAGU PURPLE	
			Codisack																Redsiva outole	
			Codifert																FURORETTO PURPLE	
			Codisne																Dantohest	
			Codimust																Heartbreaker	
			CODIAWIM																Ovation Purple	
			Codisaki																Deksampurple	
			Coditour																Vamp Time Purple	
			Codisur																Eleven Purple	
			Codisar																NAGANO PURPLE	
			Codireal																	
			Codispot																	
			Codiso																	
			Codivox CL																	
			Coditoul																	
			Codiflo																	
			Codikart																	
			Codilate																	
			Codisurf																	
			CODIMARK																	
			CODIFLOU																	
			Codikom																	
			Codival																	
			CODIARED																	
			Codimal																	
			Codisure																	
			Codipro																	
			Codigran																	
			Codirap																	
			Codiroc																	