



CAJ/68/9

ORIGINAL: Inglés

FECHA: 2 de septiembre de 2013

UNIÓN INTERNACIONAL PARA LA PROTECCIÓN DE LAS OBTENCIONES VEGETALES

Ginebra

COMITÉ ADMINISTRATIVO Y JURÍDICO

**Sexagésima octava sesión
Ginebra, 21 de octubre de 2013**

POSIBLE ELABORACIÓN DE UN INSTRUMENTO DE LA UPOV DE BÚSQUDA DE SIMILITUD A LOS FINES DE LA DENOMINACIÓN DE VARIEDADES

Documento elaborado por la Oficina de la Unión

*Descargo de responsabilidad: el presente documento no constituye
un documento de política u orientación de la UPOV*

1. El Comité Administrativo y Jurídico (CAJ), en su sexagésima séptima sesión, celebrada en Ginebra el 21 de marzo de 2013, asistió a una ponencia presentada por la Delegación de la Unión Europea sobre la experiencia de la Oficina Comunitaria de Variedades Vegetales (OCVV) en la utilización de su instrumento de búsqueda de denominaciones similares en el examen de las propuestas de denominación de variedades. Durante la presentación, la OCVV, propuso estudiar la posibilidad de elaborar un instrumento de la UPOV de búsqueda de similitud a los fines de la denominación de variedades, sobre la base del instrumento de búsqueda de la OCVV.¹ El CAJ acogió con agrado el ofrecimiento de la OCVV y acordó incluir un punto dedicado a considerar esa propuesta en su sexagésima octava sesión, que se iba a celebrar en Ginebra el 21 de octubre de 2013 (véase el documento CAJ/67/14 "Informe sobre las conclusiones", párrafos 49 y 50).

ANTECEDENTES

2. Actualmente la pestaña de búsqueda por denominación de la base de datos sobre variedades vegetales (base de datos PLUTO) (<https://www3.wipo.int/pluto/user/es/index.jsp>) proporciona las siguientes posibilidades de búsqueda para encontrar denominaciones similares:

- | | |
|---|---|
| a) Factor de similitud
[Instrumento de
búsqueda de la OCVV] | Se efectúa un análisis de la denominación que se haya introducido sobre la base de una combinación de factores, incluidas las letras en común, la longitud relativa de las palabras y la ubicación de las letras en común. Se trata del método de comparación más complejo y puede que la búsqueda tarde varios segundos en realizarse. El factor de similitud fue ideado por el Grupo de Estudio y Control de Variedades y Semillas (GEVES) (Francia) en colaboración con la Oficina Comunitaria de Variedades Vegetales (OCVV) de la Unión Europea. No obstante, debe tenerse en cuenta que los resultados de la búsqueda efectuada con arreglo al factor de similitud en la base de datos PLUTO deben ser interpretados y no constituyen una garantía en cuanto a la idoneidad de las denominaciones de variedades, que debe ser decidida por las autoridades encargadas de la concesión de derechos de obtentor en el territorio de que se trate. |
|---|---|

En el Anexo del presente documento se facilita una explicación pormenorizada del análisis.

¹ La elaboración del factor de similitud estuvo a cargo del Grupo de Estudio y Control de Variedades y Semillas (GEVES) de Francia, en colaboración con la Oficina Comunitaria de Variedades Vegetales (OCVV) de la Unión Europea.

- b) Búsqueda aproximada [fuzzy] Se buscan denominaciones que contengan palabras con uno o dos caracteres diferentes de los términos introducidos. Se parece al método de correspondencia aproximada de la pestaña de "Búsqueda de términos".
- c) Búsqueda fonética [phonetic] Se buscan denominaciones que contengan palabras que suenen de forma similar a los términos introducidos. Se parece al método de correspondencia fonética de la pestaña de "Búsqueda de términos".
- d) Contiene [contains] Se buscan denominaciones que contengan palabras que a su vez contengan la misma serie de letras que los términos que se hayan introducido. Se parece al método de correspondencia "Contiene" de la pestaña de "Búsqueda de términos".
- e) Empieza por [starts] Se buscan denominaciones que contengan palabras que empiecen con la misma serie de letras que los términos introducidos. Se parece al método de búsqueda de correspondencia "Comienza por" de la pestaña de "Búsqueda de términos".
- f) Termina por [ends] Se buscan denominaciones que contengan palabras que finalicen con la misma serie de letras que los términos introducidos. Se parece al método de búsqueda de correspondencia "Termina por" de la pestaña de "Búsqueda de términos".

3. Durante las conversaciones mantenidas con la Oficina de la Unión a fin de examinar la manera de elaborar un instrumento de la UPOV de búsqueda de similitud a los fines de la denominación de variedades, la OCVV aclaró que debían contemplarse todas las opciones y que, teniendo en cuenta los avances en la tecnología de la información, era posible que el mejor instrumento no se basara necesariamente en el de la OCVV. El principal factor a tener en cuenta debía ser elaborar un instrumento que pudiesen utilizar todos los miembros de la UPOV, a fin de minimizar las diferencias en las decisiones sobre la idoneidad.

PROPUESTA

4. Teniendo en cuenta la necesidad de disponer de conocimientos sobre el examen de la denominación de variedades y los instrumentos informáticos, se propone crear un pequeño grupo de trabajo para elaborar propuestas de un instrumento de la UPOV de búsqueda de denominaciones similares a los fines de la denominación de variedades. Dicho grupo podría estar integrado por:

- a) examinadores de denominaciones que sean miembros de la Unión (3 a 6 expertos);
- b) el Servicio de Bases de Datos Mundiales de la OMPI (responsable de la base de datos PLUTO);
- c) la Oficina Comunitaria de Variedades Vegetales (OCVV) de la Unión Europea; y
- d) la Oficina de la Unión.

5. Se prevé que el propio grupo establezca su plan de trabajo; sin embargo, cabe esperar que su primer paso sea examinar los tipos de búsqueda existentes actualmente en la pestaña de Búsqueda por denominación de la base de datos PLUTO, en especial el factor de similitud (instrumento de búsqueda de la OCVV), y analizar los tipos de búsqueda utilizados en otras situaciones (p. ej. en relación con marcas) que podrían proporcionar una base alternativa para el instrumento de búsqueda de la UPOV.

6. En el análisis de la idoneidad de los tipos de búsqueda se debería tener en cuenta especialmente el documento UPOV/INF/12 "Notas explicativas sobre las denominaciones de variedades con arreglo al Convenio de la UPOV". A ese respecto, el grupo de trabajo tendrá que solicitar al CAJ orientación adicional si su labor indicara que es necesario el análisis del documento UPOV/INF/12 para elaborar un instrumento eficaz de búsqueda de similitud para la UPOV.

7. Se propone que las reuniones del grupo de trabajo se celebren en la Oficina de la Unión en Ginebra y que dicha Oficina las presida. Se dispondrá que las reuniones no coincidan con las sesiones de la UPOV y se prevé la participación electrónica de los examinadores de denominación y de la OCVV. Es de esperar que las propuestas elaboradas por el grupo de trabajo se presenten al CAJ y al Comité Técnico (TC) y que ambos comités reciban un informe breve de dichas reuniones.

8. Se invita al CAJ a:

a) *aprobar la creación de un grupo de trabajo que elabore propuestas para un instrumento de la UPOV de búsqueda de similitud a los fines de la denominación de variedades, de conformidad con lo dispuesto en los párrafos 4 a 7 del presente documento; y*

b) *recomendar que los miembros de la Unión propongan examinadores de denominación con experiencia que resulten adecuados para participar en el grupo de trabajo.*

[Sigue el Anexo]

UNIÓN EUROPEA
Oficina Comunitaria de Variedades Vegetales

PROCEDIMIENTO DE BÚSQUEDA

1. GENERAL

Como conclusión de la fase de estudio del proyecto presentado a su Comisión de Auditoría en noviembre de 2003, la Oficina propuso, en primera instancia, utilizar programas informáticos nacionales, adaptarlos a la base de datos centralizada y ejecutarlos todos para cada comprobación. Se estima que, de esta manera, la posibilidad de no dar con una denominación similar es muy escasa.

A más largo plazo, se prevé desarrollar un programa informático de la OCVV, y la posibilidad de ampliar las funciones lingüísticas.

En la práctica, se ha partido de las especificaciones del programa informático francés para llevar a cabo el procedimiento de búsqueda en la base de datos de la OCVV.

2. NORMAS PARA DETERMINAR UNA DISTINCIÓN SUFICIENTE ENTRE 2 DENOMINACIONES DE VARIEDADES

Según el Reglamento de base relativo a la protección comunitaria de las obtenciones vegetales, una de las normas que debe cumplir la denominación de una variedad es que no coincida o pueda confundirse con una denominación de variedad bajo la cual figure en un registro oficial otra variedad de la misma especie o estrechamente relacionada con ella.

Esta norma se ha interpretado en las directrices del Consejo de Administración de la OCVV relativas a las denominaciones de variedades:

- ♣ Por lo general, una diferencia de tan solo una letra o un número o de una tilde en una letra debe considerarse como susceptible de prestarse a confusión.
- ♣ En general no debe considerarse que diferencias de dos o más letras inducen a confusión, excepto cuando simplemente se yuxtaponen las mismas letras.
- ♣ Además, una denominación de variedad no debe transmitir la falsa impresión de que la variedad está relacionada con otra o se deriva de ella;

El propósito del programa de búsqueda es descubrir denominaciones de la misma clase en la base de datos que pudieran estar en conflicto con una denominación propuesta.

3. PROCEDIMIENTO DE BÚSQUEDA

Las comprobaciones se realizaron mediante un programa interno de la base de datos ORACLE en el servidor de la OCVV (mayor rendimiento).

Para llevar a cabo una comprobación, el programa de interfaz (sitio web, ...) ejecuta un proceso denominado:

TESTDENOMINATION (comprobar denominación).

Como parámetros, el programa de interfaz transmite la denominación a comprobar y el código de la especie a la que pertenece la variedad.

Al finalizar el proceso, se obtiene un identificador de la comprobación efectuada en la interfaz (columna Testid). Con este identificador es posible leer los 2 cuadros que constituyen el resultado de la comprobación: los cuadros TESTS (comprobaciones) y TESTRESULTS (resultado de las comprobaciones).

En el cuadro TESTS figuran los datos generales sobre la comprobación: la fecha de la comprobación, el identificador de la persona que solicitó la comprobación, la denominación, el código de la especie, el código de la clase o el género, el tiempo de ejecución informática, las palabras excluidas, los mensajes de error,...

En el cuadro TESTRESULTS se presentan las listas de las denominaciones similares que se han encontrado mediante el proceso TESTDENOMINATION. Cada denominación similar está acompañada de un índice de similitud.

4. Descripción del proceso TESTDENOMINATION

Control de entrada

El código de la especie debe existir en la base de datos de denominaciones.

La longitud máxima de la denominación es de 100 caracteres (incluidos los caracteres vacíos).

Caracteres excluidos: caracteres de puntuación y /-_' . No están permitidos los caracteres con tilde.

Están permitidas las denominaciones compuestas de varias palabras: 4 palabras como máximo. Deben estar separadas por un espacio en blanco.

a) Primera operación: dividir la denominación introducida en palabras individuales

Se divide la denominación que se va a comprobar en palabras individuales. Los caracteres vacíos se consideran separadores y se eliminan.

Se reemplazan por caracteres latinos básicos aquellos caracteres latinos que no lo son (caracteres con tilde,...). Todas las letras se transforman en mayúsculas. Véase el cuadro de transformación.

Ejemplo: Déjà se transforma en DEJA

Las letras dobles se reducen a una sola letra.

Ejemplo: HELLO se transforma en HELO.

Cada palabra individual se compara con la lista de palabras excluidas para la comprobación (es decir: YELLOW, RED, PURPLE,...).

Las palabras individuales encontradas en dicha lista se excluyen de la comprobación de similitud.

En la descripción siguiente, las palabras individuales obtenidas se denominan Words Tested (WT)

Ejemplo: Denominación 'Tānau TARI YELLOW'

Se excluye la palabra YELLOW de las comprobaciones. Las palabras tenidas en cuenta (WT) son:

- TANAU,
- TARI,

- TANAUTARI,
- TARIYELLOW,
- TANAUTARIYELLOW.

b) Segunda operación: elaboración de una lista de palabras de referencia de la clase

El programa busca la clase o el género al que pertenece la especie, para determinar el alcance de la búsqueda de similitud.

En caso de denominaciones registradas como códigos, se considera una sola cadena de caracteres, sin espacios en blanco.

El programa informático confecciona una lista de todas las palabras individuales que pertenecen a denominaciones de las variedades de la clase.

No se tienen en cuenta todas aquellas palabras que tengan más de 3 caracteres que la WT más larga y menos de 3 caracteres que la WT más corta.

Se excluyen de la comprobación de similitud todas las palabras de referencia que pertenecen a la lista de palabras excluidas para la comprobación antes mencionadas (es decir: color YELLOW, RED, PURPLE,...).

Como se comentó anteriormente, las letras dobles de las palabras de referencia se reducen a una sola letra.

En la descripción siguiente, las palabras individuales incluidas en la lista se denominan palabras de referencia (WR)

c) Principios de la comprobación de similitud

Para cada WT, el proceso SIMILARITYTEST calcula un índice de similitud con respecto a cada WR incluida en la lista de denominaciones antes confeccionada.

La lista de WR se ordena según el valor del índice de similitud. Todas las WR con un índice de similitud superior a un umbral predefinido se excluyen de los resultados.

Ejemplo:

Denominación comprobada 'Tänau TARI YELLOW'

La lista de palabras de referencia se compone de 10.000 palabras individuales.

TANAU, TARI, TANAUTARI, TARIYELLOW y TANAUTARIYELLOW se contrastan con las 10.000 WR. Se efectúan 50.000 comprobaciones (5 palabras x 10.000 WR).

5. Descripción detallada de la comprobación de similitud

Para cada par (WT, WR) se siguen los siguientes pasos.

Filtro preliminar: de la lista de WR, todas las palabras con más de 3 caracteres que la WT o menos de 3 caracteres que la WT no se consideran en los cálculos siguientes.

Primer paso: cálculo de χ^2

Fórmula: $\chi^2 = \text{sum}(di)^2 / (\text{longitud}(WT)-1)(\text{longitud}(WR)-1)$

donde di = diferencia del número de letras entre la palabra comprobada y la palabra de referencia. Se tienen en cuenta todas las letras de ambas palabras.

Ejemplo:

WT: **ALADIN** longitud: 6 caracteres.

Si comparamos esta WT con una cadena de caracteres existente: **DYLAN** (5 caracteres)

	A	L	D	I	N	Y
ALADIN	2	1	1	1	1	0
DYLAN	1	1	1	0	1	1

$$\chi^2 = ((2-1)^2 + (1-1)^2 + (1-1)^2 + (1-0)^2 + (1-1)^2 + (0-1)^2) / (6-1)(5-1)$$

$$\chi^2 = (1 + 0 + 0 + 1 + 0 + 1) / (5 \cdot 4)$$

$$\chi^2 = 3/20$$

$$\chi^2 = 0,15$$

Luego, conservamos para el siguiente cálculo todas las palabras en que:

- $\chi^2 \leq 0,3$ y longitud de la WT ≥ 5 caracteres
- $\chi^2 \leq 0,4$ y longitud de la WT = 4 caracteres
- $\chi^2 \leq 0,5$ y longitud de la WT < 4 caracteres

Segundo paso: cuatro cálculos basados en la selección efectuada en el primer paso

En este segundo paso se realizan 4 cálculos:

- El cálculo del porcentaje de las letras comunes.
- El cálculo del porcentaje de las letras NO comunes.
- El cálculo del porcentaje de la diferencia de longitud.
- El cálculo de la correlación por rangos de Kendall.

Cálculo del porcentaje de las letras comunes (LC)

LC = 1 - (Nº de letras comunes) / (longitud (WT))

Ejemplo: ALADIN y DYLAN

$$LC = 1 - 4/6 = 0,33$$

LC es igual a 0 cuando todas las letras se encuentran en la WR.

Segundo ejemplo: BANANAS y BANS

Todas las letras de BANANAS se encuentran en la palabra BANS.

$$LC = 1 - 7/7 \Rightarrow LC = 0$$

Cálculo del porcentaje de las letras NO comunes (LNC)

$$\text{LNC} = (\text{Nº de letras en WR ausentes en WT}) / (\text{longitud (WT)})$$

Ejemplo: ALADIN y DYLAN

$$\text{LNC} = 1/6 = 0,16$$

LNC es igual a 0 cuando todas las letras de WR se encuentran en la WT.

Segundo ejemplo: BANANAS y BANS

Todas las letras de BANS se encuentran en la palabra BANANAS.

$$\text{LNC} = 0/7 \Rightarrow \text{LNC} = 0$$

Cálculo del porcentaje de la diferencia de longitud (DL)

$$\text{DL} = (\text{diferencia de longitud entre las 2 cadenas}) / (\text{longitud de la WT})$$

Ejemplo: BANANAS y ANANAS

$$\text{DL} = 1/7$$

DL es igual a 0 cuando las longitudes de las 2 palabras son iguales.

Cálculo de la correlación por rangos de Kendall (CK)

$$\text{Fórmula: } \text{CK} = 6 * \text{sum}(\text{Di})^2 / \text{N} * (\text{N}^2 - 1)$$

donde:

- Di es igual a la diferencia de posición de las letras comunes (Li) menos la diferencia de posición de las letras previas. Si en la WT aparece varias veces la misma letra, se utilizan las primeras letras como referencia para la posición.
Si en la WR aparece varias veces la misma letra, se utilizan las letras más próximas como la misma letra en la WT.
- N es igual al número de letras comunes entre la WT y la WR.

Ejemplo:

WT: **ALADIN** y **DYLAN**

	A	L	D	N
ALADIN	1	2	4	6
DYLAN	4	3	1	5
Diferencia	1-4 = -3	2-3 = -1	4-1 = 3	6-5 = 1
Di	-3-0 = -3	1-(-3) = 4	3-(-1) = 4	1-3 = -2

4 letras en común A, L, D y N

$$\text{CK} = 6 * ((-3)^2 + (4)^2 + (4)^2 + (-2)^2) / (4 * (4^2 - 1))$$

$$\text{CK} = 6 * (9 + 16 + 16 + 4) / 4 * 15$$

$$\text{CK} = 6 * 45 / 60$$

$$\text{CK} = 4,5$$

WT: **ALADIN** y **BALADIN**

	A	L	D	I	N
ALADIN	1	2	4	5	6
BALADIN	2	3	5	6	7
Diferencia	1-2 = -1	2-3 = -1	4-5 = -1	5-6 = -1	6-7 = -1
Li	-1-0 = -1	-1-(-1) = 0	-1-(-1) = 0	-1-(-1) = 0	-1-(-1) = 0

5 letras en común A, L, D, I y N

$$CK = 6 * ((-1)^2 + (0)^2 + (0)^2 + (0)^2 + (0)^2) / (5 * (5^2 - 1))$$

$$CK = 6 * (1) / 5 * 24$$

$$CK = 1 / 20$$

$$CK = 0,05$$

En este ejemplo, se observa que si se presenta la misma secuencia de letras en las 2 palabras (LADIN en nuestro ejemplo), la diferencia entre las 2 secuencias se tiene en cuenta una sola vez.

Si las 2 palabras, la WT y la WR, son idénticas, el rango de Kendall es igual a 0.

Tercer paso: cálculo del índice de similitud

$$\text{índice de similitud} = CK + LC + LNC + DL$$

Se **selecciona** una palabra de referencia si el valor del índice es inferior o igual a:

	CK	LC	DL	Índice de similitud
Longitud de la palabra buscada > 4 caracteres	<= 1,5	<= 0,22	<= 1,5	< 1,2
Longitud de la palabra buscada = 4 caracteres	<= 1,5	<= 0,25	<= 1,26	< 1,2
Longitud de la palabra buscada < 4 caracteres	<= 1	<= 0,34	<= 1,0	< 1,2

Se ordenan las denominaciones de referencia para presentar los resultados por índice de similitud y luego por orden alfabético.

[Fin del Anexo y del documento]