



CAJ/68/9

ORIGINAL: englisch

DATUM: 2. September 2013

INTERNATIONALER VERBAND ZUM SCHUTZ VON PFLANZENZÜCHTUNGEN

Genf

VERWALTUNGS- UND RECHTSAUSSCHUSS

Achtundsechzigste Tagung Genf, 21. Oktober 2013

ETWAIGE ENTWICKLUNG EINES UPOV-SUCHINSTRUMENTS FÜR ÄHNLICHKEITEN ZUM ZWECK DER SORTENBEZEICHNUNG

Vom Verbandsbüro erstelltes Dokument

Haftungsausschluß: dieses Dokument gibt nicht die Grundsätze oder eine Anleitung der UPOV wieder

1. Der Verwaltungs- und Rechtsausschuß (CAJ) hörte auf seiner siebenundsechzigsten Tagung am 21. März 2013 in Genf einen Vortrag der Delegation der Europäischen Union über die Erfahrung des Gemeinschaftlichen Sortenamtes (CPVO) mit der Verwendung seines Suchinstruments für Ähnlichkeiten zum Zweck der Sortenbezeichnung bei der Prüfung der vorgeschlagenen Sortenbezeichnungen. Im Verlauf des Vortrags schlug das CPVO vor, auf Grundlage des Suchinstruments des CPVO die Möglichkeiten für die Entwicklung eines Suchinstruments für Ähnlichkeiten der Sortenbezeichnung für die UPOV auszuloten¹. Der CAJ begrüßte den Vorschlag des CPVO und vereinbarte, einen Punkt zur Prüfung dieses Vorschlags auf die Tagesordnung seiner achtundsechzigsten Tagung zu setzen (vergleiche Dokument CAJ/67/14 „Bericht über die Entschließungen“, Absätze 49 und 50).

HINTERGRUND

2. Die Registerkarte Bezeichnungssuche der Datenbank für Pflanzensorten (PLUTO-Datenbank) (<http://www.upov.int/pluto/de/>) enthält derzeit folgende Suchtypen zum Auffinden ähnlicher Bezeichnungen:

- a) Ähnlichkeitsfaktor [Suchprogramm des CPVO] Dabei wird eine Analyse der von Ihnen eingegebenen Bezeichnung unter Berücksichtigung einer Kombination von Faktoren, einschließlich gemeinsamer Buchstaben, relativer Wortlänge und Position der gemeinsamen Buchstaben durchgeführt. Dies ist das komplexeste Vergleichsverfahren und die Suche kann ein paar Sekunden in Anspruch nehmen. Der Vergleichsfaktor wurde von der französischen GEVES und dem Gemeinschaftlichen Sortenamte der Europäischen Union (CPVO) entwickelt. Bitte beachten Sie jedoch, daß die Ergebnisse der Suche nach dem Ähnlichkeitsfaktor in der PLUTO-Datenbank einer Auslegung bedürfen und keine Garantie für die Eignung von Sortenbezeichnungen liefern, über die von der für die Erteilung des Züchterrechts im jeweiligen Hoheitsgebiet zuständigen Behörde entschieden werden muß.

Eine ausführliche Erklärung der Analyse ist in der Anlage dieses Dokuments enthalten.

¹ Der Ähnlichkeitsfaktor wurde von der französischen *Gruppe für Sorten- und Saatgutprüfung und -kontrolle (Groupe d'étude et de contrôle des variétés et des semences)* (GEVES) und dem Gemeinschaftlichen Sortenamte der Europäischen Union (CPVO) entwickelt.

- b) Unschärfe Suche Dabei wird nach Bezeichnungen gesucht, die Wörter enthalten, die abgesehen von einem oder zwei Buchstaben wie der von Ihnen eingegebene Begriff geschrieben werden. Dies ist vergleichbar mit der unscharfen Suche in der Registerkarte Begriffssuche.
- c) Phonetisch Hier wird nach Bezeichnungen gesucht, die Wörter enthalten, die sich ähnlich wie der von Ihnen eingegebene Begriff anhören. Dies entspricht in etwa dem phonetischen Suchverfahren in der Registerkarte Begriffssuche.
- d) Enthält Wird nach Bezeichnungen gesucht, die Wörter enthalten, die dieselbe Serie von Buchstaben, wie der von Ihnen eingegebene Begriff enthalten. Dies entspricht in etwa der Methode 'Enthält' in der Registerkarte Begriffssuche.
- e) Beginnt Dabei wird nach Bezeichnungen gesucht, die Wörter enthalten, die mit derselben Reihenfolge von Buchstaben beginnen, wie der von Ihnen eingegebene Begriff. Dies entspricht in etwa der Methode 'Beginnt' in der Registerkarte Begriffssuche.
- f) Endet Dabei wird nach Bezeichnungen gesucht, die Wörter enthalten, die mit derselben Reihenfolge von Buchstaben enden, wie der von Ihnen eingegebene Begriff. Das ist ähnlich wie das Suchverfahren 'Endet' in der Registerkarte Begriffssuche.

3. Bei Sondierungsgesprächen mit dem Verbandsbüro über die Frage, wie ein Suchinstrument zum Auffinden ähnlicher Sortenbezeichnungen für die UPOV zu entwickeln sei, stellte das CPVO klar, daß sämtliche Optionen geprüft werden sollten und daß in Anbetracht der Fortschritte in der Informationstechnologie nicht unbedingt das CPVO-Suchinstrument für die Entwicklung des besten Suchinstruments zugrunde gelegt werden müsse. Die wichtigste Überlegung sei, ein Instrument zu entwickeln, das von allen UPOV-Mitgliedern benutzt werden könnte, um Abweichungen bei den Entscheidungen zur Eignung einer Bezeichnung möglichst gering zu halten.

VORSCHLAG

4. In Anbetracht der Tatsache, daß ein gewisses Wissen über Sortenbezeichnungen und über informationstechnologische Instrumente vonnöten ist, wird vorgeschlagen, eine kleine Arbeitsgruppe zur Ausarbeitung von Vorschlägen für ein UPOV-Suchinstrument für Sortenbezeichnungen einzusetzen. Eine solche Gruppe könnte folgendermaßen zusammengesetzt sein:

- a) Bezeichnungsprüfer von Verbandsmitgliedern (3 bis 6 Sachverständige);
- b) WIPO Global Databases Service (zuständig für die PLUTO-Datenbank);
- c) Gemeinschaftliches Sortenamts der Europäischen Union (CPVO); und
- d) Verbandsbüro.

5. Der Arbeitsplan der Arbeitsgruppe würde von der Arbeitsgruppe selbst erstellt werden, wobei allerdings davon ausgegangen wird, daß der erste Schritt die Prüfung der derzeit in der Registerkarte 'Bezeichnungssuche' der PLUTO-Datenbank verfügbaren Suchtypen, insbesondere des Ähnlichkeitsfaktors (CPVO-Suchinstrument) sowie die Prüfung von Suchtypen, die in anderen Situationen (z.B. in Bezug auf Handelsmarken) verwendet werden und die eine alternative Ausgangsbasis für ein UPOV-Suchinstrument zum Auffinden ähnlicher Sortenbezeichnungen darstellen könnten, wäre.

6. Bei der Prüfung der Eignung von Suchtypen würde man sich insbesondere auf Dokument UPOV/INF/12 „Erläuterungen zu Sortenbezeichnungen nach dem UPOV-Übereinkommen“ beziehen. Diesbezüglich müßte sich die Arbeitsgruppe für weitere Anleitung an den CAJ wenden, falls bei der Arbeit deutlich würde, daß eine Überarbeitung von Dokument UPOV/INF/12 für die Entwicklung eines effektiven UPOV-Instruments zum Auffinden ähnlicher Sortenbezeichnungen erforderlich wäre.

7. Es wird vorgeschlagen, daß die Sitzungen der Arbeitsgruppe vom Verbandsbüro in Genf ausgerichtet werden und daß das Verbandsbüro den Vorsitz führt. Diese Sitzungen würden nicht zeitgleich mit den UPOV-Tagungen angesetzt werden und die Beteiligung von Sortenprüfern und des CPVO über elektronische Medien würde vorausgesetzt werden. Von der Arbeitsgruppe ausgearbeitete Vorschläge würden dem CAJ und dem Technischen Ausschuß (TC) vorgelegt werden, und der CAJ und der TC würden einen kurzen Bericht über die Sitzungen der Arbeitsgruppe erhalten.

8. *Der CAJ wird ersucht:*

a) *der Einsetzung einer Arbeitsgruppe für die Ausarbeitung von Vorschlägen für eine Suchinstrument für Ähnlichkeiten zum Zweck der Sortenbezeichnung für die UPOV, wie in den Absätzen 4 bis 7 dieses Dokuments vorgeschlagen, zuzustimmen; und*

b) *anzuregen, daß Verbandsmitglieder Vorschläge für entsprechend erfahrene Sortenbezeichnungsprüfer, die an der Arbeitsgruppe teilnehmen, einreichen.*

[Anlage folgt]

EUROPÄISCHE UNION
Gemeinschaftliches Sortenamt

SUCHVORGANG

1. ALLGEMEIN

Als Schlußfolgerung der Vorbereitungsphase des Projekts, das das Büro dem CPVO-Verwaltungsausschuß im November 2003 vorgelegt hat, schlug das Büro vor, zunächst einmal nationale Software zu übernehmen, diese an die zentralisierte Datenbank anzupassen und sämtliche Programme zu deren Prüfung auszuführen. Die Wahrscheinlichkeit, daß eine sehr ähnliche Bezeichnung nicht gefunden würde, wäre auf diese Weise sehr gering.

Langfristiger gesehen wurde die Entwicklung einer CPVO-Software mit der Möglichkeit der Entwicklung von Sprachfunktionen vorgesehen.

In der Praxis wurden bei der Implementierung des Suchvorgangs in der CPVO-Datenbank die Spezifikationen der französischen Software zugrunde gelegt.

2. BESTIMMUNGEN, DIE AUSREICHEND UNTERSCHIED ZWISCHEN ZWEI SORTENBEZEICHNUNGEN FESTLEGEN

Gemäß der Grundverordnung über den gemeinschaftlichen Sortenschutz besagt eine der Durchführungsbestimmungen, der eine Sortenbezeichnung entsprechen muß, daß sie nicht mit einer Sortenbezeichnung, unter der eine andere Sorte derselben/einer eng verwandten Art eingetragen ist, identisch oder verwechselbar sein sollte.

Diese Bestimmung wurde in den Leitlinien des Verwaltungsrates des CPVO zu Sortenbezeichnungen ausgelegt:

- ♣ Ein Unterschied von nur einem Buchstaben oder von Akzenten auf Buchstaben ist im Allgemeinen als irreführend anzusehen.
- ♣ Unterschiede von zwei oder mehr Buchstaben sollten im Allgemeinen nicht als irreführend betrachtet werden, es sei denn, zwei Buchstaben tauschen lediglich die Plätze.
- ♣ Zudem sollte eine Sortenbezeichnung nicht den falschen Eindruck erwecken, daß die Sorte mit einer anderen Sorte verwandt oder von dieser abgeleitet ist.

Zweck des Suchprogramms ist das Auffinden in der Datenbank von Bezeichnungen derselben Klasse, die in Konflikt mit einer vorgeschlagenen Bezeichnung stehen könnten.

3. SUCHVORGANG

Die Prüfungen werden von einem internen Programm der ORACLE-Datenbank auf dem CPVO-Server ausgeführt (bessere Leistung).

Zur Durchführung einer Prüfung führt das Schnittstellenprogramm (Website,...) einen Vorgang aus, der folgendermaßen genannt wird:

TESTDENOMINATION

Als Parameter übermittelt das Schnittstellenprogramm die zu prüfende Bezeichnung sowie den Code der Art, der die Sorte angehört.

Der Vorgang gibt den Identifikator der durchgeführten Prüfung an die Schnittstelle (Spalte Testid) zurück. Anhand dieses Identifikators können wir die 2 Tabellen, die die Prüfungsergebnisse enthalten, lesen: die Tabellen TESTS und TESTRESULTS.

Die Tabelle TESTS enthält allgemeine Angaben zur Prüfung: Datum der Prüfung, Kennung der Person, die die Prüfung beantragt hat, Bezeichnung, Artencode, Code der Klasse oder Gattung, Computerlaufzeit, ausgenommene Wörter, Fehlernachricht,...

Die Tabelle TESTRESULTS enthält die Listen der Bezeichnungen, die im Rahmen des Vorgangs TESTDENOMINATION als ähnliche Bezeichnungen gefunden wurden. Jede ähnliche Bezeichnung wird mit einem Ähnlichkeitsindex verknüpft.

4. Beschreibung des Vorgangs TESTDENOMINATION

Inputkontrolle

Die Art muß in der Bezeichnungs-Datenbank existieren.

Die maximale Länge der Bezeichnung beträgt 100 Zeichen (einschließlich Leerzeichen).

Ausgenommene Zeichen: Satzzeichen und /-_'. Hervorgehobene Zeichen sind nicht zulässig.

Aus mehreren Wörtern zusammengesetzte Bezeichnungen sind zulässig: maximal 4 Wörter. Sie müssen durch ein Leerzeichen voneinander getrennt sein.

a) Erster Vorgang: Zerlegung der Ausgangsbezeichnung in einzelne Wörter

Die zu prüfende Bezeichnung wird in elementare Grundwörter zerlegt. Leerzeichen werden als Separatoren betrachtet und gelöscht.

Nichtlateinische Schriftzeichen (hervorgehobene Buchstaben,...) werden durch lateinische Standardschriftzeichen ersetzt. Alle Buchstaben werden in Großbuchstaben verwandelt. Siehe Umwandlungstabelle.

Beispiel: *Déjà* wird zu DEJA

Doppelbuchstaben werden auf einen einzigen Buchstaben reduziert.

Beispiel: HELLO wird zu HELO.

Jedes Grundwort wird mit der Liste von Wörtern, die vom Prüfungsvorgang ausgenommen sind (z.B.: Farbe GELB, ROT, PURPURN,...), abgeglichen

Die in diese Liste aufgenommenen Grundwörter werden von der Ähnlichkeitsprüfung ausgenommen.

Bei folgender Beschreibung werden die erhaltenen Grundwörter als „*Words Tested (WT)*“ bezeichnet

Beispiel: Bezeichnung 'Tänau TARI YELLOW'

Das Wort gelb wird von der Prüfung ausgenommen. Wörter, die geprüft werden (WT), sind:

- TANAU,
- TARI,
- TANAUTARI,
- TARIYELLOW,
- TANAUTARIYELLOW.

b) Zweiter Vorgang: Erstellung einer Liste von Vergleichswörtern innerhalb der Klasse

Das Programm durchsucht die mit der Art verknüpfte Klasse oder Gattung, um den Anwendungsbereich der Ähnlichkeitssuche zu bestimmen.

Im Falle von als Codes eingetragenen Bezeichnungen wird eine einzige Zeichenfolge ohne Leerzeichen geprüft.

Die Software erstellt eine Liste aller Grundwörter, die zu Bezeichnungen der Sorten der Klasse gehören.

Alle Wörter, die über 3 Buchstaben mehr als das längste WT und 3 Buchstaben weniger als das kürzeste WT umfassen, werden nicht berücksichtigt.

Alle Vergleichswörter, die zu der Liste von Wörtern, die wie oben ausgeführt von der Prüfung ausgenommen werden, gehören (z.B.: Farbe GELB, ROT, PURPURN, ...), werden von der Ähnlichkeitsprüfung ausgenommen.

Genau wie oben beschrieben werden bei den Vergleichswörtern Doppelbuchstaben auf einen einzigen Buchstaben reduziert.

Bei folgender Beschreibung werden die in diese Liste aufgenommenen Grundwörter als Vergleichswörter (*Words of reference* (WR)) bezeichnet.

c) Die Grundlagen der Ähnlichkeitsprüfung

Für jedes WT wird anhand des Verfahrens SIMILARITYTEST ein Ähnlichkeitsindex in Bezug auf jedes in der oben genannten Bezeichnungsliste enthaltene Vergleichswort (WR) berechnet.

Die Liste der WR wird gemäß dem Wert des Ähnlichkeitsindex sortiert. Alle WR mit einem Ähnlichkeitsindex, der über einem zuvor festgelegten Schwellenwert liegt, werden von der Suche ausgenommen.

Beispiel:

Geprüfte Bezeichnung 'Tänau TARI YELLOW'

Die Liste der Vergleichswörter enthält 10 000 Grundwörter.

TANAU, TARI, TANAUTARI, TARIYELLOW und TANAUTARIYELLOW werden mit allen 10.000 WR abgeglichen. Es werden 50.000 Prüfungen durchgeführt (5 Wörter x 10 000 WR).

5. Detaillierte Beschreibung der Ähnlichkeitsprüfung

Folgende Schritte werden für jedes Paar (WT, WR) durchgeführt.

Vorfilter: aus der Liste der Vergleichswörter (WR) werden alle Wörter, die über 3 Buchstaben mehr als das zu prüfende Wort (WT) oder 3 Buchstaben weniger enthalten, bei folgenden Berechnungen nicht berücksichtigt.

Erster Schritt: Berechnung von Ki2

Formel: $Ki2 = \frac{\sum(di)^2}{(Länge(WT)-1)(Länge(WR)-1)}$

Dabei ist di = die Differenz zwischen der Anzahl der Buchstaben des geprüften Wortes und der des Vergleichswortes. Alle Buchstaben beider Wörter werden berücksichtigt.

Beispiel:

WT : **ALADIN** Länge 6 Zeichen.

Wenn wir dieses WT mit der bestehenden Zeichenfolge vergleichen: **DYLAN** (5 Zeichen)

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

$$Chi2 = \frac{((2-1)^2 + (1-1)^2 + (1-1)^2 + (1-0)^2 + (1-1)^2 + (0-1)^2)}{(6-1)(5-1)}$$

$$Chi2 = \frac{(1 + 0 + 0 + 1 + 0 + 1)}{(5 \cdot 4)}$$

$$Chi2 = \frac{3}{20}$$

$$Chi2 = 0,15$$

Dann behalten wir für die folgende Berechnung alle Wörter bei, für die gilt:

- $Ki2 \leq 0,3$ und Länge des WT ≥ 5 Buchstaben
- $Ki2 \leq 0,4$ und Länge des WT ≥ 4 Buchstaben
- $Ki2 \leq 0,5$ und Länge des WT ≥ 4 Buchstaben

Zweiter Schritt: Vier Berechnungen, die auf der im ersten Schritt getroffenen Auswahl basieren

In diesem zweiten Schritt werden 4 Berechnungen durchgeführt:

- Berechnung des Prozentsatzes der gemeinsamen Buchstaben
- Berechnung des Prozentsatzes der NICHT gemeinsamen Buchstaben
- Berechnung des Prozentsatzes der Differenz der Länge
- Berechnung der Rang-Korrelation nach Kendall (KK)

Berechnung des prozentualen Anteils gemeinsamer Buchstaben (*common letter* = CL)

CL = 1-(Anzahl gemeinsamer Buchstaben)/ (Länge (WT))

Beispiel: ALADIN und DYLAN

$$CL = 1 - \frac{4}{6} = 0,33$$

CL ist gleich 0 wenn alle Buchstaben im WR vorkommen.

Zweites Beispiel: BANANAS und BANS

Alle Buchstaben von BANANAS kommen auch im Wort BANS vor.

$$CL = 1 - \frac{7}{7} \Rightarrow CL = 0$$

Berechnung des Prozentsatzes der NICHT gemeinsamen Buchstaben (*non-common letter* = NCL)

NCL = (Anzahl der Buchstaben im WR, die nicht im WT vorkommen) / (Länge (WT))

Beispiel: ALADIN und DYLAN
NCL = 1/6 = 0,16

NCL ist gleich 0 wenn alle Buchstaben im WR im WT vorkommen.

Zweites Beispiel: BANANAS und BANS

Alle Buchstaben von BANS kommen im Wort BANANAS vor.
NCL= 0/7 => NCL=0

Berechnung des Prozentsatzes der Differenz der Länge (*difference of length* = DL)

DL = (Längendifferenz von 2 Zeichenfolgen) / (Länge des WT)

Beispiel: BANANAS und ANANAS
DL = 1/7
DL ist gleich 0, wenn die Länge der 2 Wörter gleich ist.

Berechnung der Rang-Korrelation nach Kendall (KK)

Formel: $KK = 6 \cdot \sum(D_i)^2 / N \cdot (N^2 - 1)$

Wobei:

- D_i die Positionsdifferenz der gemeinsamen Buchstaben (L_i) abzüglich der Positionsdifferenz von den vorhergehenden Buchstaben ist. Kommt im WT ein und derselbe Buchstabe mehrmals vor, so verwenden wir die ersten Buchstaben als Bezugspunkt für die Position. Tritt im WR derselbe Buchstabe mehrmals auf, so verwenden wir die am nächsten stehenden Buchstaben als gleichen Buchstaben im WT.
- N ist gleich die Anzahl der dem WT und einem WR gemeinsamen Buchstaben.

Beispiel:

WT : **ALADIN** und **DYLAN**.

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

4 gemeinsame Buchstaben A, L, D und N

$$KK = 6 \cdot ((-3)^2 + (4)^2 + (4)^2 + (-2)^2) / (4 \cdot (4^2 - 1))$$

$$KK = 6 \cdot (9 + 16 + 16 + 4) / 4 \cdot 15$$

$$KK = 6 \cdot 45 / 60$$

$$KK = 4,5$$

WT : **ALADIN** und **BALADIN** .

	A	L	D	I	N
ALADIN	1	2	4	5	6
BALADIN	2	3	5	6	7
Differenz	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 gemeinsame Buchstaben A, L, D, I und N

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

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

$$KK = 1 / 20$$

$$KK = 0,05$$

An diesem Beispiel können wir sehen, daß die Lücke zwischen den 2 Sequenzen nur einmal berücksichtigt wird, wenn dieselbe Buchstabensequenz in beiden Wörtern vorkommt (LADIN in unserem Beispiel).

Sind die beiden Wörter WT und WR identisch, so ist der Kendall-Rang 0.

Dritter Schritt: Berechnung des Ähnlichkeitsindexes

$$\text{Ähnlichkeitsindex} = KK + CL + NCL + DL$$

Ein Vergleichswort wird **selektiert**, wenn die Ergebnisse niedriger sind als oder gleich sind wie:

	KK	CL	DL	Ähnlichkeitsindex
Länge des gesuchten Wortes > 4 Zeichen	<=1,5	<=0,22	<=1,5	<1,2
Länge des gesuchten Wortes = 4 Zeichen	<=1,5	<=0,25	<=1,26	<1,2
Länge des gesuchten Wortes < 4 Zeichen	<=1	<=0,34	<=1,0	<1,2

Vergleichsbezeichnungen werden sortiert nach Ähnlichkeitsindex und in alphabetischer Reihenfolge angezeigt.

[Ende der Anlage und des Dokuments]