



STANDARD ST.22

EMPFEHLUNG FÜR DIE ERSTELLUNG VON PATENTANMELDUNGEN ZUR VEREINFACHUNG DER MASCHINENLESBARKEIT (OPTICAL CHARACTER RECOGNITION - OCR)

Änderungen von der SCIT Arbeitsgruppe für Standards und Dokumentation in ihrer zehnten Sitzung am 21. November 2008 angenommen

EINFÜHRUNG

1. Diese Empfehlung betrifft Patentanmeldungen, die auf Papier oder elektronisch (e-filed) eingereicht werden, wobei der Fließtext der elektronischen Anmeldung in Bildform (zB PDF oder TIFF Bilder) vorliegt.
2. Diese Empfehlung wurde zur Unterstützung bei der Vorbereitung von Patentanmeldungen in Maschinschrift erstellt, die zur anschließenden Herstellung einer elektronisch digitalisierten Aufzeichnung des Inhalts der Patentanmeldung unter Anwendung der optischen Zeichenerkennung (Optical Character Recognition - OCR) geeignet sind.
3. Diese Empfehlung wurde auf Grundlage der Erfahrungen zahlreicher Ämter bei der Benutzung von OCR-Ausrüstungen erstellt. Sie wurde mit dem Ziel abgefasst, die geringst mögliche Fehlerrate beim Schritt der automatischen Texterkennung zu erreichen und gleichzeitig ein effizientes Lesen zu erlauben. Es ist zu beachten, dass dieses Dokument keine detaillierten Empfehlungen für japanische und koreanische Sprachen vorsieht; der Prozentsatz der elektronischen Volltextanmeldungen in diesen Ländern liegt allerdings über 90%, was bedeutet, dass dieser Standard in der Praxis für die japanische und koreanische Sprachen nicht betrifft.
4. Die Hauptaufgabe bei der Herstellung einer digitalisierten Aufzeichnung einer Patentanmeldung ist die Ermöglichung einer einfachen Veröffentlichung dieser Anmeldung in einem zusammengesetzten Format unter Einsatz von Computersatztechniken und damit die Darstellung und den Wert von Patentdokumenten zum Vorteil der Benutzer zu verbessern. Eine weitere Aufgabe ist die Schaffung einer maschinenlesbaren Datenbank aus dem Volltext des veröffentlichten Dokuments, sodass zu einem späteren Zeitpunkt die Vorteile aus einer damit möglichen Computer-Volltextsuche gezogen werden können.

DEFINITIONEN

5. Für die Zwecke dieser Empfehlung bedeutet der Ausdruck „Patentanmeldung“ Anmeldungen von Patenten, Gebrauchsmustern und ähnlichen Schutzrechten (zB inventor's certificates, utility certificates, patents or certificates of addition, inventor's certificates of addition and utility certificates of addition).
6. Eine mathematische oder chemische Formel wird als „komplex“ bezeichnet, wenn sie nicht als geradlinige Folge fortlaufender Zeichen mit jeweils möglichen hoch- oder tiefgestellten Attributen darstellbar ist. Eine Formel ist jedenfalls komplex, wenn sie verschachtelte Hoch- oder Tiefstellungen enthält oder wenn sie das mathematische Summen-, Integral- oder Produktsymbol enthält.
7. Das Begrenzungsrechteck eines Zeichens / einer Zeichenfolge ist das kleinste achsausgerichtete Rechteck, das alle Teile des Zeichens / der Zeichenfolge enthält.
8. Der Ausdruck „kursiv“ verweist auf eine stilisierte Form der Handschrift, wobei die Buchstaben innerhalb der Worte verbunden sind, wodurch ein Wort zu einem einzigen komplexen Federstrich wird. Schriftarten werden als kursiv bezeichnet, wenn sie so gestaltet sind, dass sie der Handschrift ähneln.

ERSTELLUNG DES ORIGINALS

9. Eine Patentanmeldung wird häufig mittels Textverarbeitung vorbereitet. Die Erfahrung hat gezeigt, dass das effizienteste Format, mit dem eine OCR-Ausrüstung zuverlässig verwendet werden kann, jenes im Internationalen Standard Organization (ISO) Standard 1073/II, dem sogenannten OCR-B Format, definierte ist.

PAPIERTRÄGER, WENN AUF PAPIER EINGEREICHT WIRD

10. Um das Scannen zu erleichtern, soll das Papier, auf das die Anmeldung gedruckt ist, folgende Eigenschaften aufweisen:
 - (a) Das Papier soll stark, weiß und sauber sein.
 - (b) Das Gewicht des Papiers soll zwischen 70 und 120 g/m², vorzugsweise 80 g/m² betragen.



- (c) Das Papierformat soll vorzugsweise A4, das sind 210 mm x 297 mm, oder 8 1/2 x 11 Zoll (was dem de facto nordamerikanischen Standard entspricht) sein.
- (d) Die Blätter sollen frei von Falten oder Löchern und nicht gerollt sein.
- (e) Das Papier soll nicht saugend sein, um ein Verschmieren der Tinte zu vermeiden (zB bei Verwendung eines Tintenstrahldruckers).

EMPFEHLUNGEN FÜR DAS SEITENLAYOUT

- 11. Doppelseitige Ausdrücke sind zu vermeiden. Falls dies nicht möglich ist, soll hinreichend undurchsichtiges Papier verwendet werden, um ein sauberes Scannen der Vorder- und Rückseiten zu gewährleisten.
- 12. Die Druckzeichen sollen schwarz auf weißem Hintergrund sein.
- 13. Ein mindestens 2 cm breiter Rand oben, unten und rechts sowie ein mindestens 2,5 cm breiter Rand links sollen auf allen Blättern vorhanden sein. Kennzeichen (interne Aktenzeichen des Anmelders oder Vertreters) sollen im oberen Rand aufscheinen. Bitte beachten Sie Anhang 1.
- 14. Zeilennummerierungen sind zu vermeiden. Falls dies nicht möglich ist, sind die Zeilennummern in arabischen Ziffern in linken Rand, mindestens 1 cm außerhalb des Textkörperrandes, wie im Anhang 1 dargestellt, vorzusehen. Die Schriftgröße der Zeilennummern soll mindestens 12 Punkt betragen.
- 15. Die Seitennummerierung soll in arabischen Ziffern ohne begrenzende Zeichen erfolgen. Die Seitennummern sollen vorzugsweise im oberen oder unteren Rand des Blattes zentriert sein, wie im Anhang 1 dargestellt.
- 16. Die Beschreibung, die Ansprüche und die Zusammenfassung sollen jeweils auf einer neuen Seite beginnen. Weiters soll das erste Wort, das auf der ersten Seite jedes dieser drei genannten Teile gedruckt ist, den entsprechenden Teil bezeichnen (in der Sprache der Anmeldung); Die Ansprüche sollen fortlaufend nummeriert sein. Das Format der Anspruchsnummerierung soll eine saubere Trennung der Nummerierung vom Anspruchstext für jeden Anspruch erlauben. Empfohlene Formate sind entweder arabische Nummern, denen ein Punkt folgt, oder die Verwendung des Wortes „Anspruch“ – oder das entsprechende Wort in der Sprache der Anmeldung – gefolgt von einem Abstand und der arabischen Anspruchsnummer, wobei in beiden Fällen der Anspruchstext rechtseingerückt von der Anspruchsnummer um mindestens 1 cm folgt.
- 17. Die Seiten sollen aus einspaltigen Absätzen bestehen (Textabsätze oder Absätze, die ein eingebettetes Bild enthalten).
- 18. Seiten, die Absätze enthalten, sollen im Hochformat ausgerichtet werden.
- 19. Querformat ist zu vermeiden. Querformat ist nur zulässig, wenn die Seiten Grafiken oder Tabellen enthalten, die im Hochformat nicht darstellbar sind.
- 20. Jede Seite darf nur eine Textausrichtung (horizontal oder vertikal) beinhalten.
- 21. Seiten im Querformat sind um 90 Grad gegen den Uhrzeigersinn zu drehen, wenn sie in einen Satz von Seiten im Hochformat eingefügt werden.
- 22. Fußnoten, Randtexte und Überschriften sind zu vermeiden, ausgenommen jene in den Absätzen 14 (Zeilennummerierung) und 15 (Seitennummerierung) angeführten sowie zur Einfügung des internen Kennzeichens des Anmelders bzw. Vertreters in der oberen linken Ecke des Randes.

EMPFEHLUNGEN FÜR DAS ABSATZLAYOUT

- 23. Es wird empfohlen, dass im Text eingebettete Tabellen sowie komplexe chemische oder mathematische Formeln, Bilder und Zeichnungen von den Textabsätzen getrennt werden. Es wird angeraten, solche Objekte mit über die ganze Seitenbreite verlaufenden oberen und unteren Rändern von mindestens 1 cm zu umgeben.
- 24. Bilder und Zeichnungen sollen in den Abschnitt „Zeichnungen“ eingefügt werden und auf diese ist in den Abschnitten „Beschreibung“ und „Ansprüche“ zu verweisen.
- 25. Bilder und Zeichnungen sollen in Schwarz-Weiß ausgeführt werden (Grautöne sind zu vermeiden, da beim Scannen Informationen verloren gehen). Die Figuren sollen deutliche Linien enthalten, die dick genug sind, um bei einer Auflösung von 300 dpi gut dargestellt zu werden.
- 26. Handgeschriebene Texte oder Anmerkungen sind zu vermeiden. Wenn diese notwendig sind, werden sie als eingebettete Zeichnungen betrachtet und sollen den Empfehlungen des Absatzes 23 folgen.
- 27. Der Zeilenabstand soll 1,5 Zeilen betragen.



28. Absätze sollen einen Abstand voneinander haben, der mindestens dem doppelten Zeilenabstand innerhalb des Absatzes entspricht.
29. Alle Zeichen in einer Absatzzeile sollen sorgfältig an ihrer Grundlinie ausgerichtet sein, außer hoch- und tiefgestellte Zeichen, wie im Absatz 35 näher erläutert.
30. Blocksatz ist zu vermeiden. Wenn Blocksatz angewandt wird, soll der Abstand zwischen den Worten zumindest so groß sein, wie dieser ohne Blocksatz wäre. Blocksatz kann das korrekte Erkennen der Wortgrenzen in einem Absatz durch OCR-Systeme verhindern.
31. Die Abteilung von Worten mit Bindestrichen ist möglichst zu vermeiden (zB am Ende der Zeilen oder der Tabellenzellen). Dies gilt nicht für Sprachen, in denen zusammengesetzte Hauptwörter verwendet werden (zB die deutsche Sprache).

EMPFEHLUNGEN FÜR TABELLEN

32. Es soll ausschließlich weißer Hintergrund verwendet werden.
33. Tabellen müssen Ränder aufweisen. Die Zellenränder sollen mit durchgehenden Linien von mindestens 1,5 Punkten Dicke ausgeführt sein.

EMPFEHLUNGEN FÜR DIE SCHRIFTART

34. Die kleinste empfohlene Schriftgröße ist 12 Punkt, wobei eine Größe von 14 Punkt zu bevorzugen ist. Als allgemeine Empfehlung sollen alle Zeichen eines Absatzes die gleiche Zeichengröße aufweisen.
35. Absätze, die hoch- oder tiefgestellte Zeichen enthalten, sollen eine Schriftgröße von mindestens 12 Punkt aufweisen (14 Punkt ist zu bevorzugen – je größer, desto besser). Es ist sicherzustellen, dass das Umgrenzungsrechteck der hoch- oder tiefgestellten Zeichen das Umgrenzungsrechteck der normalen Zeichen in der gleichen Zeile ausreichend schneidet (Dies verhindert, dass das OCR-Verfahren die hoch- oder tiefgestellten Zeichen in eigene Zeilen setzt).
36. Es werden folgende nach Eignung geordnete Schriftarten empfohlen:
- Nichtproportionale Schriftarten (Monospaced family): OCR-B, Courier New, Free Mono.
 - Schriftarten mit Serifen: ITC Officina Serif, Times New Roman, Free Times.
 - Serifenlose Schriftarten: Verdana, ITC Officina Sans, Arial, Helvetica, DejaVu Sans.

Die Schriftarten Arial, Helvetica, DejaVu Sans, Free Times and Times New Roman werden nicht empfohlen, wenn die Anmeldung chemische und/oder mathematische Formeln sowie Abkürzungen enthält, die Buchstaben und Zahlen mischen. Für chinesische Buchstaben wird der Zeichensatz Song empfohlen.

37. Die Zeichen der Schriftart sollen gut ausgeformt und ohne Schatten sein. Der Zeichenabstand soll groß genug sein (enge Abstände sind zu vermeiden).
38. Eng gestellte (narrow) und verbundene (kursive) Schriftarten sollen nicht verwendet werden.
39. Fette und kursive Textauszeichnungen sind möglichst zu vermeiden.
40. Unübliche (nicht genormte / untypische / unregelmäßige) Zeichen sind möglichst zu vermeiden. Wenn nötig, sollen Sonderzeichen aus dem genormten griechischen Alphabet und dem Symbol-Zeichensatz (in der Reihenfolge des Auftretens) gewählt werden. Zeichen, die im UNICODE-Bereich nicht vorhanden sind, dürfen nicht verwendet werden: Diese Zeichen werden von OCR-Maschinen als eingebettete Bilder erkannt und machen daher den erkannten Text schwer lesbar. Jedes Amt sollte die Erfordernisse der Schriftarten, die zur Vorbereitung von Anmeldungen verwendet werden sollen, definieren und veröffentlichen.
41. Der Text soll nicht unterstrichen werden. Falls dies notwendig ist, darf die Unterstreichungsline das Umgrenzungsrechteck der unterstrichenen Zeichen nicht schneiden.

EMPFEHLUNGEN FÜR DIE KOMBINATION MEHRERER SPRACHEN

42. Innerhalb von Abschnitten / Seiten von Anmeldungen ist das Vermischen von asiatischen (zB basierend auf Ideogrammen) und europäischen (zB lateinische und kyrillische Alphabete) für das OCR-Verfahren problematisch und soll vermieden werden, außer dies ist notwendig.



EMPFEHLUNGEN FÜR DAS SCANNEN

43. Anmeldungen sollen entweder in Schwarz-Weiß oder in Graustufen gescannt werden.
44. Die bevorzugte Auflösung für das Scannen ist 300 dpi. Das Scannen mit einer Auflösung unter 300 dpi, selbst in Graustufen, kann zu einer schlechten Qualität der veröffentlichten Dokumente führen, da der Dokumentaustausch zwischen den Ämtern und die Veröffentlichungsverfahren häufig mit einer Umwandlung in Seiten im Format 300 dpi Schwarz-Weiß TIFF Gruppe IV verbunden ist.
45. Gescannte Dokumente sind entweder in das PDF oder in das TIFF Format umzuwandeln.

KORREKTUREN

46. Verbesserungen im Text von Anmeldungen sollen durch einen Neuausdruck der gesamten Seite erfolgen. Korrekturzeichen – wie etwa gemäß dem internationalen Standard ISO 5776 – sind nicht zulässig. Verbesserungen mit weißer Korrekturflüssigkeit, selbstklebenden Papierstreifen, Radierungen oder Durchstreichen sind nicht zulässig. Austauschseiten sollen nicht mit Standardauflösung per FAX an das Amt geschickt werden: Austauschseiten sind entweder in Papierform oder per FAX mit einer Auflösung von zumindest 300 dpi oder mittels einer Netzwerkübertragung zu übermitteln, die vom Amt unterstützt wird, vorausgesetzt jede Seite wurde mit einer Auflösung von zumindest 300 dpi gescannt.

EMPFEHLUNGEN FÜR ÄMTER

47. Das Verändern der erhaltenen Seiten durch die Patentämter soll vor der Weiterleitung zum Scannen und zur OCR-Verarbeitung vermieden werden. Zum Beispiel machen einige gegenwärtige Vorgangsweisen, die eine Stempelung umfassen, welche Zeichen auf der Seite überdecken könnten, den vom Anmelder eingereichten Text für OCR-Verfahren unlesbar. Wenn auf den Originalseiten Stempel / Änderungen angebracht werden müssen, hat das Amt dafür zu sorgen, dass die Änderungen nur innerhalb der Ränder der Dokumente, wie im Anhang 1 definiert, auftreten dürfen.

48. Zukünftig sollten die Patentämter das Erstellen von Papierformularen zur Kommunikation zwischen den Anmeldern und dem Amt vermeiden. Erfahrungsgemäß ist die Bereitstellung gesicherter Onlineformulare gegenüber dem Aufbau von Systemen zur Erkennung von Papierformularen zu bevorzugen. In Anbetracht der Erleichterung deren Erkennbarkeit, werden dennoch folgende Empfehlungen für die Gestaltung von Papierformularen gemacht:

- Linien aus kleinen Punkten als Hinweis für die Benutzer/innen, wo Text einzugeben wäre, sollen in Formularen nicht verwendet werden.
- Drop-out Farben sollen für Zeichenfelder verwendet werden (Hellgrau).
- Drop-out Farben sollen nicht für Bereiche verwendet werden, die zur Aufnahme von Inhalt mit Grautönen oder Farben dienen, wie etwa einer gescannten Unterschrift oder einer Zeichnung.

DURCHFÜHRUNG

49. Ämter, die das Einreichen von Patentanmeldungen im OCR-Format zulassen oder vorschreiben wollen, wird empfohlen, regelmäßig eine vollständige Anleitung mit der genauen Definition der erlaubten Zeichensätze und die genaue Angabe der zulässigen Papiergröße in ihren offiziellen Amtsblättern und auf ihren Websites zu veröffentlichen.

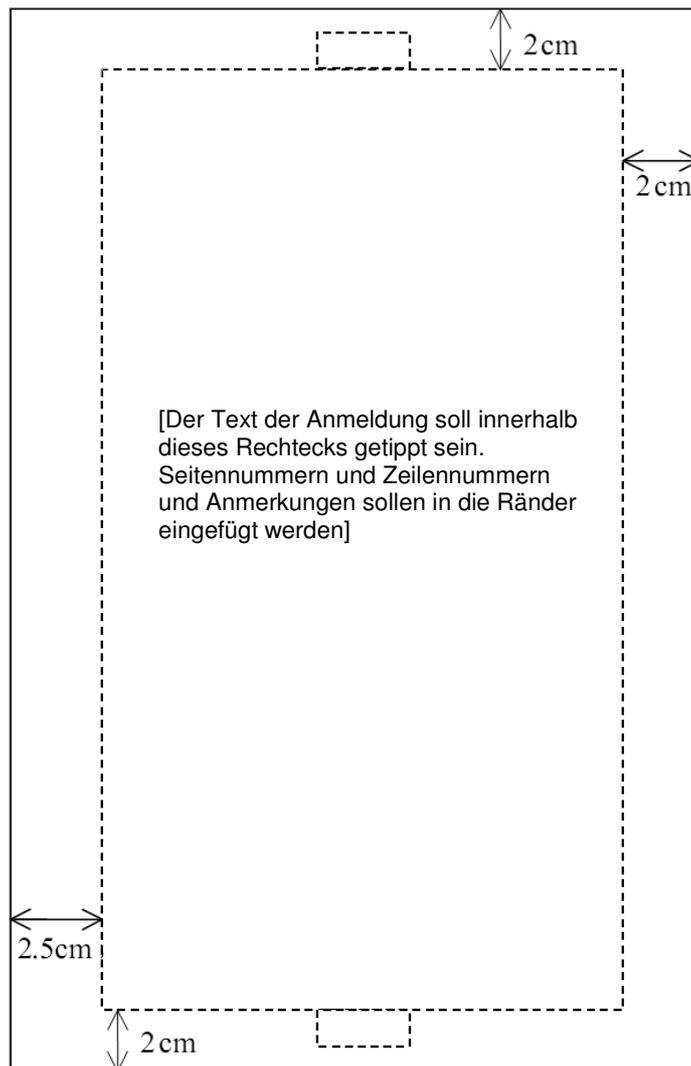
Beispiele

50. Im Anhang 2 dieser Empfehlung sind gute und schlechte Beispiele aus der Praxis im Hinblick auf OCR wiedergegeben. Die Beispiele zeigen gemeinsam mit einer kurzen Erklärung was zu tun bzw. zu unterlassen ist.

[Anhänge folgen]



ANHANG 1



Originalgröße = A4



ANHANG 2

GUTE UND SCHLECHTE BEISPIELE AUS DER PRAXIS

In diesem Anhang finden Sie gute und schlechte Beispiele von Patentdokumenten im Hinblick auf die erreichbare Genauigkeit bei der Durchführung von OCR-Verfahren auf diese.

*Gute Beispiele aus der Praxis*Beispiel 1: eine gute Beschreibungsseite

WO 2006/111319

PCT/EP2006/003401

Projection exposure system, method for manufacturing a micro-structured structural member by the aid of such a projection exposure system and polarization-optical element adapted for use in such a system

5

The invention relates to a projection exposure system, in particular for micro-lithography. The invention further relates to a method for manufacturing a micro-structured component and a polarization-optical element for the extreme ultraviolet (EUV) region.

10

For highest possible precision of the optical image to be obtained in complicated optical instruments such as a projection exposure system, the influence of the polarization of the light must be considered or, respectively, the polarization must be influenced specifically. For example, in particular

15

in case of great incidence angles, polarization effects occur in the mirror systems, which projection exposure systems in the EUV region are based on, for lack of suitable transparent materials. These polarization effects are in particular due to the varying reflectivity of the mirrors for s-polarized and p-polarized light and can give rise to imaging errors or other undesired

20

effects. Efforts have been made to measure possible polarization effects in the individual components of projection exposure systems.

25

For example, EP 1 306 665 A2 discloses an optical instrument for measuring polarization-dependent properties which comprises a light source in the EUV or X-radiation region and a rotatable polarizer. The polarizer is substantially comprised of a set of mirrors that reflects the incident light at least three times. The mirrors are arranged in such a way that the optical axes of the incident and emergent light are on the same straight line.

Anmerkung des Übersetzers: Auf die Zeilenummerierung sollte verzichtet und eine oben zentrierte Seitennummerierung eingefügt werden. Die Schriftart ist nicht optimal (besser nichtproportional, zB Courier New und größer, zB 14 Punkt). Siehe Absätze 14, 15 und 34 bis 36.



Anhang 2, Seite 2

Beispiel 2: eine gute Anspruchsseite

WO 2008/015644

PCT/IB2007/053030

5

CLAIMS

1. A sports or games apparatus (10), characterized in that it comprises a flexible grid (11) of rectangular shape, which has two base sides (12) and two height sides (13), and is formed from a plurality of grid elements (20, 30, 40), in which the said grid elements comprise a pair of rigid bars (20) forming respectively the said two height sides of the grid, capable of being attached to corresponding support structures (S); a pair of elastic base cords or tapes (30), forming respectively the said two base sides of the grid, with their ends (31) attached to the rigid bars; and a plurality of elastic grid cords or tapes (40), the ends of each of the said elastic grid cords or tapes being attached to another two corresponding grid elements, in such a way that, in an installed condition of the said apparatus, in which the said rigid bars are attached to the said support structures in such a way as to bring the said elastic base cords or tapes and the said elastic grid cords or tapes into tension, the said grid elements are positioned so as to form a grid with a rectangular mesh.
2. An apparatus according to Claim 1, in which the ends (31, 41) of each elastic base cord or tape (30) and of each elastic grid cord or tape (40) are movable along the direction of extension of the corresponding grid element to which they are attached.
3. An apparatus according to Claim 2, in which the ends (31, 41) of each elastic base cord or tape (30) and of each elastic grid cord or tape (40) are bent back to form a noose and attached to themselves, and the corresponding grid elements to which they are attached in a movable way are inserted into the corresponding nooses (42) formed by them.
4. An apparatus according to any one of the preceding claims, in which at least one elastic grid cord or tape (40) comprises at least one intermediate point (43) attached to another elastic grid cord or tape (40).
5. An apparatus according to Claim 4, in which at least one intermediate point is movable along the direction of extension of the corresponding elastic grid cord or tape (40) to which it is attached.

Alle Empfehlungen sind eingehalten: Ränder, Standardschriftart (Times New Roman), gute Schriftgröße, keine Zeilennummern, eingeschränkte Verwendung des Fettdrucks, keine Kursivschrift, kein unterstrichener Text, Anspruchsnummerierung angemessen und gut vom Text der Ansprüche beabstandet.

Anmerkung des Übersetzers: Auf den Blocksatz sollte verzichtet und eine größere Nichtproportionalschrift gewählt werden (zB Courier New, 14 Punkt). Siehe Absätze 30 und 34 bis 36.



Anhang 2, Seite 3

Beispiel 3: eine gute Beschreibungsseite mit komplexen Formeln

WO 2006/102655

PCT/US2006/011076

[0134] When performing the first iteration of step S9-4, the values of D_b , A_a , D_b and A_b are the values previously calculated at step S7-2, while all values of λ_n are zero.

[0135] The equations used by solver 244 at step S9-6 comprise the following in this embodiment:

$$\begin{aligned} \text{5} \quad & \text{if } (\lambda_{x,y,z-\max}^{n+1})_{\text{avg \& lln}} < 0 \text{ then } \lambda_{x,y,z-\max}^{n+1} = 0 & (46) \\ & \text{if } (\lambda_{x,y,z-\min}^{n+1})_{\text{avg \& lln}} > 0 \text{ then } \lambda_{x,y,z-\min}^{n+1} = 0 & (47) \\ & \lambda_{ln}^{n+1} = \lambda_{ln-\min}^{n+1} + \lambda_{ln-\max}^{n+1} & (48) \\ & \lambda_{\text{avg}}^{n+1} = \lambda_{\text{avg}-\min}^{n+1} + \lambda_{\text{avg}-\max}^{n+1} & (49) \end{aligned}$$

10 [0136] The equations used by solver 244 at step S9-8 comprise the following in this embodiment:

$$\begin{aligned} & D_a^{n+1} = D_a^n + L \frac{(\lambda_{ln}^{n+1} - \lambda_{ln}^n)}{m_a} & (50) \\ & A_a^{n+1} = A_a^n + I_a^{-1} [r_a^y] L (\lambda_{ln}^{n+1} - \lambda_{ln}^n) + I_a^{-1} T (\lambda_{\text{avg}}^{n+1} - \lambda_{\text{avg}}^n) & (51) \\ & D_b^{n+1} = D_b^n - L \frac{(\lambda_{ln}^{n+1} - \lambda_{ln}^n)}{m_b} & (52) \\ & A_b^{n+1} = A_b^n - I_b^{-1} [r_b^y] L (\lambda_{ln}^{n+1} - \lambda_{ln}^n) - I_b^{-1} T (\lambda_{\text{avg}}^{n+1} - \lambda_{\text{avg}}^n) & (53) \end{aligned}$$

15 [0137] Referring again to Figure 7, at step S7-6, solver 244 performs a convergence test. In this embodiment, solver 244 performs processing to determine whether the values of λ calculated for the current iteration differ from the values of λ calculated for the previous iteration by more than a predetermined threshold, in accordance with the following equation:

$$\sum_x \frac{(\lambda^{n+1} - \lambda^n)^2}{\lambda^{n2}} \leq \text{Threshold} \quad (54)$$

20 [0138] In this embodiment, the threshold employed in Equation (54) is set to 10^{-4} .

[0139] At step S7-8, solver 244 determines whether a predetermined number of iterations of the processing at steps S7-2 to S7-8 have been performed. In this embodiment, solver 244 determines whether 50 iterations have been performed.

Die Textabsätze enthalten keine unüblichen Zeichen. Die mathematischen Formeln sind korrekt eingebunden, wodurch den OCR-Maschinen eine einfache Aufteilung der eingebetteten Grafiken ermöglicht wird. Ein mögliches Ergebnis der Aufteilung ist durch blaue Rechtecke angedeutet.



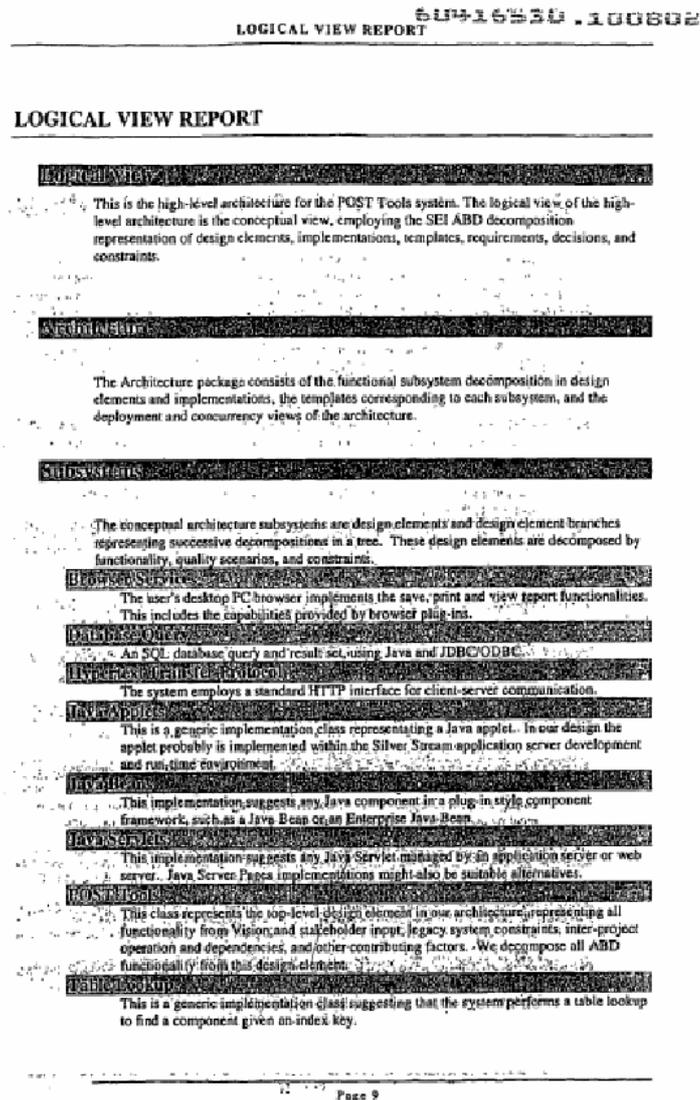
Anhang 2, Seite 4

Schlechte Beispiele aus der Praxis

Beispiel 1: eine dürrtige Seite mit zahlreichen Mängeln

WO 2005/060413

PCT/US2004/033203



Dieses Beispiel ist nicht im Einklang mit Absatz 10 (die Seite wurde vermutlich per FAX mit einer Auflösung von 200 dpi an das Amt übermittelt – siehe das Rauschen – und der Text ist teilweise auf dunkelgrauem Hintergrund). Ebensovienig erfüllt das Beispiel die Absätze 13 und 47: Das interne Aktenzeichen (604115530.100802) findet sich im Textbereich der Seite (es sollte im Rand sein). Die Seitennummerierung ist nicht korrekt (sollte „9“ nicht „page 9“ lauten, siehe Absatz 34). Schließlich ist die Schriftart zu klein (Absatz 15). Solche Seiten sollten idealer Weise von den Ämtern nicht angenommen und Ersatzblätter eingefordert werden (es ist unmöglich diese Seite korrekt mit OCR zu verarbeiten).



Anhang 2, Seite 5

Beispiel 2: eine Seite mit nicht weißem Hintergrund

WO 2005/097403

- 13 -

PCT/FR2005/050194

REVENDICATIONS

1. Dispositif d'usinage (D) du type de celui associant une machine-outil d'usinage (100) à un dispositif porte-pièce (200) équipé d'un axe de mise en mouvement de rotation transversal (A) par rapport à l'axe de plongée (Z), CARACTERISÉ PAR LE FAIT QUE le dispositif porte-pièce (200) est constitué par un bâti (210) supportant deux paliers de guidage (210 et 230) en rotation selon ledit axe de rotation transversal (A), la structure formée par le bâti (210) et les deux paliers (220 et 230) étant formée par la pièce à usiner (300) dont les extrémités viennent se fixer auxdits paliers (310 et 220), la pièce à usiner (300) étant une pièce longue du type de celle comportant des surfaces à usiner concentrées à ses deux extrémités ET PAR LE FAIT QUE la machine-outil (100) est du type de celle assurant la mise en mouvement de deux coulants porte-outil indépendants (110 et 120) de façon à ce que les usinages des deux extrémités de la pièce (300) soient réalisés par un coulant différent.

2. Dispositif d'usinage (D) selon la revendication 1, CARACTERISÉ PAR LE FAIT QUE chaque palier (220 et 230) comprend et guide un plateau tournant (221 et 231) équipé d'un moyen de mise en mouvement motorisé, la rotation des deux plateaux (221 et 231) étant synchronisée.

3. Dispositif (D) selon la revendication 2, CARACTERISÉ PAR LE FAIT QUE chaque plateau (220 et 230) est équipé de deux appuis (410, 420 et 510, 520) pour accueillir et maintenir en position l'extrémité de la pièce (300).

4. Dispositif (D) selon la revendication 1, CARACTERISÉ PAR LE FAIT QUE le bâti (210) du dispositif porte-pièce (200) est lui-même monté mobile en rotation selon un axe (B) perpendiculaire à l'axe (A) de rotation défini par les deux paliers (220 et 230) qu'il supporte.

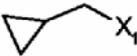
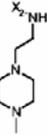
5. Procédé d'usinage d'une pièce longue (300) du type de celle comportant des surfaces à usiner concentrées à

Dieses Beispiel ist nicht im Einklang mit Absatz 12. Die Seite muss gefiltert werden, um zu versuchen, den verrauschten Hintergrund vor der Durchführung des OCR-Verfahrens zu entfernen. Wenn ein OCR-Vorgang in der vorliegenden Form erfolgt, ist der erhaltene Text unlesbar.



Anhang 2, Seite 6

Beispiel 3: eine Seite mit durchscheinenden Zeichen

#	R2	A	UV max [nm]:	MS (ESI) (M+H) ⁺	
25			305, 350	476	Trihyc 1,41 (m, 2f) (m, 1f)

Beispiele 26-40

Die folgenden Verbindungen sind über ein analoges Verfahren beschrieben, hergestellt. Die Herstellung des Benz

5 beschrieben. Das für die Darstellung des Amids ei

Ein kleiner Bereich der Seite ist vergrößert, um die Zeichen zu zeigen: die Farbe des Originaltextes ist vermutlich Grau, was nach dem Scannen mit 300 dpi Schwarz-Weiß zu Zeichen führt, die nicht ausgefüllt sind. Daraus ergibt sich eine dürftige OCR-Genauigkeit (dieses Beispiel ist nicht im Einklang mit Absatz 12).

Beispiel 3: eine Seite mit handschriftlichem Text

TITLED : JIG HEAD SWAY BAR

BACK GROUND

IN THE ART OF FISHING THERE IS A PIECE OF TACKLE KNOWN AS A PIVOT-HEAD JIG WHICH USES SPECIALIZED OR SPECIFICALLY SHAPED HOOKS TO PROVIDE AN ACTION PRODUCING LURE COMBINATION. MY INVENTION THE SWAYBAR ALLEVIATES THIS NEED FOR SPECIAL HOOKS BY BEING ABLE TO BOTH SUPPORT THE JIG HEAD AND ALLOW FOR CONNECTION OF OTHER REQUIRED TACKLE

Wie zu erwarten ist, ergibt sich bei einem OCR-Verfahren aufgrund dieser Seite ein unlesbarer Text. Die Ämter sollten maschineschriebene Texte verlangen, um eine minimale Qualität der Veröffentlichung sicherzustellen.

Anhang 2, Seite 7

Beispiel 5: eine Seite mit empfehlungswidrigem Layout und weiteren Mängeln

WO 2005/086760

PCT/US2005/007335

relation to the determination of AN by FTIR spectroscopy

This concept is illustrated in Figure 1 for AN, the BN analysis being analogous but using a different reagent. Differential spectroscopy is then used to eliminate the spectral contributions from the base oil and any additives and/or contaminants and breakdown products present in the oil that may spectrally interfere with the measurement of the signal from the reaction product. This is achieved by treating a portion of the sample with a blank reagent, this portion effectively serving as a reference oil. Figure 2 illustrates the general analytical protocol.

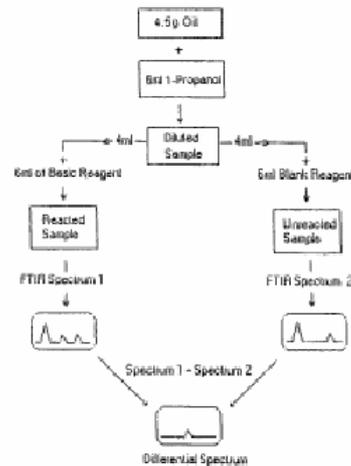


Figure 2. Analytical protocol for the determination of AN by FTIR spectroscopy.

In this procedure, the sample is first diluted with an innocuous solvent (1-propanol), then split and treated with a

reactive and a blank reagent to produce two samples for spectral analysis. Since these two samples are the same except for the reaction products, subtraction of their spectra leaves only the spectral contribution related to AN.

The COAT AN/BN Analyzer

The COAT AN/BN Analyzer has been designed and programmed to automate AN/BN analyses based on the concepts laid out above. Figure 3 illustrates key components of the COAT AN/BN Analyzer: an FTIR spectrometer, a sample handling accessory, an autosampler, and the computer that controls the system.

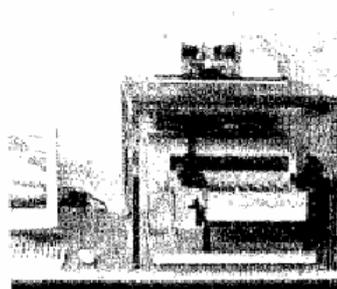


Figure 3. The COAT AN/BN Analyzer and its key components

The compact nature of the sample handling system is made possible by the dilution of the sample in the analytical protocol (Figure 2), allowing a micropump to be substituted for the peristaltic pump employed in most FTIR used oil analyzers. The resulting low viscosity of the sample dramatically

Diese Seite ist nicht im Einklang mit den folgenden Empfehlungen: Absatz 17 (einspaltiges Textformat), Absatz 39 (Kursiv- und Fettdruck), Absatz 46 (manuelle Korrekturen nach dem Ausdruck). Der Blocksatz ist ebenfalls empfehlungswidrig (Absatz 30), obwohl in diesem Fall dadurch keine negativen Auswirkungen auf das OCR-Verfahren auftreten würden, da die Wortabstände ausreichend sind. Ebensovienig erfüllt das Beispiel schließlich Absatz 27 (Zeilenabstand 1,5).



Anhang 2, Seite 8

Beispiel 6: eine Seite mit zu kleinen Zeilennummern

WO 2004/110497

PCT/US2004/013820

- 5 [0028] Figs. 9A-9B are plots showing the percent of mitomycin C released from liposomes comprised of HSPC/mPEG-DSPE/lipid-DTB-mitomycin C (Fig. 9A) and HSPC/cholesterol/mPEG-DSPE/lipid-DTB-mitomycin C (Fig. 9B) as a function of time of incubation in the presence of cysteine at concentrations of 150 μ M (closed symbols) and at 1.5 mM (open symbols);
- 10 [0029] Fig. 10 is a plot of growth rate of M109 cells, expressed as a percentage based on growth of M109 cells in the absence of drug and cysteine, as a function of mitomycin C amount, in nM, for free mitomycin c (open triangles), liposomes comprised of HSPC/mPEG-DSPE/lipid-DTB-mitomycin C (closed squares), and liposomes comprised of HSPC/cholesterol/mPEG-DSPE/lipid-DTB-mitomycin C (open circles);
- 15 [0030] Fig. 11A is a plot of growth rate of M109 cells, expressed as a percentage based on growth of M109 cells in the absence of drug or cysteine, as a function of mitomycin C concentration in nM. Shown are cells treated mitomycin C in free form (open triangles) and with mitomycin C in free form plus 1000 μ M cystein (closed triangles). Also shown are cells treated with the liposome formulation comprised of HSPC/PEG-DSPE/lipid-DTB-mitomycin C (open circles) and with the liposome formulation with additional cysteine added at concentrations of 150 μ M (open diamonds), 500 μ M (closed circles) and 1000 μ M (open squares);
- 20 [0031] Fig. 11B is a plot of growth rate of M109 cells, expressed as a percentage based on growth of M109 cells in the absence of drug or cysteine, as a function of mitomycin C concentration in nM. Shown are cells treated mitomycin C in free form (open triangles) and with mitomycin C in free form plus 1000 μ M cysteine (closed triangles). Also shown are cells treated with the liposome formulation comprised of
- 25 HSPC/cholesterol/mPEG-DSPE/lipid-DTB-mitomycin C (open circles) and with the liposome formulation with additional cysteine added at concentrations of 150 μ M (open diamonds), 500 μ M (closed circles) and 1000 μ M (open squares);
- 30 [0032] Fig. 12 is a plot showing the percent increase in cytotoxicity (as determined by $(IC_{50_{no\ cysteine}}/IC_{50_{cysteine}}) \times 100$) of free mitomycin C (closed squares), mitomycin C associated with liposomes comprised of HSPC/cholesterol/mPEG-DSPE/lipid-DTB-mitomycin C (closed circles), and liposomes comprised of HSPC/mPEG-DSPE/lipid-DTB-mitomycin C (open triangles) to M109 cells *in vitro* at various concentrations of cysteine;
- [0033] Fig. 13A is a plot showing the concentration of mitomycin C in the blood of

Zeilennummern führen bei OCR-Maschinen zu mehreren Schwierigkeiten (Absatz 14):

- Sie können gegenüber der zugeordneten Zeile versetzt sein, was zu Problemen bei der Erkennung der Grundlinie führt.
- Sie können zu klein sein, was zu Erkennungsfehlern führt, welche das XML-Umwandlungsverfahren daran hindert, die Zeilennummern richtig aus dem Textkörper der Seite zu entfernen.
- Sie können innerhalb des Textkörperbereichs falsch angeordnet sein oder innerhalb des Seitenrandesandes zu nahe am Textkörperbereich, wodurch sie nach der OCR-Ausgabe innerhalb des Fließtextes erscheinen.

In diesem Beispiel sind sie zu klein.

Die tiefgestellten Zeichen in diesem Beispiel sind zu klein (Absatz 35).



Anhang 2, Seite 9

Beispiel 7: eine Seite die Text in mehreren Ausrichtungen enthält

WO 2005/081642

PCT/JP2005/003688

Table 11 (continued-2)

	Amount in retardation-controlling agent solution (mass parts)					Amount in UV absorber solution (mass parts)						Mixing ratio of solutions			
	Retardation-control agent A-2	Retardation-control agent A-12	Retardation-control agent B	Retardation-control agent C	Retardation-control agent D	UV absorber A	UV absorber B	UV absorber C	UV absorber D	UV absorber E	UV absorber F	Cellulose acetate solution	Matting agent solution	Retardation-controlling agent solution	UV absorber solution
This invention			6	14							15	94.6	1.2	6.2	1.3
This invention	3		3	14							15	94.6	1.2	7.0	3.2
This invention	5		5	10							15	94.6	1.2	6.2	0.8
This invention		59.8			15	4.8	10.2					94.6	1.2	6.2	0.8
This invention			10		10	4.8	10.2					94.6	1.2	6.2	0.8
This invention					15	4.8	10.2					94.6	1.2	6.2	0.8
Comparative example	10		10									94.6	1.2	6.6	0
Comparative example				20								94.6	1.2	4.1	0
Comparative example								5	10			94.6	1.2	0	6.3
Comparative example	10		10			10.5	4.5					94.6	1.2	7.1	0.8
Comparative example	10		10			10.5	4.5					94.6	1.2	7.1	0.8

Dieses Beispiel ist nicht im Einklang mit Absatz 20:

Eine der Einschränkungen der besten derzeit verfügbaren OCR-Maschinen ist, dass diese nur Text von einzelnen Seiten lesen können, der in eine Richtung ausgerichtet ist (ein Vorverfahren bestimmt die Hauptausrichtung des Textes auf der Seite). Daraus ergibt sich, dass alle Worte unberücksichtigt bleiben, die nicht in der Hauptausrichtung verlaufen. Es ist natürlich zulässig, auf einer Seite mit Anmerkungen im Rand, die im Hochformat ausgerichtet sind (Seitennummer, interne Aktenzeichen etc.), eine Tabelle im Querformat oder sogar den Textkörper im Querformat anzuordnen.



Anhang 2, Seite 10

Beispiel 8: eine Seite mit gemischten mathematischen Formeln und Fließtext

WO 2005/116630

PCT/US2005/017216

$$\Delta L = \frac{\hbar}{2} - r \times eA \tag{33}$$

$$= \left[\frac{\hbar}{2} - \frac{e\phi}{2\pi} \right]_z \tag{34}$$

In order that the change of angular momentum, ΔL , equals zero, ϕ must be $\Phi_0 = \frac{h}{2e}$, the magnetic flux quantum. The magnetic moment of the electron is parallel or

5 antiparallel to the applied field only. During the spin-flip transition, power must be conserved. Power flow is governed by the Poynting power theorem,

$$\nabla \cdot (\mathbf{E} \times \mathbf{H}) = -\frac{\partial}{\partial t} \left[\frac{1}{2} \mu_0 \mathbf{H} \cdot \mathbf{H} \right] - \frac{\partial}{\partial t} \left[\frac{1}{2} \epsilon_0 \mathbf{E} \cdot \mathbf{E} \right] - \mathbf{J} \cdot \mathbf{E} \tag{35}$$

Eq. (36) gives the total energy of the flip transition which is the sum of the energy of reorientation of the magnetic moment (1st term), the magnetic energy (2nd term), the electric energy (3rd term), and the dissipated energy of a fluxon treading the orbitsphere (4th term), respectively,

$$\Delta E_{\text{avg}}^{\text{spin}} = 2 \left(1 + \frac{\alpha}{2\pi} + \frac{2}{3} \alpha^2 \left(\frac{\alpha}{2\pi} \right) - \frac{4}{3} \left(\frac{\alpha}{2\pi} \right)^3 \right) \mu_B B \tag{36}$$

$$\Delta E_{\text{mag}}^{\text{spin}} = g \mu_B B \tag{37}$$

15 where the stored magnetic energy corresponding to the $\frac{\partial}{\partial t} \left[\frac{1}{2} \mu_0 \mathbf{H} \cdot \mathbf{H} \right]$ term increases,

the stored electric energy corresponding to the $\frac{\partial}{\partial t} \left[\frac{1}{2} \epsilon_0 \mathbf{E} \cdot \mathbf{E} \right]$ term increases, and the

$\mathbf{J} \cdot \mathbf{E}$ term is dissipative. The spin-flip transition can be considered as involving a magnetic moment of g times that of a Bohr magneton. The g factor is redesignated the fluxon g factor as opposed to the anomalous g factor. Using $\alpha^{-1} = 137.03603(82)$, the

20 calculated value of $\frac{g}{2}$ is 1.001 159 652 137. The experimental value [23] of $\frac{g}{2}$ is 1.001 159 652 188(4).

1.G. SPIN AND ORBITAL PARAMETERS

The total function that describes the spinning motion of each electron orbitsphere

25 is composed of two functions. One function, the spin function, is spatially uniform over the orbitsphere, spins with a quantized angular velocity, and gives rise to spin angular momentum. The other function, the modulation function, can be spatially uniform—in which case there is no orbital angular momentum and the magnetic moment of the electron orbitsphere is one Bohr magneton—or not spatially uniform—in which case

30 there is orbital angular momentum. The modulation function also rotates with a quantized angular velocity.

The spin function of the electron corresponds to the nonradiative $n = 1, \ell = 0$

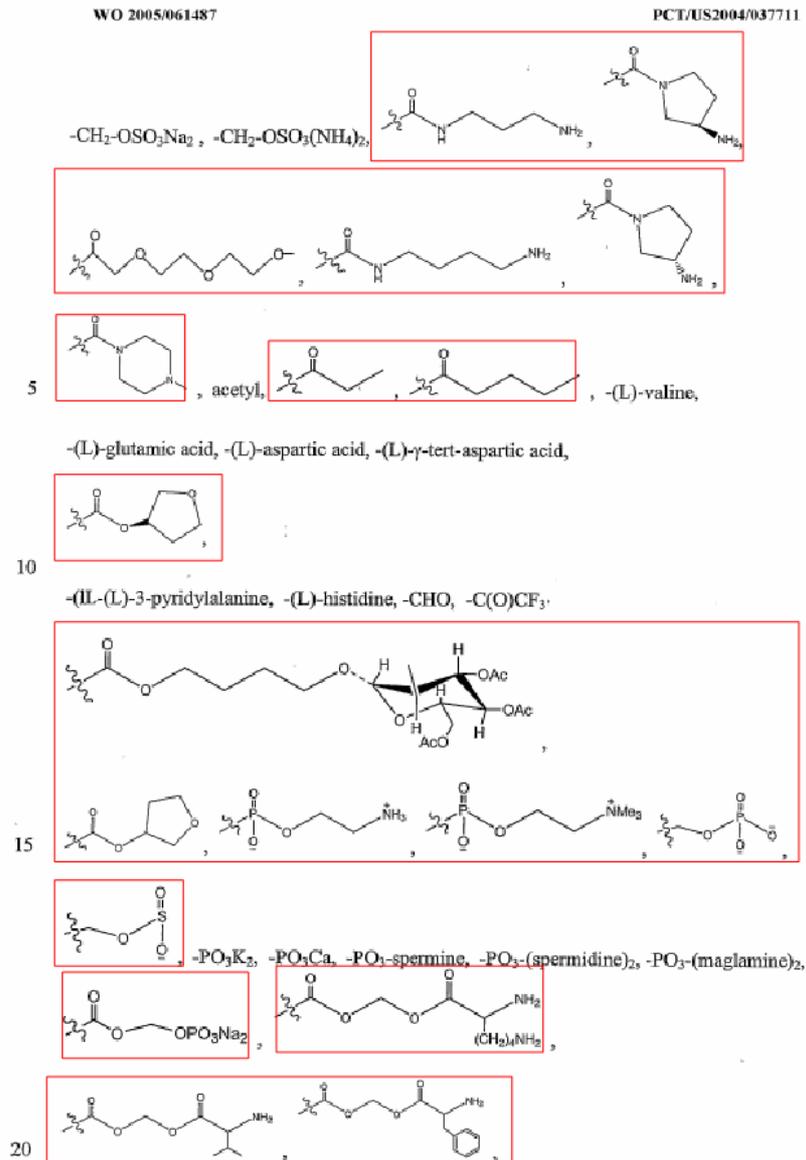
Dieses Beispiel ist nicht im Einklang mit Absatz 23. Die OCR-Maschine ist nicht in der Lage, den Text richtig von den Formeln zu trennen (siehe die rote manuelle Trennung: sogar die eingebetteten Formeln überschneiden einander).

Allgemeiner Hinweis zu diesem Beispiel: der Text und die Formeln liegen zu eng beisammen, um eine gute Erkennung zu ermöglichen; ebenso ist das Beispiel nicht im Einklang mit den Absätzen 27 und 28.

Dieses Beispiel verwendet auch unübliche Zeichen: Griechische Zeichen können verwendet werden, obwohl sie die Erkennung der Seite erschweren (siehe Ansatz 40). Dennoch wird nachdrücklich empfohlen, keine kursiven, fetten oder unterstrichenen Zeichensätze mit unüblichen Zeichen zu kombinieren (Absatz 39).

Anhang 2, Seite 11

Beispiel 9: eine Seite mit gemischten chemischen Formeln und Fließtext



Dieses Beispiel ist nicht im Einklang mit Absatz 23. Das Ergebnis der erwarteten (manuell erfolgten) Aufteilung der Zeichnungen ist rot dargestellt. Diese Aufteilung kann von einer OCR-Maschine nicht erreicht werden, da die Formeln zu eng am umgebenden Text liegen.



Anhang 2, Seite 12

Beispiel 10: eine Seite mit zu kleinen tiefgestellten Zeichen

WO 2005/110416

PCT/US2005/015897

R₁ is hydrogen, C₁-C₆alkyl, C₂-C₆alkenyl, C₂-C₆alkynyl, C₁-C₆alkoxy, C₁-C₆haloalkyl, C₁-C₆haloalkoxy, (C₃-C₇cycloalkyl)C₂-C₆alkyl;

R₃ is selected from alkoxy, cycloalkoxy, phenyl, 4- to 7-membered heterocycles, -O(CH₂)_nphenyl, -O(CH₂)_npyridyl, -E-(CR_AR_B)_n-Q, and Q, each of which is substituted with between 0 and 3 substituents selected from halogen, cyano, hydroxy, oxo, (CR_AR_B)-T, C₁-alkyl, C₁-alkoxy, C₁-haloalkyl, C₁-haloalkoxy, mono- and di-(C₁-alkyl)amino, (C₁-alkyl)((CR_AR_B)-T)amino, benzyl, S(O)_n(C₁-alkyl), α,ω-C₁-alkylene, α,ω-C₁-alkyleneoxy, α,ω-C₁-alkylenedioxy, -E-(CH₂)_n-Q, and Q;

T is CO₂H, CONH₂, C₁-alkoxycarbonyl, mono- or di-(C₁-alkyl)aminocarbonyl, SO₃H, SO₂NH₂ or SO₂(C₁-alkyl);

j is an integer ranging from 0 to 6;

Q is a saturated heterocyclic ring comprising between 4 and 7 ring members, in which the point of attachment is a carbon or nitrogen atom;

E is O, NR_D, or a single covalent bond;

R₄ and R₉ are independently chosen from hydrogen, halogen, hydroxy, C₁-C₆alkyl, C₁-C₆alkenyl, (C₃-C₆cycloalkyl)C₂-C₆alkyl and C₁-C₆alkoxy; and

Ar is phenyl which is mono-, di-, or tri-substituted; or 1-naphthyl, 2-naphthyl, pyridyl, pyrimidinyl, pyrazinyl, pyridizynyl, thienyl, thiazolyl, pyrazolyl, imidazolyl, tetrazolyl, oxazolyl, isoxazolyl, pyrrolyl, furanyl, indolyl, indazolyl, or triazolyl, each of which is optionally mono-, di-, or tri-substituted.

Yet other compounds of Formula VIII include those compounds in which the group designated:



is chosen from naphthyl, tetrahydronaphthyl, benzofuranyl, benzodioxolyl, indanyl, indolyl, indazolyl, benzodioxolyl, benzo[1,4]dioxanyl and benzoxazolyl, each of which is substituted with from 0 to 3 substituents independently chosen from R₂.

Certain compounds of Formula IX include those in which

Ar is mono-, di-, or tri-substituted phenyl, which phenyl group is substituted with one to three substituents independently chosen from hydroxy, halogen, cyano, amino, nitro, -COOH, aminocarbonyl, -SO₂NH₂, C₁-alkyl, C₁-alkenyl, C₁-alkynyl, C₁-haloalkyl, C₁-aminoalkyl, C₁-hydroxyalkyl, C₁-carboxyalkyl, C₂-alkoxy, C₁-haloalkoxy, C₁-alkylthio, C₁-alkanoyl, C₁-alkanoyloxy, C₂-alkanone, C₁-alkyl ether, mono- or di-(C₁-alkyl)aminoC₂-alkyl, -NHC(=O)(C₁-alkyl), -(C₁-alkyl)C(=O)(C₁-alkyl), -NHS(O)_n(C₁-alkyl), -(C₁-alkyl)C(=O)NH₂, -(C₁-alkyl)C(=O)NH(C₁-alkyl), -(C₁-alkyl)C(=O)NH(C₁-alkyl)(C₁-alkyl), -S(O)_n(C₁-alkyl), -S(O)_nNH(C₁-alkyl), -S(O)_nN(C₂-alkyl)(C₁-alkyl) and Z; or

Dies ist ein typisches Beispiel, in dem die tiefgestellten Zeichen zu klein für eine genaue Erkennung sind. Dieses Phänomen ist bei Patenten am Gebiet der Chemie häufig festzustellen.



Anhang 2, Seite 13

Beispiel 11: eine Seite mit schlecht formatierten Tabellen

WO 2005/063765

PCT/US2004/043492

Table D

Other compounds of the invention result from selecting appropriate features from the table of possible features below. For example, compound A77 results from the following selections: none-morpholino-aryl-OCH₂(CO)-piperazine-CH₃.

Left-hand substituent	Left-hand ring	Aryl or heteroaryl	Ring substituent	Nitrogen feature	Right-hand substituent
CH ₃	morpholino	aryl	OCH ₂	NHM	alkyl
isopropyl	piperazine	thiophene	OCH ₂ (CO)	NMM	alkoxy
CH ₃ CH ₂ O(CO)CH ₂			SO ₂	morpholino	alcohol
none			OCH ₂ (CO)OCH ₂	piperazine	substituted amine
				piperidine	acid
				pyrazole	ester
				pyrrolidine	CH ₂ CH ₂ OCH ₃
					CH ₂ CH ₂ OH
					CH ₂ NH ₂
					CH ₂ NHCH ₂ CH ₂ CH ₃
					CH ₂ NHCH ₃
					CH ₂ NHCHCH ₃ CH ₃
					CH ₃
					CHCH ₂ CH ₃
					COOCH ₂ CH ₃
					none

Table E

Other compounds of the invention result from selecting appropriate features from the table of possible features below. For example, compound B3 results from the following selections: none-morpholino-aryl-CH₂-piperazine-CH₂CH₂OH.

Left-hand substituent	Left-hand ring	Aryl or heteroaryl	Ring substituent	Nitrogen feature	Right-hand substituent
CH ₃	morpholino	aryl	CH ₂	NHM	alkyl
isopropyl	piperazine	thiophene	CH ₂ CH ₂	NMM	alkoxy
CH ₃ CH ₂ O(CO)CH ₂			CH ₂ CH ₂ CH ₂	morpholino	alcohol
none			CH ₂ CH ₂ CH ₂ CH ₂	piperazine	substituted amine
				piperidine	acid
				pyrazole	ester
				pyrrolidine	CH ₂ CH ₂ OCH ₃
					CH ₂ CH ₂ OH
					CH ₂ NH ₂
					CH ₂ NHCH ₂ CH ₂ CH ₃
					CH ₂ NHCH ₃
					CH ₂ NHCHCH ₃ CH ₃
					CH ₃

In diesem Beispiel fehlen die Tabellenbegrenzungen (ist nicht im Einklang mit Absatz 33). Daraus ergibt sich, dass die OCR-Maschine versuchen wird, den Inhalt der Tabelle als Absatz zu erkennen. Dies führt zu mehreren weiteren Problemen:

- Die Zeichengröße in der Tabelle ist zu klein (Absätze 34 und 35)
- Die Grundlinien der Spaltenüberschriften werden vertauscht (Absatz 29). In der Folge wird die Maschine falsche Hoch- und Tiefstellungen erkennen.
- Der ausgegebene Text wird keine Rücksicht auf die Spalten nehmen:

Left-hand Substituent ring CH3 ...
 Left-Hand ring morpholino
 Aryl or heteroaryl aryl
 Nitrogen feature Ring substituent
 Right-hand substituent



Anhang 2, Seite 14

Beispiel 12: eine ausgerichtete Seite (Blocksatz)

WO 2005/087962

1

PCT/EP2005/002268

GKSS-Forschungszentrum Geesthacht GmbH, Max-Planck-Strasse 1, 21502 Geesthacht

Verfahren zur Herstellung von Profilen aus Leichtmetallwerkstoff mittels Strangpressen

Beschreibung

Die Erfindung betrifft ein Verfahren zur Herstellung von Profilen aus Leichtmetallwerkstoff, insbesondere Magnesiumwerkstoff, mittels Strangpressen, bei dem ein Werkstoffvolumen durch eine Matrize, die die Form des gewünschten Profils bestimmt, zur Ausbildung des Profils gepreßt wird.

Die Herstellung von Profilen aus Leichtmetall- bzw. Leichtmetall-Legierungswerkstoffen mittels eines Strangpreßverfahrens ist eine allgemein eingeführte, bekannte Technologie und wird industriell angewendet. So ist es bekannt, daß konventionell verfügbare Leichtmetall- bzw. Leichtmetall-Knetlegierungen in Form von Gußblöcken durch konventionelles Strangpressen in Profilformen gepreßt werden. Dabei wird der Leichtmetall- bzw. Leichtmetall-Legierungsblock, im folgenden zusammenfassend kurz mit Werkstoffvolumen bezeichnet, bei Temperaturen

In diesem Beispiel sind die Absätze rechts und links ausgerichtet (Blocksatz). Auch wenn dies dem Text ein ästhetisches Aussehen verleiht, erschwert es manchmal die OCR-Verfahren, wenn die Wortabstände zu klein werden (Absatz 30). Ebenso ist dieses Beispiel nicht im Einklang mit Absatz 31, der festlegt, dass eine Wortabteilung am Zeilenende möglichst vermieden werden sollte (die OCR Maschine hat manchmal Schwierigkeiten, zwischen Bindestrichen und Trennzeichen zu unterscheiden, was zu unerwünschten Bindestrichen in der Ausgabe führt).



Anhang 2, Seite 15

Beispiel 13: eine Tabelle mit mangelhaften Begrenzungslinien

WO 2004/110415

- 60 -

PCT/EP2004/051048

Comp. No.	Exp. No.	AIK*	Y	AIK ^b	L	Physical data
106	B2	cb	C=O	-CH ₂ -		2R-trans
107	B3b	cb	C=O	-CH ₂ -		2R-trans
108	B8	cb	C=O			2R-trans, HCl(1:3); H ₂ O(1:1)
109	B2	cb	C=O			2R-trans HCl(1:2) H ₂ O(1:1)
110	B3b	cb	C=O			2R-trans
111	B2	cb	C=O			[2R-(2a,4B(E))] 2R-trans

In diesem Beispiel sind die Begrenzungslinien der Tabelle im Original unzureichend. Nach dem Scannen kann das OCR-Verfahren die Tabelle nicht richtig erkennen und zur Aufteilung der Seite ist eine manuelle Nachbearbeitung nötig. Wenn eine derartige Seite nicht nachbearbeitet wird, enthält der Ausgabertext unerwünschte Zeichen, wodurch die Indizierung des Dokuments durch Suchmaschinen ineffizienter wird.



Anhang 2, Seite 16

Beispiel 14: mangelhafte hoch- und tiefgestellte Zeichen

WO 2005/100305

PCT/IB2005/000872

-9-

thiazolyl, pyrazolyl, pyridinyl, pyrimidinyl, purinyl, quinolinyl, benzofuran and isoquinolinyl.

- 5 p. "heteroaryl, optionally substituted," refers to a heteroaryl moiety as defined immediately above, in which up to 4 carbon atoms of the heteroaryl moiety may be substituted with a substituent, each substituent is independently selected from the group consisting of halogen, cyano, hydroxy, (C₁-C₈)alkyl, (C₁-C₈)alkoxy, (C₁-C₂)alkyl substituted with one or more halogens, (C₁-C₂)alkoxy substituted with one or more halogens, SR^a, and NR^aR^b, in which R^a and R^b are as defined above.
- 10 q. "heterocycle" or "heterocyclic ring" refers to any 3- or 4-membered ring containing a heteroatom selected from oxygen, nitrogen and sulfur; or a 5-, 6-, 7-, 8-, 9-, or 10- membered ring containing 1, 2, or 3 nitrogen atoms; 1 oxygen atom; 1 sulfur atom; 1 nitrogen and 1 sulfur atom; 1 nitrogen and 1 oxygen atom; 2 oxygen atoms in non-adjacent positions; 1 oxygen and 1 sulfur atom in non-adjacent positions; or 2 sulfur atoms in non-adjacent positions. The 5-membered ring has 0 to 1 double bonds, the 6- and 7-membered rings have 0 to 2 double bonds, and the 8, 9, or 10 membered rings may have 0, 1, 2, or 3 double bonds. The term "heterocyclic" also includes bicyclic groups in which any of the above heterocyclic rings is fused to a benzene ring, a cyclohexane or cyclopentane ring or another heterocyclic ring (for example, indolyl, quinolyl, isoquinolyl, tetrahydroquinolyl, benzofuryl, dihydrobenzofuryl or benzothienyl and the like). Heterocyclics include: pyrrolidinyl, tetrahydrofuranlyl, tetrahydrothiophenyl, piperidinyl, piperazinyl, azepane, azocane, morpholinyl, isochromanlyl and quinolinyl.
- 15 r. "heterocyclic, optionally substituted" refers to a heterocyclic moiety as defined immediately above, in which up to 4 carbon atoms of the heterocycle moiety may be substituted with a substituent, each substituent is independently selected from the group consisting of halogen, cyano, hydroxy, (C₁-C₈)alkyl, (C₁-C₈)alkoxy, (C₁-C₂)alkyl substituted with one or more halogens, (C₁-C₂)alkoxy substituted with one or more halogens, SR^a, and NR^aR^b, in which R^a and R^b are as defined above. Any nitrogen atom within such a heterocyclic ring
- 20
- 25
- 30
- 35

In diesem Beispiel treten die folgenden Probleme auf (Absatz 35):

- Die hoch- und tiefgestellten Zeichensind zu klein.
- Tiefgestellte Zeichen sind zu tief bezüglich der Grundlinie der Zeile angeordnet.
- Hochgestellte Zeichen sind zu hoch bezüglich der Grundlinie der Zeile angeordnet.

Dadurch werden die Zeilen 33 und 34 des Beispiels vom OCR-Verfahren wie folgt erkannt:

„Substituted with one or more halogens, (C -C)alkoxy substituted
1 2
8 8 9 8 9
with one or more halogens, SR , and NR R , in which R and R are“



Anhang 2, Seite 17

Beispiel 15: ein Beispiel mit unüblichen Zeichen

WO 2006/057705 **PCT/**

...

c = speed of sound in water;

\tilde{z}_u = initial altitude for beam pair u;

$\Delta\varepsilon_{z,u} = \varepsilon_{z,p+1,u} - \varepsilon_{z,p,u}$ = comparable to sway-reduced altitude difference;

$\Delta\varepsilon_{\gamma,u} = \varepsilon_{\gamma,p+1,u} - \varepsilon_{\gamma,p,u}$ = comparable to sway-reduced horizontal displacement;

5 $\varepsilon_{z,p,u}$ = difference of vertical linearization point in ping p, beam pair u, from nominal \tilde{z}_u ;

$\varepsilon_{z,p+1,u}$ = difference of vertical linearization point in ping p+1, beam pair u, nominal \tilde{z}_u ;

$\varepsilon_{\gamma,p,u}$ = difference of horizontal-range sample v linearization point in ping p

10 u, from the nominal $\gamma_{v,u}$. Note that this is the same for all horizontal samples;

$\varepsilon_{\gamma,p+1,u}$ = difference of horizontal-range sample v linearization point in ping pair u, from the nominal $\gamma_{v,u}$. Note that this is the same for all horizontal samples;

15 $\gamma_{v,u}$ = nominal horizontal offset to horizontal-range sample u for beam pair

Bei diesem Beispiel treten die folgenden Probleme auf:

- Unübliche Zeichen: kursive griechische Buchstaben und sogar Zeichen mit einer Tilde.
- Die Tiefstellungen sind auch hier zu klein.

Mit den meisten OCR-Maschinen werden alle unüblichen Zeichen nicht richtig erkannt.



Anhang 2, Seite 18

Beispiel 16: ein Beispiele mit engen Schriftzeichen und engem Zeichenabstand

WO 2006/036330

PCT/US2005/028798

23. The method of claim 18, wherein the data is encoded onto the representative transmission symbol by using a modulation method selected from a group consisting of: amplitude modulation, phase modulation, frequency modulation, single-sideband modulation, vestigial-sideband modulation, quadrature amplitude modulation, orthogonal frequency division modulation, pulse-code modulation, pulse-width modulation, pulse-amplitude modulation, pulse-position modulation, pulse-density modulation, frequency-shift keying, and phase-shift keying.
24. The method of claim 18, wherein each of the at least two communication signals is transmitted through a communication medium selected from a group consisting of: a wire medium, a wireless medium, an optical fiber ribbon, a fiber optic cable, a single mode fiber optic cable, a multi-mode fiber optic cable, a twisted pair wire, an unshielded twisted pair wire, a plenum wire, a PVC wire, and a coaxial cable.
25. The method of claim 18, wherein the at least two communication signals are both transmitted wirelessly.
26. The method of claim 18, wherein the at least two communication signals are both transmitted through a wire medium.
27. The method of claim 18, wherein the at least two communication signals are transmitted through a wire medium, and wirelessly.

Dieses Beispiel ist nicht im Einklang mit den Absätzen 37 und 38. Dadurch kann die OCR-Maschine die Wortgrenzen nicht richtig erkennen, was zu einem völlig unbrauchbaren OCR-Ergebnis führt.



Anhang 2, Seite 19

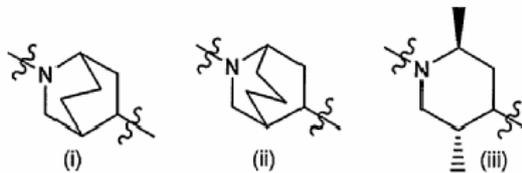
Beispiel 17: schlechte Stempelung vor dem Scannen durch das Anmeldeamt

WO 2006/058294

PCT/US2005/042931

Reaction of a piperidine ring with a reagent such as diborane or alkylation of the piperidine nitrogen with an alkyl halide or sulfonate ester provides the desired compounds.

[00176] Additionally, compounds of formulae (I, Ia, and Ib) in which the piperidine ring is replaced by:



Dieses Beispiel ist nicht im Einklang mit Absatz 47. Dadurch können die ersten sechs Worte der Seite vom OCR-Verfahren nicht gelesen werden. Darüber hinaus werden durch den Stempel zusätzliche falsche Zeichen eingefügt, welche die Indizierungsmaschinen verderben, falls die Qualität der Seite nicht manuell überprüft wird.



Anhang 2, Seite 20

Beispiel 18: eine weitere Seite mit schlechtem Layout mathematischer Formeln

WO 2006/079181

24

PCT/AU2006/000108

probability of the statistical outlier event of a noise only FFT bin magnitude being larger than a FFT bin containing both signal and noise is negligible.

Define,

$$\alpha = \sum_{n=0}^{N-1} r[n] \exp[-j2\pi(\frac{\hat{f}}{f_s} - \frac{1}{2N})n] \quad (9)$$

5

$$\beta = \sum_{n=0}^{N-1} r[n] \exp[-j2\pi(\frac{\hat{f}}{f_s} + \frac{1}{2N})n] \quad (10)$$

Then the discriminant, or distance metric, of frequency estimation error is defined as,

$$D(\varepsilon, \hat{\varepsilon}) = \frac{|\beta| - |\alpha|}{|\beta| + |\alpha|} \quad (11)$$

$$\text{where, } \varepsilon = fT_s - \frac{k_{\max}}{N} \quad (12)$$

10 and,

$$\hat{\varepsilon} = \hat{f}T_s - \frac{k_{\max}}{N}$$

For the initial frequency estimate using the FFT, $\hat{f}_0 T_s = \frac{k_{\max}}{N}$ and $\hat{\varepsilon} = 0$.

In the noiseless case,

$$D(\varepsilon, \hat{\varepsilon}) = \begin{cases} -1, & \varepsilon - \hat{\varepsilon} = \frac{-1}{2N} \\ 0, & \varepsilon - \hat{\varepsilon} = 0, \\ 1, & \varepsilon - \hat{\varepsilon} = \frac{1}{2N} \end{cases} \quad (13)$$

15 $D(\varepsilon, \hat{\varepsilon})$ is a monotonically increasing function of $\varepsilon - \hat{\varepsilon}$. Therefore, each $D(\varepsilon, \hat{\varepsilon})$, there is a unique inverse mapping to $\varepsilon - \hat{\varepsilon}$. Clearly, $D(\varepsilon, \hat{\varepsilon})$ may be used as a discriminant for fine frequency interpolation between FFT bin center frequencies.

There exists some functional relationship such that,

$$\hat{f}_1 T_s = \frac{k_{\max}}{N} + \psi[D(\varepsilon, \hat{\varepsilon})] \quad (14)$$

20 where, $\psi(\cdot)$ is a monotone increasing function. $\psi(\cdot)$ is called the frequency interpolation function and \hat{f}_1 is the first interpolated frequency estimate.

The requirement that \hat{f}_1 has zero error in the noiseless case is,

$$\psi[D(\varepsilon, \hat{\varepsilon})] = \varepsilon - \hat{\varepsilon}, \text{ for } -1 \leq D \leq 1. \text{ Therefore, } \psi^{-1}(\varepsilon - \hat{\varepsilon}) = D(\varepsilon, \hat{\varepsilon}).$$

25

THE FREQUENCY INTERPOLATION FUNCTION

Da diese Seite viele Empfehlungen ignoriert, ist das OCR-Ergebnis unbrauchbar:

- eingebettete mathematischen Formeln sind nicht in vom Fließtext getrennt (Absatz 23);
- unübliche Zeichen liegen im Text vor (Absatz 40);
- kursiver Zeichensatz kombiniert mit griechischen Buchstaben (Absatz 39).

Es wird empfohlen, diese Seite mit zusätzlichen Abständen zwischen den eingebetteten Formeln und dem Fließtext zu formatieren. Griechische Buchstaben in Formeln und Absätzen sollten nicht kursiv sein. Zirkumflex (^) ist möglichst zur Kennzeichnung von Variablen zu vermeiden: stattdessen können Hochstellungen verwendet werden: „Epsilon Zirkumflex“ könnte durch ε^\wedge oder $\varepsilon^{\text{zirkumflex}}$ dargestellt werden.



Anhang 2, Seite 21

Beispiel 19: eine Seite mit kursiven und unterstrichenen Zeichen

WO 2006/038001

PCT/GB2005/003827

- 132 -

2-(3-*[5-Chloro-4-(1*H*-indol-3-yl)pyrimidin-2-yl]amino* piperidin-1-yl)-*N*-methylacetamide (S Enantiomer)

LCMS 399/401 [M+H]⁺, RT 1.88 min.**EXAMPLE 320**

5 3-*[5-Chloro-4-(1*H*-indol-3-yl)pyrimidin-2-yl]amino*}-*N*-isopropylpiperidine-1-carboxamide (Enantiomer 1)

LCMS 413/415 [M+H]⁺, RT 3.20 min.**EXAMPLE 321**

10 3-*[5-Chloro-4-(1*H*-indol-3-yl)pyrimidin-2-yl]amino*}-*N*-isopropylpiperidine-1-carboxamide (Enantiomer 2)

LCMS 413/415 [M+H]⁺, RT 3.19 min.**EXAMPLE 322**

2-*[3-*[4-*[5-Chloro-4-(1*H*-indol-3-yl)pyrimidin-2-yl]amino* piperidin-1-yl]carbonyl]pyrrolidin-1-yl]-*N*-methylacetamide (Racemate)**

15 LCMS (pH 5.8) 496/498 [M+H]⁺, RT 2.79 min.

In dieser PCT-Veröffentlichung wird ein häufiges OCR-Problem aufgezeigt:

- Absatz 41: Der Text soll nicht unterstrichen sein. Von Unterstreichungen wird speziell bei chemischen Formeln abgeraten (in diesen Fällen sind Wörterbücher nutzlos). Vor allem entstehen Probleme mit allen Zeichen, die von der Unterstreichung geschnitten werden:]) y p ... werden nicht richtig erkannt.
- Absatz 39: kursive Schriftart wird nicht empfohlen. Es wird besonders nahegelegt, innerhalb von Worten die Schriftart nicht zu wechseln (OCR-Maschinen nehmen an, dass alle Zeichen in einem Wort die gleiche Formatierung haben). Als Ergebnis werden alle „1*H*“ und „-*N*-“ falsch erkannt.



Anhang 2, Seite 22

Beispiel 20: eine vollkommen unlesbare Seite

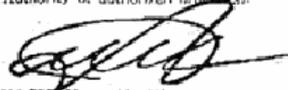
WO 2005/071074

PCT/KR2005/000214

BUDAPEST TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT
OF MICROORGANISMS FOR THE PURPOSE OF PATENT PROCEDURE

INTERNATIONAL FORM -
RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT
issued pursuant to Rule 7.1

TO : **Magen Biotechnology Research Institute
6311, Hwang-m, Namsang-eup, Yongin-city, Kyonggi-do 449-910,
Republic of Korea**

I. IDENTIFICATION OF THE MICROORGANISM	
Identification reference given by the DEPOSITOR: Saccharomyces cerevisiae HJ35017/M&KB #36	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: KCTC 105828P
II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION	
The microorganism identified under I above was accompanied by: <input type="checkbox"/> a scientific description <input type="checkbox"/> a proposed taxonomic designation (Mark with a cross where applicable)	
III. RECEIPT AND ACCEPTANCE	
This International Depositary Authority accepts the microorganism identified under I above, which was received by it on January 13 2004 .	
IV. RECEIPT OF REQUEST FOR CONVERSION	
The microorganism identified under I above was received by this International Depositary Authority on _____ and a request to convert the original deposit to a deposit under the Budapest Treaty was received by it on _____	
V. INTERNATIONAL DEPOSITARY AUTHORITY	
Name: Korean Collection for Type Cultures Address: Korea Research Institute of Bioscience and Biotechnology (KRISS) #52, Oun-dong, Yusong-ku, Taejeon 305-380, Republic of Korea	Signature(s) of person(s) having the power to represent the International Depositary Authority of authorized official(s):  PARK, Yong-Hu Director Date: January 17 2004

Form 1054 (KCTC Form 17)

Diese Seite sollte von den Ämtern nicht angenommen werden. Sie wurde per FAX mit 100 dpi gesendet und kann nicht einmal mit bloßem Auge gelesen werden. Bei der Bearbeitung dieses Falles wird der gesamte Inhalt als Bild deklariert, da kein Text ausgelesen werden kann.

[Ende des Anhangs 2 und des Standards]