

ŠTANDARD ST.22

ODPORÚČANIE STANOVUJÚCE ZÁSADY ÚPRAVY PATENTOVÝCH PRIHLÁŠOK NA ÚČELY UMOŽNENIA OPTICKÉHO ROZPOZNÁVANIA ZNAKOV (OCR)

*Revízia schválená na 10.zasadnutí Pracovnej skupiny pre štandardy a dokumentáciu SCIT
21.novembra 2008*

ÚVOD

1. Odporúčanie sa vzťahuje na patentové prihlášky podané na papieri alebo elektronicky (e-filed), ktorých hlavný text je podaný v obrazovom formáte (napr. PDF alebo TIFF).
2. Odporúčanie bolo vytvorené ako návod na prípravu patentovej prihlášky v strojopisnej forme vhodnej na následné spracovanie digitalizovaného záznamu obsahu prihlášky použitím zariadenia na optické rozpoznávanie znakov (OCR).
3. Odporúčanie bolo vytvorené na základe skúseností rôznych úradov s používaním zariadenia na OCR. Jeho cieľom je dosiahnutie čo najnižšej chybovosti pri automatickom „čítaní“ textu patentových prihlášok a zároveň umožnenie efektívneho čítania dokumentu používateľom. Tento dokument neposkytuje podrobné odporúčania pre japonský a kórejský jazyk; percentá počtu plnotextových elektronických podaní z celkového počtu podaní za rok je v týchto krajinách vyšší ako 90% čo znamená, že tento štandard nie je pre japonský a kórejský jazyk v praxi využiteľný.
4. Hlavným cieľom spracovania digitalizovaného záznamu patentovej prihlášky je umožnenie jednoduchej publikácie prihlášky v zloženom formáte použitím počítačových techník sádzania textu a tým zlepšenie prezentácie patentového dokumentu a zvýšenie jeho hodnoty v prospech všetkých používateľov. Ďalším cieľom je vytvorenie strojom čitateľnej databázy úplných textov vydaných dokumentov, pričom neskôr možno využívať výhody, ktoré plnotextové vyhľadávanie ponúka.

DEFINÍCIE

5. Na účely tohto odporúčania výraz „patentová prihláška“ znamená prihlášku pre patenty na vynálezy, autorské osvedčenia na vynálezy, osvedčenia o užitočnosti, úžitkové vzory, dodatkové patenty alebo dodatkové osvedčenia, dodatkové autorské osvedčenia a dodatkové osvedčenia o užitočnosti.
6. Matematický alebo chemický vzorec sa nazýva „zloženým“, ak nemôže byť zobrazený ako lineárna postupnosť znakov, pričom každý znak má ľubovoľný dolný alebo horný index. Vzorec sa považuje za obzvlášť zložený ak obsahuje vložené dolné/ horné indexy alebo obsahuje matematické symboly (operátory): integrál, suma alebo súčin.
7. Ohraničujúci box (z angl. bounding box) so znakmi/ súborom znakov je najmenším obdĺžnikom zarovnaným podľa osi, ktorý zahŕňa všetky časti znakov/ súborov znakov.
8. Výraz „kurzíva“ sa vzťahuje na štylizovanú formu rukopisu, pričom písmená sú spojené do slov tak, aby slovo tvorilo jednoduchý súvislý ťah. Typy písma sa nazývajú kurzíva, ak sa podobajú na rukopis.

TVORBA ORIGINÁLU

9. Patentová prihláška sa často pripravuje použitím zariadenia na spracovanie textu. Skúsenosti ukazujú, že najvhodnejším používaným formátom umožňujúcim spoľahlivé využitie zariadenia na OCR je tzv. OCR-B formát definovaný Medzinárodnou organizáciou pre normalizáciu (ISO) v rámci štandardu 1073/II.

VLASTNOSTI PAPIERA PRI PODANÍ V PAPIEROVEJ FORME

10. Prihláška písaná v papierovej forme má mať na účely skenovania nasledovné vlastnosti:

- (a) Papier má byť pevný, biely a čistý.
- (b) Papier má mať hmotnosť medzi 70, prenostne 80 a 120 g/m².
- (c) Veľkosť papiera má byť prednostne A4, vid'. 210 mm x 297 mm alebo 8 1/2 až 11 palcov (čo je v skutočnosti severoamerický štandard).
- (d) Listy majú byť bez rýh, dier a nemajú byť stočené.
- (e) Papier nemá byť savý, aby sa zabránilo rozmazaniu atramentu (napríklad pri použití tlačiarne s atramentovou dýzou).

ODPORÚČANIA TÝKAJÚCE SA ROZLOŽENIA TEXTU NA STRANE

11. Nemala by sa používať obojstranná tlač. Ak to nie je možné má sa použiť dostatočne nepriehľadný papier, aby sa zabezpečilo čisté obojstranné skenovanie.

12. Znaký majú byť úplne čierne na bielom pozadí.

13. Horný, dolný okraj a bočné okraje každého listu majú mať minimálne 2 centimetre, pričom minimálny okraj na ľavej strane každého listu má mať 2,5 centimetra. Referenčné údaje o prihlasovateľovi alebo zástupcovi sa majú uvádzať v hornom okraji listu. Pozrite prílohu 1.

14. Neodporúča sa číslovať riadky. Ak je to potrebné, riadky sa majú číslovať arabskými číslicami, v ľavom okraji aspoň 1 cm mimo rámčeka uvedeného v prílohe 1. Veľkosť čísla riadku má byť aspoň 12 bodov.

15. Odporúča sa číslovanie strán arabskými číslicami bez ďalších znakov a čísla strán uvádzať v strede horného alebo spodného okraja listu, ako je uvedené v prílohe 1.

16. Opis, patentové nároky a anotácia majú byť uvedené vždy na novej strane. Okrem toho, prvé slovo na prvej strane každej z troch uvedených častí prihlášky má identifikovať zodpovedajúcu časť (v jazyku prihlášky); patentové nároky sa číslojú priebežne. Formát čísiel patentových nárokov má umožniť pri každom nároku oddelenie čísla nároku od jeho textu. Odporúčaným formátom sú buď arabské číslice s bodkou alebo slovo „nárok“ (alebo jeho ekvivalent v jazyku prihlášky), za ktorým nasleduje arabská číslica nároku, pričom nasledujúci text nároku je v oboch prípadoch odsadený vpravo od čísla nároku vo vzdialenosti aspoň 1 cm.

17. Jednotlivé strany (obsahujúce odseky s textom alebo vloženým obrázkom) majú byť usporiadané do jedného stĺpca.

18. Strany obsahujúce odseky majú byť orientované na výšku.

19. Nie je vhodné používať orientáciu strany na šírku. Táto je prípustná len v prípade strán obsahujúcich vložené obrázky a tabuľky, ktoré by sa nezmestili na výšku.

20. Každá strana má obsahovať text orientovaný len v jednom smere.

21. Strany orientované na šírku majú byť otočené o 90 stupňov proti smeru hodinových ručičiek, aby ich bolo možné spojiť do súboru strán orientovaných na výšku.

22. Nemajú sa používať poznámky pod čiarou, texty popri okrajoch a záhlaviach, okrem uvedeného v odseku 14 (číslovanie riadkov), odseku 15 (číslovanie strán) a na účely uvádzania spisovej značky v ľavom hornom okraji strany.

ODPORÚČANIA TÝKAJÚCE SA ROZLOŽENIA ODSEKOV

23. Odporúča sa oddeliť tabuľky, zložené chemické vzorce, zložené matematické vzorce, obrázky a výkresy od odsekov s textom, pričom tieto prvky má obklopovať zhora a zdola čistý okraj široký aspoň 1 cm na celú šírku strany.
24. Obrázky a výkresy sa majú v maximálnej miere zahrnúť do časti patentovej prihlášky s názvom „Výkresy“ s uvedením vzťahových značiek v častiach „Opis“ a „Patentové nároky“.
25. Obrázky a výkresy majú byť čiernobiele (nemajú sa používať obrázky v odtieňoch sivej z dôvodu ich straty pri skenovaní alebo konverzii na čiernu a bielu). Čísla majú mať jasné, dostatočne hrubé kontúry, aby boli čitateľné pri rozlíšení 300 dpi (bodov na palec).
26. Neodporúča sa písať odseky s textom alebo anotácie ručne. Ak je to nevyhnutné, ručne písaný text sa má považovať za vložený obrázok a má sa pri ňom postupovať podľa odporúčaní odseku 23.
27. Pri písaní textu sa má použiť riadkovanie 1,5.
28. Odseky majú byť oddelené riadkovaním, ktoré je najmenej dvakrát väčšie ako riadkovanie vo vnútri odseku.
29. Všetky znaky v rámci odseku okrem dolných a horných indexov majú mať základné línie starostlivo zarovnané ako je uvedené v odseku 35.
30. Odseky v texte sa nemajú zarovnávať. V prípade zarovnávanía sa odporúčajú medzery medzi slovami aspoň v takej vzdialenosti ako v rámci nezarovnaného textu. Zarovnaný text môže pri OCR zabrániť správnej identifikácii hraníc slova v odseku.
31. Odporúča sa podľa možnosti nerozdeľovať slová pomlčkami (napr. na konci riadkov alebo buniek v tabuľke). To sa nevzťahuje na jazyky, ktoré obsahujú zložené podstatné mená (napr. nemčina).

ODPORÚČANIA TÝKAJÚCE SA TABULIEK

32. Používať sa má len biele pozadie.
33. Tabuľky musia byť ohraničené. Hranice tabuľky majú byť hrubšie ako 1,5 bodu a majú mať len plné čiary.

ODPORÚČANIA TÝKAJÚCE SA VEĽKOSTI PÍSMO

34. Minimálne odporúčaná veľkosť písma je 12 bodov, pričom sa uprednostňuje 14 bodová veľkosť písma. Všeobecne sa odporúča rovnaká veľkosť všetkých znakov v odseku.
35. Pri odsekoch s textom obsahujúcim dolné a horné indexy sa má používať veľkosť písma aspoň 12 bodov (odporúča sa 14 bodov, čím väčšie tým lepšie). Uistite sa, že ohraničujúci box so symbolmi dolných a horných indexov dostatočne pretína box s bežnými znakmi na tom istom riadku (Týmto sa zabráni pri OCR uloženiu dolných/ horných indexov na rôznych riadkoch.)
36. Odporúčajú sa nasledovné rodiny písma v tomto poradí:
 - (a) monospace (neproporcionálne), s rovnakou šírkou znakov (písací stroj): OCR-B, Courier New, Free Mono
 - (b) serif (pätkové): ITC Officina Serif, Times New Roman, Free Times
 - (c) sans-serif (bezpätkové): Verdana, ITC Officina Sans, Arial, Helvetica, DejaVu Sans.

Avšak typy (fonty) písma Arial, Helvetica, DejaVu Sans, Free Times a Times New Roman sa neodporúčajú použiť v prihláškach obsahujúcich chemické a/alebo matematické vzorce ako aj skratky obsahujúce zmes písmen a číslíc. Pre čínske znaky sa odporúča použiť font Song.

37. Písmo má byť zreteľne tvarované, bez tieňov. Medzery medzi znakmi majú byť dostatočne široké (je vhodné vyvarovať sa úzkych medzier).

38. Úzke fonty a kurzíva sa nemajú používať.

39. Znakovým štýlom - tučné (bold) a kurzíva sa treba vyhýbať čo možno najviac.

40. Neobvyklým (neštandardným / netypickým / nepravidelným) znakom sa treba vyhýbať čo možno najviac. Ak je to nutné, majú to byť znaky štandardnej gréckej abecedy s použitím symbolového fonu (v prioritnom poradí). Nesmú sa používať znaky, ktoré nepatria do škály znakov UNICODE: tieto znaky sa považujú za vložené obrázky pomocou zariadenia na OCR a preto robia rozpoznaný text ťažko čitateľným. Každý úrad stanoví a zverejní svoje požiadavky na súbor znakov, ktoré možno použiť na prípravu patentovej prihlášky.

41. Neodporúča sa podčiarkovať text. Ak je to nevyhnutné má sa zabezpečiť, aby podčiarknutie nepretínalo ohraničujúci box podčiarknutých znakov.

ODPORÚČANIA TÝKAJÚCE SA KOMBINÁCIE JAZYKOV

42. Nakoľko pri zmiešaní ázijských (t.j.založených na ideogramoch) a európskych (t.j.latinka a azbuka) jazykov v rámci častí/strán patentovej prihlášky dochádza k problémom pri OCR, odporúča sa vyhýbať sa ich kombinovaniu, okrem prípadov, kde je to nutné.

43. Patentové prihlášky sa majú skenovať buď čiernobielo alebo v odtieňoch sivej.

44. Pri skenovaní sa uprednostňuje rozlíšenie 300 dpi. Výsledkom skenovania pri rozlíšení menej ako 300 dpi a to aj v odtieňoch sivej môžu byť nekvalitné dokumenty vydávané úradmi, pretože výmena dokumentov medzi úradmi a proces ich vydávania si často vyžaduje konverziu strán do formátu TIFF (skupina IV) pri rozlíšení 300 dpi v čiernobieli.

45. Naskenované dokumenty sa majú konvertovať buď do formátu PDF alebo TIFF.

OPRAVY

46. Opravy textu prihlášky sa majú vykonať novým vydaním celej strany. Nie sú prípustné symboly na korekciu textu - ako napríklad symboly určené medzinárodným štandardom ISO 5776. Nie sú prípustné ani opravy prostredníctvom korekčného bielidla, papierovej samolepiacej pásky, vymazanie alebo prečiarknutie. Opravené strany sa nemajú zasielať použitím štandardného faxového rozlíšenia: strany sa majú zasielať fyzicky alebo faxom použitím rozlíšenia vyššieho alebo rovného 300 dpi alebo prostredníctvom akéhokoľvek prostriedku sieťového prenosu podporovaného úradom za podmienky, že každá strana bola naskenovaná pri rozlíšení vyššom alebo rovnom 300 dpi.

ODPORÚČANIA PRE ÚRADY

47. Patentové úrady sa majú vyhýbať zámene prijatých strán prihlášky pred ich postúpením na skenovanie a OCR. Napríklad niektoré súčasné praktiky zahŕňajú činnosti súvisiace s vybavením prihlášky podacím číslom a dátumom podania (napr.perforovaním, pečiatkou), ktoré môžu spôsobiť prekrytie znakov na strane a tým urobiť text prihlášky predloženej prihlasovateľom nečitateľným pomocou OCR. V prípade použitia pečiatok na originálnych stranách dokumentu, úradom sa odporúča prijať opatrenia na zabezpečenie toho, aby sa pečiatky vyskytovali len na okrajoch dokumentu ako je uvedené v prílohe 1.

48. Patentové úrady sa v budúcnosti majú vyhýbať používaniu papierových formulárov na udržiavanie korešpondencie medzi prihlasovateľmi a úradom. Podľa skúseností vytvorenie a zavedenie zabezpečených systémov online formulárov je vhodnejšie na budovanie systémov na rozpoznávanie papierových formulárov. Nasledovné odporúčania sa však týkajú používania papierových formulárov na účely uľahčenia ich rozpoznávania:

- Na usmernenie používateľa, kde má byť vložený text sa neodporúčajú používať vo formulároch riadky s malými bodkami.
- Okienka pre znak majú byť vyšrafované sivou farbou.
- V oblastiach, kde sa predpokladá výskyt odtieňov sivej alebo iného farebného textu ako naskenovaný podpis alebo výkres sa nemá šrafovanie používať.

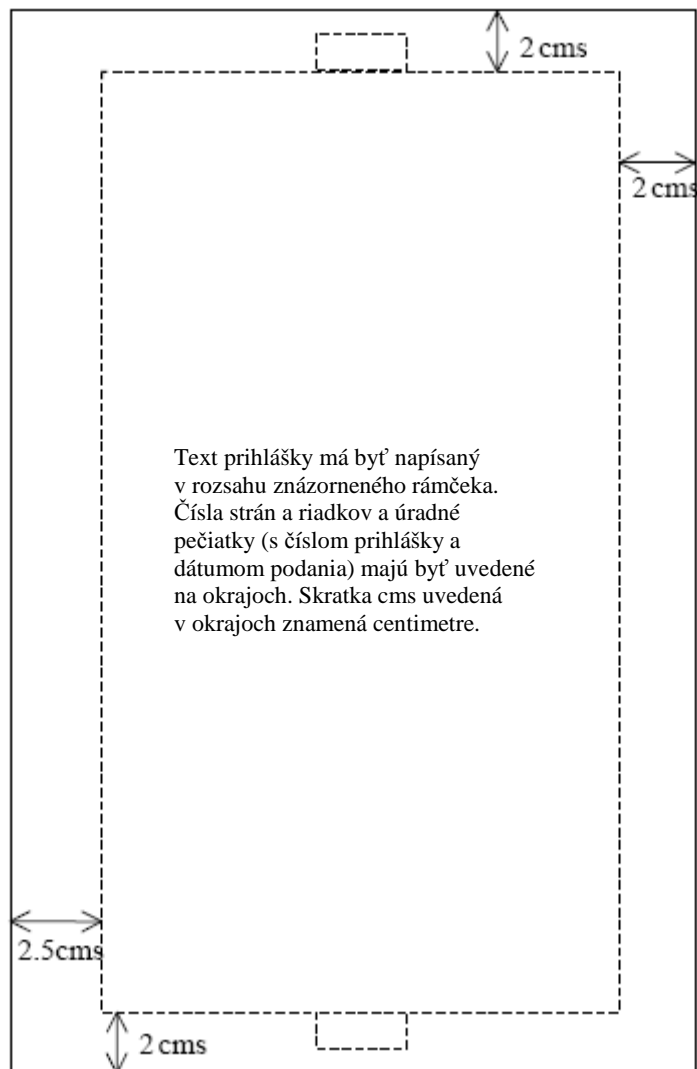
UVEDENIE DO PRAXE

49. Úradom, ktoré majú v úmysle začať prijímať alebo požadovať podanie patentovej prihlášky napísanej vo formáte OCR sa odporúča pravidelne zverejňovať kompletnú príručku v úradnom vestníku a na svojich web stránkach s presným vymedzením dovoleného typu písma a prípustnej veľkosti papiera.

Príklady

50. Príklady správnych a nesprávnych praktík súvisiacich s OCR sú uvedené v prílohe 2 tohto odporúčania. Príklady ukazujú, správne a nesprávne praktiky spolu s krátkym vysvetlením.

[Nasledujú prílohy]

PRÍLOHA 1

Pôvodný rozmer = A4

PRÍLOHA 2**PRÍKLADY SPRÁVNÝCH A NESPRÁVNÝCH PRAKTÍK**

Príloha obsahuje správne a nesprávne príklady strán patentových dokumentov vzhľadom na presnosť dosiahnutú pri ich spracovaní do OCR formátu.

*Príklady správnych praktík*Príklad 1: správne napísaná strana s opisom vynálezu

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.

For example, EP 1 306 665 A2 discloses an optical instrument for measuring polarization-dependent properties which comprises a light source in the
25 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.

Príklad 2: správne napísaná strana s patentovými nárokmi

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.

Strana je v súlade so všetkými odporúčaniami štandardu: okraje a štandardný typ písma (Times New Roman), vyhovujúca veľkosť písma, neočíslované riadky, obmedzené použitie znakového štýlu bold, žiadna kurzíva a podčiarknutý text, primerané číslovanie nárokov a oddelenie čísel od textu s nárokmi.

Príklad 3 správne napísaná strana so zloženými vzorcami

WO 2006/102655

PCT/US2006/011076

[0134] When performing the first iteration of step S9-4, the values of D_a , 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:

$$5 \quad \text{if } (\lambda_{x,y,z-\max}^{n+1})_{ang \& lin} < 0 \text{ then } \lambda_{x,y,z-\max}^{n+1} = 0 \quad (46)$$

$$\text{if } (\lambda_{x,y,z-\min}^{n+1})_{ang \& lin} > 0 \text{ then } \lambda_{x,y,z-\min}^{n+1} = 0 \quad (47)$$

$$\lambda_{lin}^{n+1} = \lambda_{lin-\min}^{n+1} + \lambda_{lin-\max}^{n+1} \quad (48)$$

$$\lambda_{ang}^{n+1} = \lambda_{ang-\min}^{n+1} + \lambda_{ang-\max}^{n+1} \quad (49)$$

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

$$10 \quad D_a^{n+1} = D_a^n + L \frac{(\lambda_{lin}^{n+1} - \lambda_{lin}^n)}{m_a} \quad (50)$$

$$A_a^{n+1} = A_a^n + I_a^{-1} [r_a^s] L (\lambda_{lin}^{n+1} - \lambda_{lin}^n) + I_a^{-1} T (\lambda_{ang}^{n+1} - \lambda_{ang}^n) \quad (51)$$

$$D_b^{n+1} = D_b^n - L \frac{(\lambda_{lin}^{n+1} - \lambda_{lin}^n)}{m_b} \quad (52)$$

$$A_b^{n+1} = A_b^n - I_b^{-1} [r_b^s] L (\lambda_{lin}^{n+1} - \lambda_{lin}^n) - I_b^{-1} T (\lambda_{ang}^{n+1} - \lambda_{ang}^n) \quad (53)$$

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:

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

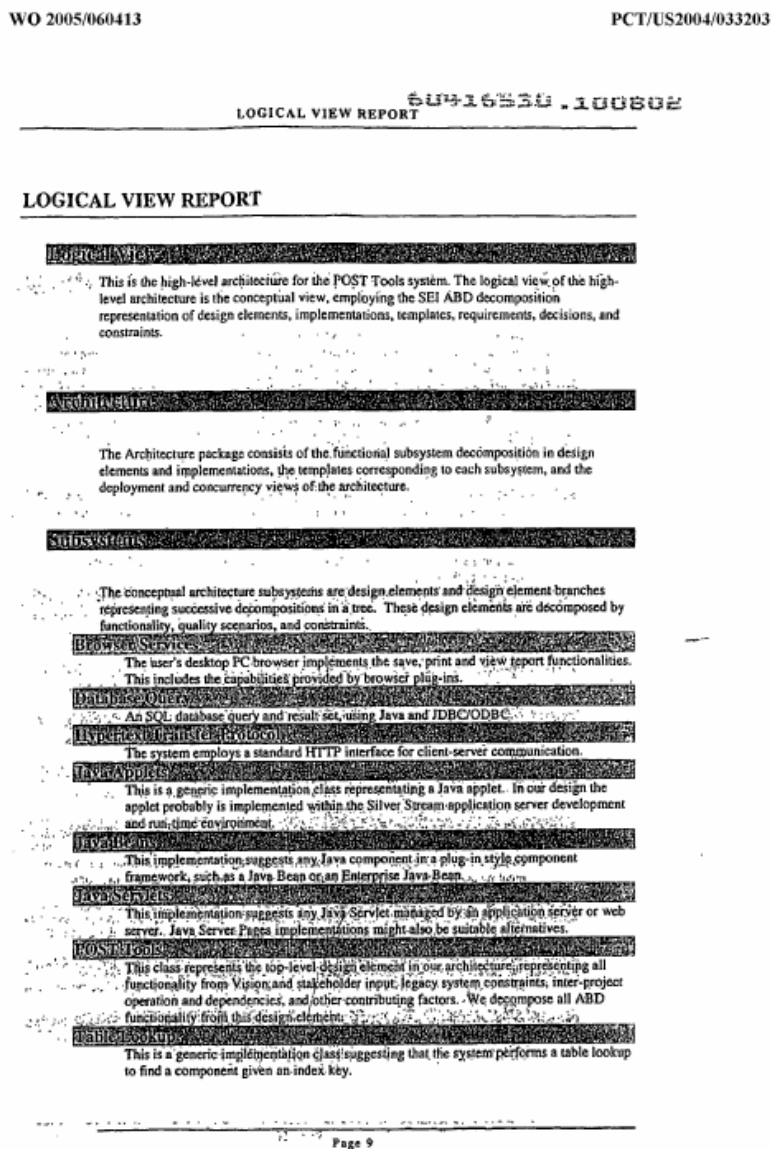
20 [0138] In this embodiment, the threshold employed in Equation (54) is set to 10^{-4} .

[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.

Odseky s textom neobsahujú nezvyčajné matematické symboly. Matematické vzorce sú správne vložené, pričom umožňujú jednoduchú segmentáciu vložených obrázkov pomocou zariadenia OCR. Možný výsledok tejto segmentácie je vyznačený modrou farbou.

Príklady nesprávnych praktík

Príklad 1: nekvalitná strana s viacerými nedostatkami



Tento príklad nie je v súlade s odsekom 10 (strana bola pravdepodobne podaná faxom pri rozlíšení 200 dpi – pozrite „šum“ – niektoré časti textu majú tmavosivé pozadie) ani s odsekmi 13 a 47: číslo spisu (60411530.100802) je vyznačené v hlavnom texte strany (odporúča sa na okrajoch). Číslovanie je nesprávne (odporúča sa „9“ nie „strana 9“, pozrite odsek 34) a veľkosť písma je prímalá (odsek 15). V ideálnom prípade by úrady nemali prijímať takéto strany a mali by požadovať ich náhradu (túto stranu nemožno správne spracovať do formátu OCR).

Príklad 2: strana s iným ako bielym pozadím

WO 2005/097403

- 13 -

PCT/FR2005/050194

REVENDEICATIONS

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), CARACTÉRISÉ 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 fermée 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, CARACTÉRISÉ 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.


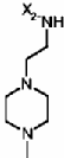
3. Dispositif (D) selon la revendication 2, CARACTÉRISÉ PAR LE FAIT QUE chaque plateau (220 et 230) est équipé de deux appuis (410, 420 et 510, 520) pour accueillir et maintenir en position l'extrémité de la pièce (300).

4. Dispositif (D) selon la revendication 1, CARACTÉRISÉ 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

Tento príklad nie je v súlade s odsekom 12. Strana vyžaduje pokusiť sa o odstránenie nečistého pozadia pred postúpením na OCR, inak bude text po vykonaní OCR nečitateľný.

Príklad 3: strana s nejasnými znakmi

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

Beispiele 26-40

Die folgenden Verbindungen sind über ein analoges Verfahren
beschrieben, hergestellt. Die Herstellung des Benzol-
5 beschrieben. Das für die Darstellung des Amids ei

Na účely zobrazenia znakov je priblížená malá časť strany: farba pôvodného textu je pravdepodobne sivá, ako výsledok skenovania pri rozlíšení 300 dpi v čiernobielym pričom znaky nie sú úplne tmavé. Výsledkom je nepresnosť textu spracovaného do formátu OCR (tento príklad nie je v súlade s odsekom 12).

Príklad 4: strana s ručne písaným textom

TITLED : JIG HEAD SWAY BAR

BACKGROUND
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

Podľa očakávania text získaný spracovaním tejto strany do formátu OCR je nečitateľný. Úradom sa odporúča požadovať strojom písaný text na zabezpečenie minima kvality vydaného dokumentu.

Príklad 5: strana neusporiadaná podľa odporúčania a iné nedostatky

WO 2005/086760

PCT/US2005/007335

38

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.

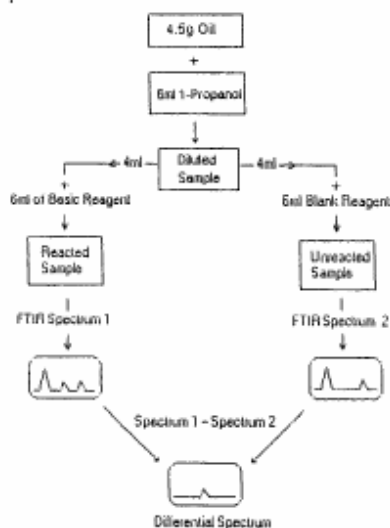


Figure 2. Analytical protocol for the determination of AN by FTIR spectroscopy.

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.

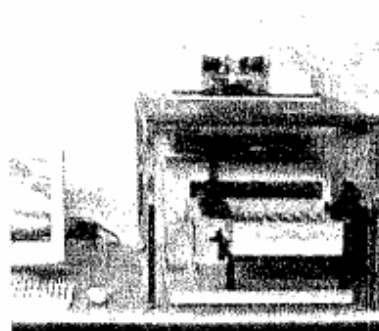


Figure 3. The COAT AN/BN Analyzer and its key components

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

3

Táto strana nie je v súlade s nasledujúcimi odporúčaniami: odsek 17 (formátovanie textu do dvoch stĺpcov), odsek 39 (použitie písma v kurzíve a bold), odsek 46 (opravy po vytlačení strany vykonané ručne). Neodporúča sa tiež zarovnanie textu vľavo a vpravo (odsek 30), hoci v tomto prípade by to nemalo negatívny vplyv nakoľko slová sú dostatočne oddelené bielymi medzermi. Napokon príklad nie je v súlade ani s odsekom 27 (riadkovanie 1,5).

Príklad 6: strana s primalými číslami riadkov

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 μM (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 μM 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 μM (open diamonds), 500 μM (closed circles) and 1000 μM (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 μM 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 μM (open diamonds), 500 μM (closed circles) and 1000 μM (open squares);
- 30 [0032] Fig. 12 is a plot showing the percent increase in cytotoxicity (as determined by $(\text{IC}_{50_{\text{no cysteine}}}/\text{IC}_{50_{\text{cysteine}}}) \times 100$) 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

6

Čísla riadkov spôsobujú problémy zariadeniam OCR z niekoľkých príčin (odsek 14):

- nie sú zarovnané s riadkami, ku ktorým patria, čo vedie k základným nedostatkom;
- možno sú primalé, čoho následkom je zabránenie korektnému odstráneniu chýb z textu strany v procese XML extrakcie dát
- mohli by sa stratiť v rámci textu strany alebo na okrajoch s výnimkou, ak sú príliš blízko pri texte, a výsledkom je, že sa objavajú v texte exportovanom systémom OCR.

V tomto príklade sú čísla riadkov primalé.

Dolné indexy sú v tomto príklade tiež primalé (odsek 35).

Príklad 7: strana obsahujúca viaceré orientácie textu

WO 2005/081642

PCT/JP2005/003688

Table 11 (continued-2)

	Saponification degree	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
Cellulose acetate film 25																
This invention		3	3	14								15	94.6	1.2	7.0	3.2
Cellulose acetate film 26																
This invention		5	5	10								15	94.6	1.2	6.2	0.8
Cellulose acetate film 27																
This invention	59.8		5			15	4.8	10.2					94.6	1.2	6.2	0.8
Cellulose acetate film 28																
This invention		10				10	4.8	10.2					94.6	1.2	6.2	0.8
Cellulose acetate film 29																
This invention						15	4.8	10.2					94.6	1.2	6.2	0.8
Cellulose acetate film 30																
Comparative example	60.9	10	10										94.6	1.2	6.6	0
Cellulose acetate film 31																
Comparative example	60.5			20									94.6	1.2	4.1	0
Cellulose acetate film 32																
Comparative example	60.9								5	10			94.6	1.2	0	6.3
Cellulose acetate film 33																
Comparative example	60.8	10	10				10.5	4.5					94.6	1.2	7.1	0.8
Cellulose acetate film 34																
Comparative example	60.8	10	10				10.5	4.5					94.6	1.2	7.1	0.8
Cellulose acetate film 35																

Tento príklad nie je v súlade s odsekom 20. Jedným z obmedzení najlepších zariadení na OCR v súčasnosti je, že dokážu čítať text na jednej strane orientovaný len v jednom smere (predbežným spracovaním sa zisťuje hlavná orientácia textu na strane). Výsledkom je, že všetky slová, ktoré nie sú orientované v smere ako hlavný text sa ignorujú. Je samozrejme prijateľné mať na strane tabuľku orientovanú na šírku alebo dokonca hlavný text s poznámkami na okrajoch orientovanými na výšku (číslo strany, číslo prihlášky atď.).

Príklad 8: strana obsahujúca kombináciu vložených matematických vzorcov a textu

WO 2005/116630 PCT/US2005/017216

19

$$\Delta \mathbf{L} = \frac{\hbar}{2} - \mathbf{r} \times e \mathbf{A} \quad (33)$$

$$= \left[\frac{\hbar}{2} - \frac{e \phi}{2\pi} \right] \hat{z} \quad (34)$$

In order that the change of angular momentum (ΔL_z) equals zero, ϕ must be $\Phi_0 = \frac{\hbar}{2e}$, the magnetic flux quantum. The magnetic moment of the electron is parallel or antiparallel to the applied field only. During the spin-flip transition, power must be conserved. Power flow is governed by the Poynting power theorem,

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

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,

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

$$\Delta E_{\text{mag}}^{\text{spin}} = g \mu_B B \quad (37)$$

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 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 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 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 $n = 1, \ell = 0$

Tento príklad nie je v súlade s odsekom 23. Zariadenie na OCR nedokáže korektne oddeliť text a vzorec (pozrite výsledok manuálnej segmentácie vzorca červenou farbou: vložené vzorce sa dokonca pretínajú).

Nie je v súlade ani s odsekmi 27 a 28. Vo všeobecnosti možno poznamenať, že text a vzorce sú príliš nahusto, aby ich bolo možno dobre rozpoznať.

V tomto príklade sú tiež použité nezvyčajné znaky: grécke symboly možno používať aj keď zvyšujú obtiažnosť rozpoznávania strany (pozri odsek 40), hoci sa veľmi odporúča nekombinovať kurzívu, tučné alebo podčiarknuté písmo s nezvyčajnými znakmi (odsek 39).

Príklad 10: strana obsahujúca príliš malé symboly dolných indexov

WO 2005/110416

PCT/US2005/015897

R₁ is hydrogen, C₁-C₆alkyl, C₂-C₆alkenyl, C₂-C₆alkynyl, C₁-C₆alkoxy, C₁-C₆haloalkyl, C₁-C₆haloalkoxy, (C₃-C₇cycloalkyl)C₀-C₄alkyl;

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)_j-T, C₁₋₆alkyl, C₁₋₆alkoxy, C₁₋₆haloalkyl, C₁₋₆haloalkoxy, mono- and di-(C₁₋₆alkyl)amino, (C₁₋₆alkyl)((CR_AR_B)_j-T)amino, benzyl, S(O)_n(C₁₋₆alkyl), α,ω-C₁₋₄alkylene, α,ω-C₁₋₄alkyleneoxy, α,ω-C₁₋₄alkylenedioxy, -E-(CH₂)_n-Q, and Q;

T is CO₂H, CONH₂, C₁₋₆alkoxycarbonyl, mono- or di-(C₁₋₆alkyl)aminocarbonyl, SO₃H, SO₂NH₂ or SO₂(C₁₋₆alkyl);

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;

R₈ and R₉ are independently chosen from hydrogen, halogen, hydroxy, C₁-C₆alkyl, C₁-C₆alkenyl, (C₃-C₆cycloalkyl)C₀-C₄alkyl 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 R₈.

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, C₁₋₆alkyl, C₁₋₆alkenyl, C₁₋₆alkynyl, C₁₋₆haloalkyl, C₁₋₆aminoalkyl, C₁₋₆hydroxyalkyl, C₁₋₆carboxyalkyl, C₁₋₆alkoxy, C₁₋₆haloalkoxy, C₁₋₆alkylthio, C₁₋₆alkanoyl, C₁₋₆alkanoyloxy, C₃₋₆alkanone, C₁₋₆alkyl ether, mono- or di-(C₁₋₆alkyl)aminoC₀₋₆alkyl, -NHC(=O)(C₁₋₆alkyl), -N(C₁₋₆alkyl)C(=O)(C₁₋₆alkyl), -NHS(O)_n(C₁₋₆alkyl), -(C₁₋₆alkyl)C(=O)NH₂, -(C₁₋₆alkyl)C(=O)NH(C₁₋₆alkyl), -(C₁₋₆alkyl)C(=O)NH(C₁₋₆alkyl)(C₁₋₆alkyl), -S(O)_n(C₁₋₆alkyl), -S(O)_nNH(C₁₋₆alkyl), -S(O)_nN(C₁₋₆alkyl)(C₁₋₆alkyl) and Z; or

Je to typický príklad, kde sú symboly dolných indexov príliš malé na to, aby umožnili ich presné rozpoznanie. S týmto javom sa často stretávame v oblasti chemických patentov.

Príklad 11: strana so zle naformátovanými tabuľkami

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.

5	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 NHCH(CH ₃)CH ₃
						CH ₃
						CHCH ₃ 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.

10	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 NHCH(CH ₃)CH ₃
						CH ₃

V tomto príklade sa stratili hranice tabuľky (nie je v súlade s odsekom 33). Výsledkom je, že zariadenie OCR sa pokúsilo rozpoznať obsah tabuľky ako odsek s textom, čo vedie k viacerým problémom:

- Veľkosť znakov v tabuľke je príliš malá (odseky 34 a 35)
- Základné línie záhlaví stĺpcov sú zmiešané (odsek 29). Výsledkom je, že zariadenie odhalilo chybné horné a dolné indexy
- Získaný text neberie do úvahy stĺpce:

Left-hand Substituent **Left-Hand ring** **Aryl or heteroaryl** **Nitrogen feature** **Right-hand substituent**
 CH3...

Príklad 12: zarovnaná strana

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

V tomto príklade sa použilo zarovnanie odsekov vľavo a vpravo. Za účelom dosiahnutia estetickejšie vyzerajúceho textu, niekedy môžu príliš malé medzery medzi slovami spôsobovať ťažkosti pri vykonávaní OCR (odsek 30). Tento príklad nie je v súlade ani s odsekom 31 odporúčajúcim podľa možnosti nerozdeľovať slovo na konci riadku (zariadenie na OCR má niekedy ťažkosti rozlišovať tvrdú a mäkkú pomlčku, dôsledom čoho sú slová obsahujúce nežiadúce pomlčky vo výstupe).

Príklad 13: tabuľka s chybnými hranicami

WO 2004/110415

- 60 -

PCT/EP2004/051048

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

V tomto príklade, pred skenovaním neboli k dispozícii kvalitné hranice tabuľky v pôvodnom úradom prijatom dokumente. Výsledkom skenovania je, že systém OCR nedokázal správne rozpoznať tabuľku a vyžaduje sa manuálny zásah na segmentáciu strany. Ak takáto strana nebude skontrolovaná operátorom kvality výstup bude obsahovať v texte skomoleniny znakov spôsobujúce, že indexácia dokumentu vyhľadávacími nástrojmi bude menej efektívna.

Príklad 14: chybné znaky dolných a horných indexov

WO 2005/100305

PCT/IB2005/000872

-9-

thiazolyl, pyrazolyl, pyridinyl, pyrimidinyl, purinyl, quinoliny,
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-C₆)alkyl, (C₁-C₆)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, tetrahydrofuranyl, tetrahydrothiophenyl, piperidinyl, piperazinyl, azepane, azocane, morpholinyl, isochroamyl 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-C₆)alkyl, (C₁-C₆)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

V tomto príklade existujú nasledujúce problémy (odsek 35)

- príliš malé znaky dolných a horných indexov
- znaky dolných indexov sú umiestnené príliš nízko vzhľadom na základnú líniu riadku
- znaky horných indexov sú umiestnené príliš vysoko vzhľadom na základnú líniu riadku

Výsledkom je, že riadky 33 a 34 v texte sú systémom OCR rozpoznané nasledovne:

"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"

Príklad 15: príklad s nezvyčajnými znakmi**WO 2006/057705****PCT/**

c = speed of sound in water;

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

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

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

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

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

10 $\epsilon_{\gamma,p,u}$ = difference of horizontal-range sample v linearization point in ping p , beam pair u , from the nominal $\gamma_{v,u}$. Note that this is the same for all horizontal-range samples;

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

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

V tomto príklade sa vyskytujú nasledujúce problémy:

- Nezvyčajné znaky: grécke znaky v kurzíve a dokonca znaky s vlnovkou.
- Dolné indexy sú znovu primálne.

Väčšina zariadení OCR nerozpozná tieto neobvyklé znaky korektne.

Príklad 16: príklad s úzkymi typmi písma a úzkymi medzerami

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.

Tento príklad nie je v súlade s odsekmi 37 a 38. Výsledkom je, že zariadenie na OCR nedokáže správne rozlíšiť hranice slova a výstup po vykonaní OCR je nepoužiteľný.

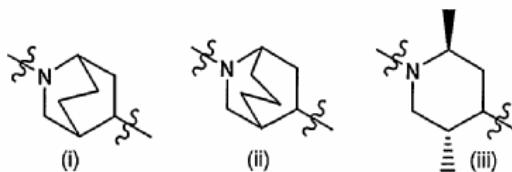
Príklad 17: nevhodne umiestnená pečiatka prijímacieho úradu pred skenovaním

WO 2006/058294

PCT/US2005/042931

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:



Tento príklad nie je v súlade s odsekom 47. Výsledkom je, že technológia OCR nedokáže rozpoznať v texte prvých šesť slov. Okrem toho pokiaľ, strana nebola skontrolovaná operátorom kvality, pečiatka obsahuje ďalšie chybné znaky, ktoré znečistia indexovacie zariadenie.

Príklad 18: ďalšia strana s nesprávne rozloženými matematickými vzorcami

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,

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

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

5

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

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

where, $\varepsilon = f\Gamma_s - \frac{k_{max}}{N}$ (12)

10 and,

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

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

In the noiseless case.

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

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,

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

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,

$$\psi[D(\varepsilon, \hat{\varepsilon})] = \varepsilon - \hat{\varepsilon}, \text{ for } -1 \leq D \leq 1 \tag{15}$$

25 Therefore, $\psi^{-1}(\varepsilon - \hat{\varepsilon}) = D(\varepsilon, \hat{\varepsilon})$.

THE FREQUENCY INTERPOLATION FUNCTION

Výsledok OCR je nepoužiteľný, pretože táto strana nie je v súlade s viacerými odporúčaniami:

- vložené matematické vzorce nie sú oddelené od odsekov v texte (odsek 23);
- nezvyčajné znaky v texte (odsek 40);
- kurzíva kombinovaná s gréckou abecedou (odsek 39).

Odporúčaný spôsob rozloženia tejto strany je použitie ďalších medzier na oddelenie vložených vzorcov od odsekov. Grécke písmená sa nemajú uvádzať vo vzorcoch a texte. Znakom „vokáň“ (^) by sa podľa možnosti nemali označovať premenné v odsekoch textu: namiesto toho sa môže použiť horný index: symbol "epsilon vokáň " môže byť uvedený ako ε^{\wedge} alebo $\varepsilon^{circumflex}$.

Príklad 19: strana obsahujúca symboly v kurzíve a podčiarknuté

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**

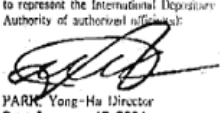
2-{3-*[(4-*[5-Chloro-4-(1*H*-indol-3-yl)pyrimidin-2-yl]amino**]-piperidin-1-yl)carbonylpyrrolidin-1-yl}-*N*-methylacetamide (Racemate)

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

Je to častý problém pri OCR vyskytujúci sa vo vydaných dokumentoch PCT. Táto strana nie je v súlade s nasledujúcimi odporúčaniami:

- odsek 41: text nemá byť podčiarknutý. Podčiarkovanie sa pri chemických vzorcoch neodporúča. Najmä chemické vzorce spôsobujú problémy týkajúce sa všetkých znakov, ktoré sa prekrývajú s podčiarknutím:]) y p ...sú nesprávne rozpoznané
- odsek 39: kurzíva sa neodporúča. Veľmi sa odporúča nemeniť typ písma v rámci slova (zariadenia na OCR často predpokladajú, že všetky znaky slova sú napísané v jednom type písma). Výsledkom je, že symboly "1H" a "-N-" sú nesprávne rozpoznané.

Príklad 20: úplne nečitateľná strana

WO 2005/071074	PCT/KR2005/000214
<small>HEALTHY TREATY ON THE INTERNATIONAL DEPOSITARY OF THE IDENTIFICATION OF MICROORGANISMS FOR THE PURPOSES OF PATENT PROCEDURE</small> INTERNATIONAL FORM - RECEIPT IN THE CASE OF AN ORIGINAL DEPOSIT <small>Issued pursuant to Rule 7.1</small>	
TO: Maxim Biotechnology Research Institute #301, Hujung-ro, Kosaung-eup, Yongin-city, Kyonggi-do 449-910, Republic of Korea	
I. IDENTIFICATION OF THE MICROORGANISM	
Identification reference given by the DEPOSITOR: Saccharomyces cerevisiae 013501//MHLK8 #36	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: KCTC 10582BP
II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION The microorganism identified under I above was accompanied by: <input checked="" type="checkbox"/> a scientific description <input type="checkbox"/> a proposed taxonomic designation <small>(Mark with a cross where applicable)</small>	
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 (KRIBB) #52, Oun-dong, Yusong-ku, Taejeon 305-385, Republic of Korea	Signature(s) of person(s) having the power to represent the International Depositary Authority of authorized official(s):  PARK Yong-Ha Director Date: January 17 2004
<small>FORM DEPOSITARY Form 01</small>	<small>FORM 01</small>

Takúto stranu by úrady nemali prijať: bola odoslaná faxom pri rozlíšení 100 dpi a dokonca nie je čitateľná ľudským okom. Operátor v snahe vyriešiť takýto prípad, považuje celý obsah strany za obrázok, pretože text nie je možné extrahovať.

[Konec prílohy 2 a štandardu]