

СТАНДАРТ ST.22

РЕКОМЕНДАЦІЇ ЩОДО ПРЕДСТАВЛЕННЯ ПАТЕНТНИХ ЗАЯВОК, СПРЯМОВАНІ НА ПОЛЕГШЕННЯ ОПТИЧНОГО РОЗПІЗНАВАННЯ СИМВОЛІВ (OCR)

*Редакція прийнята Робочою групою зі стандартів та документації ПКІТ
на її десятій сесії 21 листопада 2008 року*

ВСТУП

1. Даний рекомендований стандарт стосується патентних заявок, що подаються на папері або в електронному вигляді (електронне подання), текстова частина яких подається у формі зображення (наприклад, зображення у форматі PDF або TIFF).
2. Даний рекомендований стандарт запроваджено для сприяння підготовці патентної заявки у машинописній формі, придатній для подальшого створення електронного цифрового запису вмісту патентної заявки за допомогою обладнання для оптичного розпізнавання символів (OCR).
3. Даний рекомендований стандарт запроваджено на основі досвіду використання обладнання для OCR у різних відомствах. Він був розроблений з метою досягнення найнижчої можливої частоти появи помилок на етапі автоматичного зчитування текстів патентних заявок і, у той же час, надання особі можливості ефективного ознайомлення з документом. Зауважте, що цей документ не надає детальних рекомендацій для японської та корейської мов; відсоток заявок, що подаються в електронному вигляді, від загальної кількості заявок на рік у цих країнах складає більше 90%, а це означає, що цей стандарт на практиці не може застосовуватися для японської та корейської мов.
4. Головна мета створення цифрового запису патентної заявки полягає у полегшенні публікації заявки у скомп'юнованому форматі з використанням техніки комп'ютерного набору даних і, таким чином, у покращенні представлення і збільшення корисності патентних документів на користь усіх користувачів. Ще одна мета полягає у створенні машинозчитуваної бази даних повних текстів опублікованих документів для отримання у подальшому переваг, що надає повнотекстовий комп'ютерний пошуку.

ВИЗНАЧЕННЯ

5. Для цього рекомендованого стандарту вираз «патентна заявка» означає заявки на патенти на винаходи, авторські свідоцтва на винаходи, свідоцтва про корисність, корисні моделі, додаткові патенти або свідоцтва, додаткові авторські свідоцтва на винаходи або додаткові свідоцтва про корисність.
6. Математична або хімічна формула вважається «складною», якщо її не можна відтворити у вигляді лінійної послідовності символів, кожен з яких має необов'язковий нижній або верхній індекс. Формула є винятково складною, якщо вона містить вбудовані нижні/верхні індекси, або якщо вона містить математичні символи додавання, інтеграла або добутку.

7. Обмежуюча рамка символу/набору символів – це найменший прямокутник, вирівняний відповідно до осей координат, який включає усі частини символу/набору символів.

8. Термін «курсив» означає стилізований рукописний формат, коли букви в словах пов'язані так, що відображають слово як одну єдину складну лінію. Шрифти називаються курсивними, якщо вони створені для відтворення рукописної форми написання.

СТВОРЕННЯ ОРИГІНАЛУ

9. Підготовка патентної заявки часто здійснюється з використанням засобів обробки тексту. Досвід показав, що найефективнішим із використовуваних форматів, які дозволяють застосовувати обладнання для OCR, є так званий формат OCR-B, визначений у стандарті Міжнародної організації зі стандартизації (ISO) 1073/II.

ПАПЕРОВИЙ НОСІЙ У ВИПАДКУ ПОДАННЯ ЗАЯВКИ НА ПАПЕРІ

10. Для полегшення сканування паперовий носій надрукованої заявки повинен мати наступні характеристики:

- (a) Папір повинен бути міцним, білим і чистим;
- (b) Вага паперу має становити від 70, краще 80, до 120 г/м²;
- (c) Розмір аркуша має бути А4, а саме 210 мм x 297 мм або 8½ x 11 дюймів (що є де-факто північно-американським стандартом);
- (d) Аркуші не повинні мати загинів, дірок, і не повинні бути скручені;
- (e) Папір не повинен бути поглинаючим, щоб уникнути розпливчатості чорнил (наприклад у разі використання струменевого принтера).

РЕКОМЕНДАЦІЇ ЩОДО РОЗМІТКИ СТОРІНКИ

11. Необхідно уникати двостороннього друку. Якщо це неможливо, необхідно використовувати достатньо непрозорий папір, щоб забезпечити якісне сканування з обох сторін аркуша.

12. Символи мають бути повністю чорними на білому фоні.

13. Зверху, знизу і з правого боку кожного аркуша необхідно залишити мінімальне поле шириною 2 см і мінімальне поле шириною 2,5 см з лівої сторони. Будь-які примітки заявника або його представника має наводитися у верхньому полі (див. додаток 1).

14. Необхідно уникати нумерації рядків. За необхідності, номери рядків повинні бути надруковані арабськими цифрами у лівому полі, на відстані щонайменше 1 см від рамки, як показано у додатку 1. Розмір шрифту номерів рядків має бути щонайменше 12 кеглів.

15. Нумерацію сторінок необхідно вказувати арабськими цифрами без будь-яких інших символів розмежування. Номери сторінок мають бути вирівняні по центру у верхньому або нижньому полі сторінки, як показано у додатку 1.

16. Опис, формула і реферат мають бути надруковані, кожен починаючи з нової сторінки. Крім того, перше слово, що друкується на першій сторінці кожної з вищезгаданих частин

заявки, має містити назву відповідної частини (мовою заявки); пункти формули мають бути пронумеровані послідовно. Формат нумерації пунктів формули має забезпечувати відмежування номера кожного пункту від його тексту. Рекомендованими форматами є арабські цифри, після яких ставиться крапка; слово «Пункт» - або його еквівалент мовою патентної заявки, - після якого ставиться пробіл і його номер арабською цифрою. При цьому в обох випадках подальший текст пункту формули друкується з відступом вправо відносно номера пункту формули щонайменше на 1 см в обох випадках.

17. Абзаци на сторінці (текстові абзаци або абзаци, що містять зображення) розташовуються в одну колонку.

18. Сторінки з абзацами повинні мати книжкову орієнтацію.

19. Необхідно уникати альбомної орієнтації. Вона допускається тільки для сторінок, які містять вбудовані креслення або таблиці, для яких не підходить книжкова орієнтація.

20. У межах однієї сторінки текст повинен розміщуватися тільки в одному напрямку.

21. Сторінки з альбомною орієнтацією повинні бути перевернуті на 90 градусів проти годинникової стрілки для об'єднання їх із комплектом сторінок з книжковою орієнтацією.

22. Рекомендовано уникати використання виносок, тексту і заголовків у полях, окрім випадків, вказаних у параграфі 14 (нумерація рядків), параграфі 15 (нумерація сторінок), і для включення реєстраційного номера заявника у верхньому лівому куті поля.

РЕКОМЕНДАЦІЇ ЩОДО РОЗМІТКИ АБЗАЦІВ

23. Рекомендовано, щоб таблиці, складні хімічні та математичні формули, креслення і зображення відділялися від текстових абзаців. Бажано, щоб такі вбудовані об'єкти мали порожні поля зверху і знизу висотою щонайменше 1 см на ширину сторінки.

24. Креслення і зображення повинні бути максимально включені до розділу «Креслення», а посилання на них повинні міститися у розділах «Опис» і «Формула» патентної заявки.

25. Креслення і зображення повинні бути чорно-білими (необхідно уникати зображень із сірими тонами, так як під час їх сканування або перетворення у чорно-білі ця інформація втрачається). Рисунки повинні бути виконані чіткими лініями, товщина яких має бути достатньою для належного відтворення у разі застосування роздільної здатності 300 точок/дюйм.

26. Необхідно уникати рукописного тексту в абзацах або примітках. Якщо необхідно, вони повинні розглядатися як вбудовані зображення і відповідати рекомендаціям, викладеним у параграфі 23.

27. Текст має друкуватися з інтервалом 1,5.

28. Абзаци повинні відділятися проміжком, щонайменше вдвічі більшим за відстань між рядками в абзаці.

29. Усі символи у межах рядка параграфа повинні бути розташовані чітко на одній прямій, крім нижніх та верхніх індексів, як вказано у параграфі 35.

30. Необхідно уникати вирівнювання тексту абзаців за шириною. Якщо воно все ж таки застосовується, відстань між словами має бути щонайменше такої ж ширини, як і в невирівняному тексті. Текст, вирівняний за шириною, може заважати системам OCR належним чином ідентифікувати межі слова у тексті.

31. По можливості необхідно уникати переносу слів за допомогою дефіса (наприклад наприкінці рядків або комірок таблиці). Це не стосується мов, які використовують складені іменники (наприклад німецької мови).

РЕКОМЕНДАЦІЇ СТОСОВНО ТАБЛИЦЬ

32. Необхідно використовувати тільки білий фон.

33. Таблиці повинні мати границі. Границі мають бути товщими за 1,5 пункти і виконані тільки безперервними лініями.

РЕКОМЕНДАЦІЇ СТОСОВНО ШРИФТІВ

34. Мінімальний рекомендований розмір шрифту – 12 кеглів, бажано 14. Загалом, рекомендується, щоб усі символи тексту мали один і той самий розмір шрифту.

35. У текстових абзацах, що містять нижні і верхні індекси, слід використовувати щонайменше 12-й розмір шрифту (рекомендовано 14 – чим більше, тим краще). Переконайтесь, що обмежуюча рамка нижніх та верхніх індексів належним чином перетинає обмежуючу рамку звичайних символів у тому ж рядку (це запобігає друкуванню нижніх/верхніх індексів у інших рядках під час OCR).

36. Рекомендовано використовувати такі шрифти (у наведеній послідовності):

- (a) Сімейство шрифтів з однаковою шириною символів (Monospaced): OCR-B, Courier New, Free Mono;
- (b) Сімейство шрифтів із зарубками (Serif): ITC Officina Serif, Times New Roman, Free Times;
- (c) Сімейство шрифтів без зарубок (Sans Serif): Verdana, ITC Officina Sans, Arial, Helvetica, DejaVu Sans.

Однак шрифти Arial, Helvetica, DejaVu Sans, Free Times і Times New Roman не рекомендовані для заявок, що містять хімічні та/або математичні формули, а також акроніми, у яких поєднано букви і цифри. Для китайських символів рекомендовано використовувати шрифт Song.

37. Символи шрифтів повинні мати чітку форму і не мати тіней. Проміжки між символами повинні бути достатньо широкими (необхідно уникати вузької розрядки).

38. Не слід використовувати вузькі та рукописні шрифти.

39. По можливості слід уникати напівгрубих і курсивних шрифтів.

40. По можливості слід уникати нестандартних символів. Якщо необхідно, це мають бути стандартні літерами грецького алфавіту і символи шрифту (у порядку їх переваги). Не слід використовувати символи, відсутні у системі UNICODE: такі символи розпізнаються засобами

OCR як вбудовані зображення і через це роблять розпізнаний текст складним для читання. Кожне відомство має визначити і опублікувати свої вимоги до тих символів, які можуть використовуватися для підготовки патентних заявок.

41. Текст не повинен бути підкресленим. Якщо необхідно, слід упевнитись, що підкреслювання не перекриває обмежуючі рамки символів.

РЕКОМЕНДАЦІЇ СТОСОВНО ПОЄДНАННЯ МОВ

42. Поєднання азійських (тобто заснованих на ідеографічному письмі) і європейських (тобто заснованих на латиниці і кирилиці) мов у розділах/на сторінках патентних заявок становить проблему для процедури OCR і його необхідно уникати, окрім випадків, коли воно необхідне.

РЕКОМЕНДАЦІЇ СТОСОВНО СКАНУВАННЯ

43. Патентні заявки необхідно сканувати у чорно-білій або напівтоновій (сірій) шкалі.

44. Найкращою роздільною здатністю для сканування є 300 точок/дюйм. Сканування з роздільною здатністю, меншою за 300 точок/дюйм, навіть у напівтоновій (сірій) шкалі, може призвести до низької якості документів, що публікуються відомством, так як обмін документами між відомствами і процеси публікації часто включають процедуру перетворення на чорно-білі сторінки з роздільною здатністю 300 точок/дюйм у форматі TIFF групи IV.

45. Відскановані документи необхідно перетворювати у формат PDF або TIFF.

ВИПРАВЛЕННЯ

46. виправлення у тексті заявки необхідно здійснювати шляхом передрукування всієї сторінки. Коректурні символи як, наприклад, такі, що вказані у міжнародному стандарті ISO 5776, не допускаються. виправлення за допомогою білої коригувальної рідини, клейких смужок паперу, шляхом витирання або закреслення не допускаються. Аркуші заміни не повинні надсилатися до відомства факсом, що використовує стандартну роздільну здатність факсів: аркуші повинні надсилатися фізично, факсом, що використовує роздільну здатність не меншу за 300 точок/дюйм, або іншими засобами мережевого передавання, що підтримуються відомством, за умови, що кожний аркуш було відскановано з роздільною здатністю не меншою за 300 точок/дюйм.

РЕКОМЕНДАЦІЇ ДЛЯ ВІДОМСТВ

47. Патентні відомства повинні уникати заміни отриманих аркушів до їх передання на сканування і процедуру OCR. Наприклад, сучасний досвід передбачає процедуру проставлення штапу, що може спричинити накладання символів на сторінках, перетворюючи наданий заявником текст на непридатний для розпізнавання під час процедури OCR. Якщо на аркушах оригіналу необхідно проставити штампи/внести зміни, відомству необхідно вжити заходів для того, щоб такі зміни вносилися тільки у межах полів документа відповідно до додатку 1.

48. У подальшому патентні відомства повинні уникати розробки паперових бланків для взаємодії між заявником і відомством. Як показує досвід, розробка і впровадження безпечних онлайн систем є кращим варіантом, ніж побудова системи розпізнавання бланків на папері. Тим не менш, для вже розроблених бланків на папері надаються наступні рекомендації для цілей полегшення їх розпізнавання:

- На бланках не повинні використовуватися рядки, що складаються з маленьких крапок, які вказують користувачеві, куди необхідно вводити текст.
- Для знакомісць необхідно використовувати кольори, що пропускаються під час сканування (світло-сірий).
- Кольори, що пропускаються під час сканування, не повинні використовуватися для полів, у яких формою передбачено заповнення з використанням кольорів або відтінків сірого, як у випадку з від сканованим підписом або зображенням.

ВПРОВАДЖЕННЯ

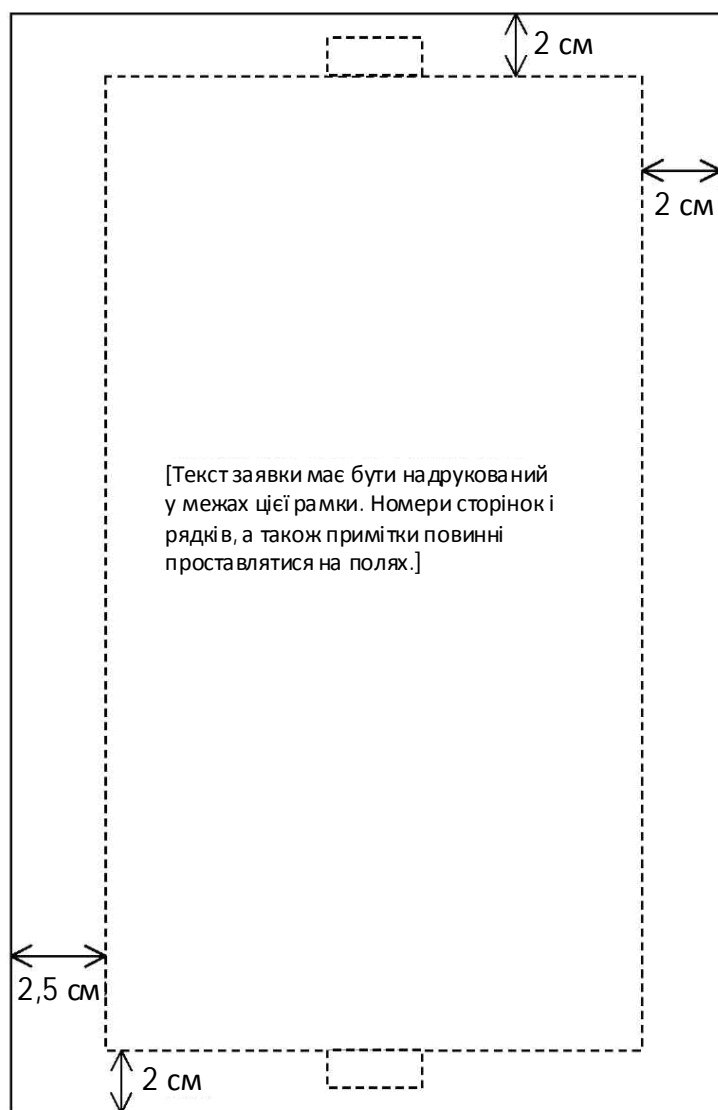
49. Рекомендовано, щоб відомства, які мають намір розпочати прийом або вимагатимуть подання патентних заявок, роздрукованих у форматі, придатному для OCR, регулярно публікували довідники у повному обсязі у своїх офіційних бюлетенях і на веб-сайтах, чітко визначаючи допустимі шрифти і розмір паперу.

Приклади

50. Приклади позитивного і негативного практичного досвіду стосовно OCR у патентних заявках наведено у додатку 2 до цих рекомендацій. Наведені приклади показують, що потрібно, а що не потрібно робити, і супроводжуються поясненнями.

[Додатки наведено далі]

ДОДАТОК 1



Розмір оригіналу = A4

ДОДАТОК 2

ЗРАЗКИ ПАТЕНТНИХ ДОКУМЕНТІВ, ЩО ВІДПОВІДАЮТЬ АБО НЕ ВІДПОВІДАЮТЬ РЕКОМЕНДАЦІЯМ ДАНОГО СТАНДАРТУ

У цьому додатку представлено зразки сторінок патентних документів, що відповідають або не відповідають рекомендаціям даного стандарту в частині точності їх відтворення під час процедури OCR.

Зразки, що відповідають рекомендаціям даного стандарту

Приклад 1: сторінка опису гарної якості

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 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.

25

Додаток 2, сторінка 2

Приклад 2: сторінка з формулою гарної якості

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.

Відповідає усім рекомендаціям: поля, стандартний шрифт (Times New Roman), належний розмір шрифту, без номерів рядків, обмежене використання напівгрубого шрифту, відсутній курсив, відсутній підкреслений текст, пункти формули пронумеровані належним чином і чітко відокремлені від тексту формули.

Додаток 2, сторінка 3

Приклад 3: сторінка зі складним описом гарної якості

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:

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

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

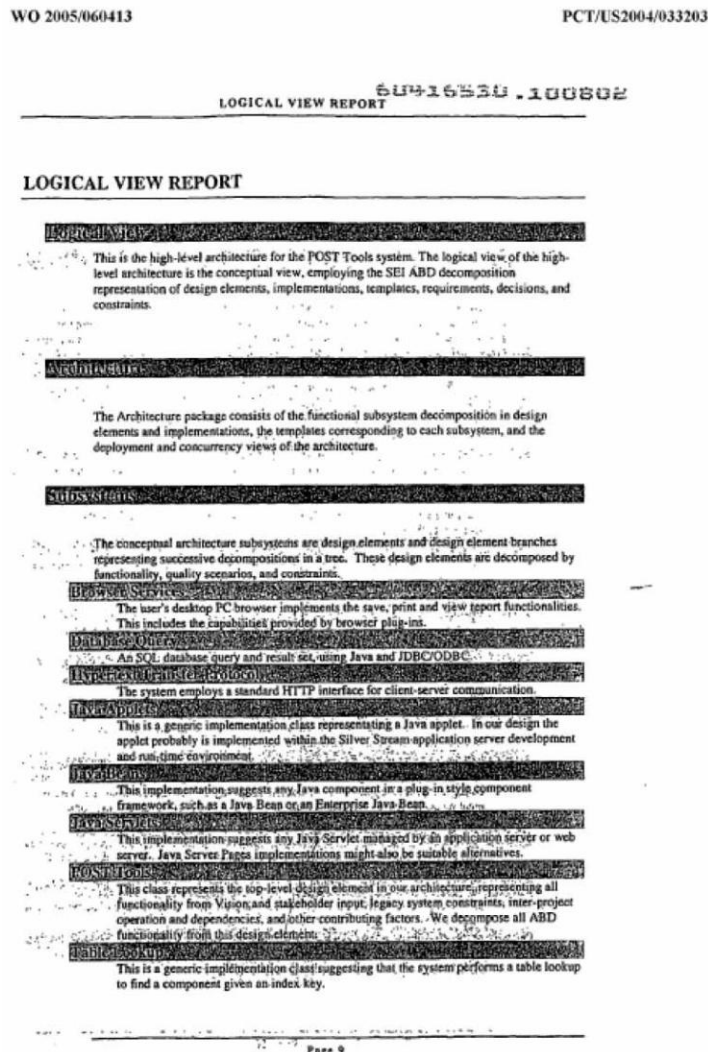
[0139] At step S7-8, solver 244 determines whether a predetermined number of iterations of the processing at steps S7-2 to S7-8 have been performed. In this embodiment, solver 244 determines whether 50 iterations have been performed.

Абзаци тексту не містять нетипових математичних символів. Математичні формули розміщені належним чином, дозволяючи системам OCR легко виділяти вбудовані зображення. Можливий результат їх виділення показано синьою лінією.

Додаток 2, сторінка 4

Зразки, що не відповідають рекомендаціям даного стандарту

Приклад 1: сторінка поганої якості з великою кількістю дефектів



Цей приклад не відповідає вимогам параграфу 10 (ймовірно, сторінка була надіслана у відомство факсом з роздільною здатністю 200 точок/дюйм – проглядається «шум» – і деякий текст надруковано на інтенсивному сірому фоні). Цей приклад також не відповідає вимогам параграфів 13 і 47: присвоєний номер (60415530.100802) проштамповано на основній ділянці сторінки (він повинен розміщуватися на полях). Нумерація сторінок неправильна (має бути «9» замість «сторінка 9», див. параграф 34). І нарешті, розмір шрифту занадто малий (параграф 15). Взагалі-то відомства не повинні приймати подібні сторінки і мають вимагати сторінки для заміни (неможливо виконати OCR належним чином).

Додаток 2, сторінка 5

Приклад 2: сторінка з небілим фоном

WO 2005/097403

- 13 -

PCT/FR2005/050194


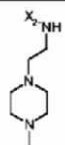
REVENDEICATIONS

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

Цей приклад не відповідає вимогам параграфу 12. Сторінка потребує фільтрації, щоб спробувати видалити фоновий «шум» перед передачею її на процедуру OCR. Якщо виконати OCR на цій сторінці, як вона є, отриманий текст буде непридатним для читання.

Додаток 2, сторінка 6

Приклад 3: сторінка з нечіткими символами

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

Beispiele 26-40

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

5 beschrieben. Das für die Darstellung des Amids ei

Фрагмент сторінки збільшено, щоб показати символи: ймовірно, колір оригінального тексту – сірий, що в результаті чорно-білого сканування з роздільною здатністю 300 точок/дюйм призводить до втрати цілісності символів. Як наслідок, точність розпізнаного тексту дуже низька (цей приклад не відповідає вимогам параграфу 12).

Приклад 4: сторінка з рукописним текстом

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

Цілком імовірно, що текст, отриманий після OCR на цій сторінці, буде непридатним для читання. Для забезпечення мінімальної якості публікації відомства повинні вимагати машинописний текст.

Додаток 2, сторінка 7

Приклад 5: сторінка з nereкомендованою розміткою та іншими недоліками

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.

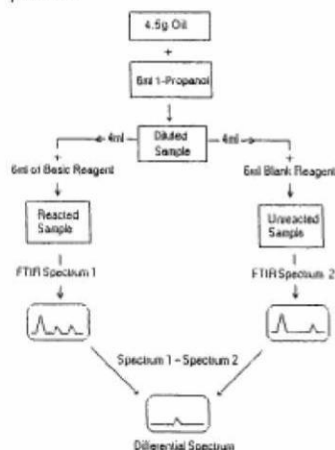


Figure 2. Analytical protocol for the determination of AN by FTIR spectroscopy.

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.

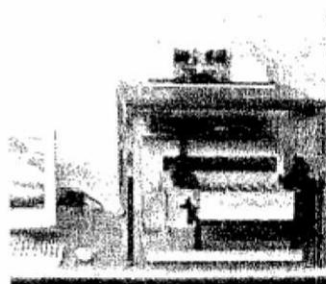


Figure 3. The COAT AN/BN Analyzer and its key components

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

Ця сторінка не відповідає таким рекомендаціям: параграф 17 (форматування в один стовпчик), параграф 39 (використання курсивних і напівгрубих шрифтів), параграф 46 (присутні виправлення, виконані вручну після друкування). Вирівнювання абзацу по ширині також не рекомендовано (параграф 30), хоча у даному випадку це не призвело б до негативних наслідків після OCR, оскільки слова все одно достатньо відокремлені білими проміжками. І, нарешті, цей приклад не відповідає вимогам параграфу 27 (міжрядковий інтервал 1,5).

Додаток 2, сторінка 8

Приклад 6: сторінка із надто маленьким розміром номерів рядків

WO 2004/110497 PCT/US2004/013820

[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);

5 [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);

10 [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);

15 [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 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);

20 [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;

25 [0033] Fig. 13A is a plot showing the concentration of mitomycin C in the blood of

30

6

Номери рядків становлять проблему для систем OCR з декількох причин (параграф 14):

- вони можуть бути не на одному рівні з відповідними рядками, що призводить до помилок у визначенні основного рядка;
- вони можуть бути занадто малими, що призведе до помилок під час розпізнавання і перешкоджатиме XML процедурам видобування усунути їх належним чином з текстової частини сторінки;
- вони можуть бути помилково розміщені всередині тексту сторінки або на полях, але занадто близько до текстової частини, і, як наслідок, з'являтимуться в текстовому потоці, що створюється під час процедури OCR.

У даному прикладі вони занадто малі.

У цьому прикладі також занадто малі символи нижнього індексу (параграф 35).

Додаток 2, сторінка 9

Приклад 7: сторінка, що містить декілька напрямків тексту

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	Cellulose acetate film 25		6	14								15	94.6	1.2	6.2	1.3
This invention	Cellulose acetate film 26	3		14								15	94.6	1.2	7.0	3.2
This invention	Cellulose acetate film 27	5	5	10								15	94.6	1.2	6.2	0.8
This invention	Cellulose acetate film 28		5			15	4.8	10.2					94.6	1.2	6.2	0.8
This invention	Cellulose acetate film 29		10			10	4.8	10.2					94.6	1.2	6.2	0.8
This invention	Cellulose acetate film 30					15	4.8	10.2					94.6	1.2	6.2	0.8
Comparative example	Cellulose acetate film 31	10	10										94.6	1.2	6.6	0
Comparative example	Cellulose acetate film 32			20									94.6	1.2	4.1	0
Comparative example	Cellulose acetate film 33									5	10		94.6	1.2	0	6.3
Comparative example	Cellulose acetate film 34	10	10				10.5	4.5					94.6	1.2	7.1	0.8
Comparative example	Cellulose acetate film 35	10	10				10.5	4.5					94.6	1.2	7.1	0.8

Цей приклад не відповідає вимогам параграфу 20.

Одне з обмежень найкращих з існуючих систем OCR полягає у тому, що вони можуть зчитувати тільки один напрямок тексту в межах однієї сторінки (попередня обробка сторінки полягає у визначенні основної орієнтації тексту сторінки). Як наслідок, усі слова, що не співпадають з основним напрямком тексту, пропускаються. Проте на сторінці допускається присутність таблиць і навіть основного тексту з альбомною орієнтацією і з книжковим розміщенням приміток на полях (номер сторінки, номер заявки тощо).

Додаток 2, сторінка 10

Приклад 8: сторінка з комбінацією вбудованих математичних формул і тексту

WO 2005/116630

PCT/US2005/017216

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

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

In order that the change of angular momentum, $\Delta \mathbf{L}$, 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,

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

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
10 electric energy (3rd term), and the dissipated energy of a fluxon treading the orbitsphere (4th term), respectively,

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

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

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

the stored electric energy corresponding to the $\frac{\partial}{\partial t} \left[\frac{1}{2} \epsilon_s \mathbf{E} \cdot \mathbf{E} \right]$ 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 $n=1, \ell=0$

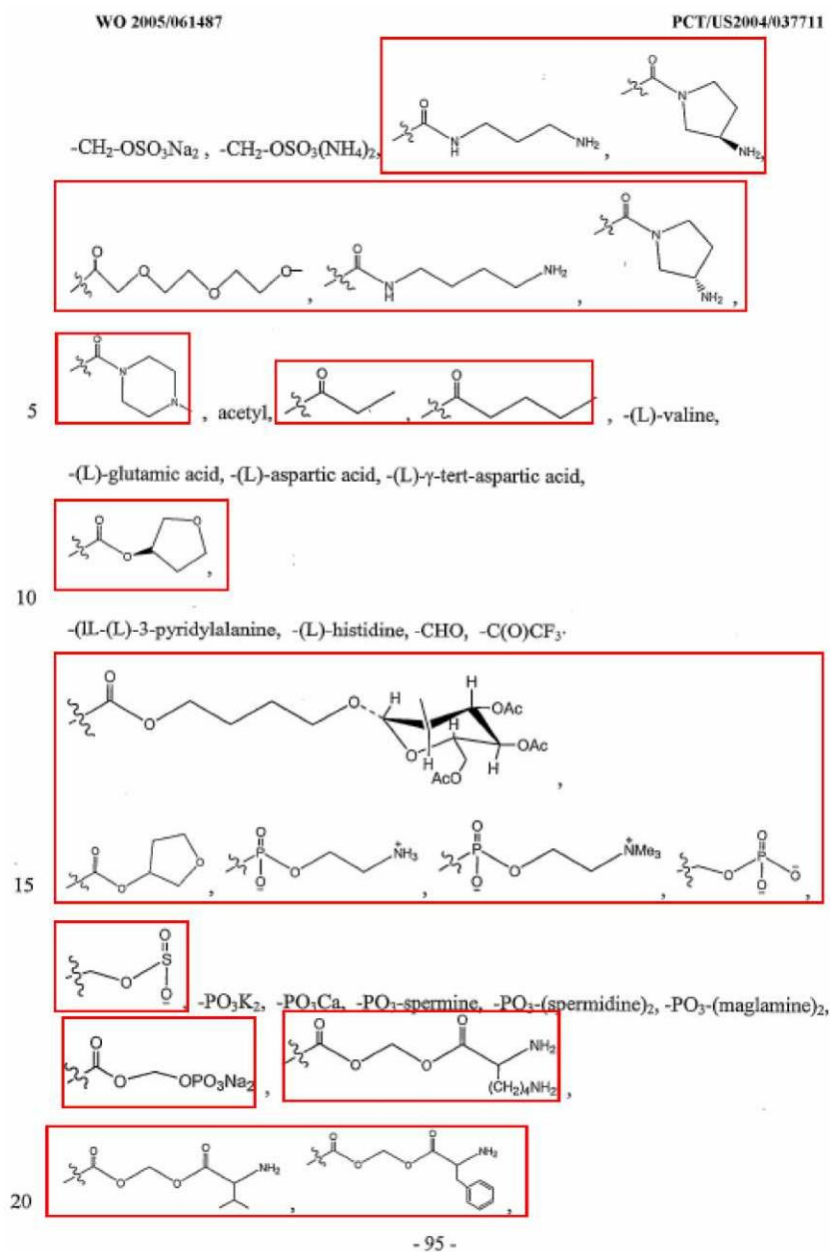
Цей приклад не відповідає вимогам параграфу 23. Система OCR не може належним чином відокремити текст від формул (див. результат ручної сегментації формул червоною лінією: вбудовані формули навіть накладаються).

Загалом, у даному прикладі текст і формули розміщені занадто щільно для належного розпізнавання; цей приклад також не відповідає вимогам параграфів 27 і 28.

Також у даному прикладі використовуються нестандартні символи: грецькі символи можуть використовуватися, навіть якщо вони ускладнюють розпізнавання сторінки (див. параграф 40). Однак, категорично не рекомендується комбінувати курсивні, напівгрубі і підкреслені шрифти з нестандартними символами (параграф 39).

Додаток 2, сторінка 11

Приклад 9: сторінка з комбінацією вбудованих хімічних формул і тексту



Цей приклад не відповідає вимогам параграфу 23. У червоних рамках можна побачити очікуваний результат сегментації зображень (виконаної вручну). Але така сегментація не може бути виконана системою OCR належним чином, оскільки формули розміщені занадто близько до оточуючого тексту.

Додаток 2, сторінка 12

Приклад 10: сторінка із занадто малими символами нижнього індексу

WO 2005/110416

PCT/US2005/015897

R_1 is hydrogen, C_1 - C_6 alkyl, C_2 - C_6 alkenyl, C_2 - C_6 alkynyl, C_1 - C_6 alkoxy, C_1 - C_6 haloalkyl, C_1 - C_6 haloalkoxy, $(C_3$ - C_7 cycloalkyl) C_3 - C_4 alkyl;

R_3 is selected from alkoxy, cycloalkoxy, phenyl, 4- to 7-membered heterocycles, $-O(CH_2)_n$ phenyl, $-O(CH_2)_n$ pyridyl, $-E-(CR_4R_5)_n-Q$, and Q , each of which is substituted with between 0 and 3 substituents selected from halogen, cyano, hydroxy, oxo, $(CR_4R_5)_j-T$, C_1 - C_4 alkyl, C_1 - C_4 alkoxy, C_1 - C_4 haloalkyl, C_1 - C_4 haloalkoxy, mono- and di- $(C_1$ - C_4 alkyl)amino, $(C_1$ - C_4 alkyl)((CR_4R_5) $_j-T$)amino, benzyl, $S(O)_n(C_1$ - C_4 alkyl), α,ω - C_1 - C_4 alkylene, α,ω - C_1 - C_4 alkyleneoxy, α,ω - C_1 - C_4 alkylenedioxy, $-E-(CH_2)_n-Q$, and Q ;

T is CO_2H , $CONH_2$, C_1 - C_6 alkoxycarbonyl, mono- or di- $(C_1$ - C_4 alkyl)aminocarbonyl, SO_3H , SO_2NH_2 or $SO_2(C_1$ - C_4 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;

R_8 and R_9 are independently chosen from hydrogen, halogen, hydroxy, C_1 - C_6 alkyl, C_1 - C_6 alkenyl, $(C_3$ - C_7 cycloalkyl) C_3 - C_4 alkyl and C_1 - C_6 alkoxy; 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 R_4 .

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, $-SO_2NH_2$, C_1 - C_6 alkyl, C_1 - C_6 alkenyl, C_1 - C_6 alkynyl, C_1 - C_6 haloalkyl, C_1 - C_6 aminoalkyl, C_1 - C_6 hydroxyalkyl, C_1 - C_6 carboxyalkyl, C_1 - C_6 alkoxy, C_1 - C_6 haloalkoxy, C_1 - C_6 alkylthio, C_1 - C_6 alkanoyl, C_1 - C_6 alkanoyloxy, C_1 - C_6 alkanone, C_1 - C_6 alkyl ether, mono- or di- $(C_1$ - C_4 alkyl)amino C_4 - C_6 alkyl, $-NHC(=O)(C_1$ - C_4 alkyl), $-N(C_1$ - C_4 alkyl) $C(=O)(C_1$ - C_4 alkyl), $-NHS(O)_n(C_1$ - C_4 alkyl), $-(C_1$ - C_4 alkyl) $C(=O)NH_2$, $-(C_1$ - C_4 alkyl) $C(=O)NH(C_1$ - C_4 alkyl), $-(C_1$ - C_4 alkyl) $C(=O)NH(C_1$ - C_4 alkyl)(C_1 - C_4 alkyl), $-S(O)_n(C_1$ - C_4 alkyl), $-S(O)_nNH(C_1$ - C_4 alkyl), $-S(O)_nN(C_1$ - C_4 alkyl)(C_1 - C_4 alkyl) and Z ; or

Це типовий приклад, коли символи нижнього індексу занадто малі для їх точного розпізнавання. Такий випадок часто зустрічається у патентах в галузі хімії.

Додаток 2, сторінка 13

Приклад 11: сторінка з погано відформатованою таблицею

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	NH	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 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.

10	Left-hand substituent	Left-hand ring	Aryl or heteroaryl	Ring substituent	Nitrogen feature	Right-hand substituent
	CH ₃	morpholino	aryl	CH ₂	NH	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

У цьому прикладі відсутні границі таблиці (не відповідає вимогам параграфу 33). Як наслідок, система OCR намагатиметься розпізнати вміст таблиці як текстовий абзац. Це призведе до декількох інших проблем:

- Занадто малий розмір шрифту символів у таблиці (параграфи 34 і 35).
- Змішення основних ліній заголовків стовпців (параграф 29). Як наслідок, система помилково розпізнаватиме символи нижнього і верхнього індексу.
- Отриманий текстовий потік не враховуватиме стовпчики:

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

Додаток 2, сторінка 14

Приклад 12: сторінка, вирівняна по ширині

WO 2005/087962

1

PCT/EP2005/002268

GKSS-Forschungszentrum Geesthacht GmbH, Max-Planck-Stra-
Be 1, 21502 Geesthacht

Verfahren zur Herstellung von Profilen aus Leichtmetall-
werkstoff 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

У цьому прикладі до абзаців застосовується вирівнювання по ширині. Хоча це надає тексту більш естетичного вигляду, іноді це викликає труднощі під час процедури OCR, коли проміжки між словами стають занадто малими (параграф 30). Цей приклад також не відповідає вимогам параграфу 31, який визначає, що необхідно, наскільки це можливо, уникати поділу слова наприкінці рядка (іноді це становить проблему для системи OCR під час розрізнення жорстких і м'яких переносів, внаслідок чого слова містять небажані переноси).

Додаток 2, сторінка 15

Приклад 13: таблиця з неналежними границями

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

У цьому прикладі границі таблиці в оригіналі, отриманому перед скануванням, вже були поганої якості. Після сканування процедура OCR не здатна розпізнати таблицю належним чином, і необхідно проводити ручне сегментування сторінки. Якщо подібна сторінка не перевірятиметься оператором на якість, вихідний текст міститиме небажані «переривчасті» символи, що призведе до зниження ефективності індексації документа пошуковими системами.

Додаток 2, сторінка 16

Приклад 14: неналежні символи нижнього і верхнього індексу

WO 2005/100305

PCT/IB2005/000872

-9-

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

- 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, SR⁸, and NR⁸R⁹, in which R⁸ and R⁹ are as defined above.
- 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, isochroamyl and quinolinyl.
- 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, SR⁸, and NR⁸R⁹, in which R⁸ and R⁹ are as defined above. Any nitrogen atom within such a heterocyclic ring

У цьому прикладі наявні наступні недоліки (параграф 35):

- Символи нижнього і верхнього індексу занадто малі
- Символи нижнього індексу розміщені занадто низько відносно основного рядка
- Символи верхнього індексу розміщені занадто високо відносно основного рядка

Як наслідок, рядки 34 і 35 розпізнаються під час OCR наступним чином:

"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"

Додаток 2, сторінка 17

Приклад 15: приклад з нестандартними символами

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 displacem

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 , 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 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 samples;

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

У цьому прикладі присутні наступні недоліки:

- Нестандартні символи: курсивні грецькі літери і навіть символи з тильдою
- Символи нижнього індексу тут знову занадто малі

У більшості систем OCR усі нестандартні символи не розпізнаватимуться належним чином.

Додаток 2, сторінка 18

Приклад 16: приклад з вузькими шрифтами і вузькою розрядкою

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.

Цей приклад не відповідає вимогам параграфів 37 і 38. Як наслідок, система OCR не може розрізнити межі слів належним чином і, як наслідок, процедура OCR взагалі непридатна для застосування.

Додаток 2, сторінка 19

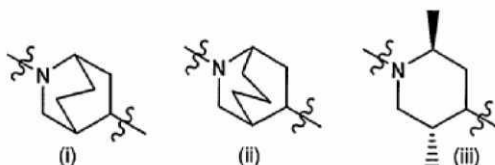
Приклад 17: невдале проставлення штампа відомством, що приймає заявку, перед скануванням

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:



Цей приклад не відповідає вимогам параграфу 47. Як наслідок, перші шість слів тексту на сторінці не можуть бути зчитані за допомогою процедури OCR. Крім того, штамп призводить до надмірного спотворення символів, що заважатиме роботі систем індексування, якщо якість сторінки не буде перевірена оператором.

Додаток 2, сторінка 20

Приклад 18: ще одна сторінка з невдалим розташуванням математичних формул

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,

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

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

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

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

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

and,

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

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,

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

$D(\varepsilon, \hat{\varepsilon})$ 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,

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

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,

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

THE FREQUENCY INTERPOLATION FUNCTION

Оскільки ця сторінка не відповідає багатьом рекомендаціям, результат OCR не придатний для використання:

- вбудовані математичні формули, не відокремлені від текстових абзаців (параграф 23);
- нестандартні символи у текстових абзацах (параграф 40);
- поєднання курсивного стилю з символами грецького алфавіту (параграф 39).

Рекомендований шлях для форматування цієї сторінки полягає у використанні додаткових проміжків для відокремлення вбудованих формул від абзаців. Грецькі літери у формулах і абзацах не повинні бути курсивними. Необхідно, наскільки це можливо, уникати використання циркумфлексів (^) для позначення змінних у текстових абзацах: замість них можуть використовуватися символи верхнього індексу: ε^{\wedge} («іпсилон циркумфлекс») може бути представлений як $\varepsilon^{\text{циркумфлекс}}$.

Додаток 2, сторінка 21

Приклад 19: сторінка з курсивними і підкресленими символами

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

15 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)*

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

Це проблема, що часто зустрічається в публікаціях РСТ. Дана сторінка не відповідає наступним рекомендаціям:

- Параграф 41: текст не повинен бути підкресленим. Підкреслювання особливо не рекомендується для хімічних формул (у таких випадках словники не допоможуть). Особливо це викликає проблеми зі всіма символами, що пересікаються з лінією підкреслення:]] у р – вони не розпізнаються належним чином.
- Параграф 39: не рекомендується використовувати курсив. Особливо не рекомендується змінювати стиль шрифту всередині слова (системи OCR часто припускають, що всі літери у слові мають однаковий стиль шрифту). Як наслідок, усі «1*H*» і «-*N*-» не розпізнаються належним чином.

Додаток 2, сторінка 22

Приклад 20: абсолютно непридатна для читання сторінка

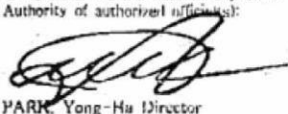
WO 2005/071074

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: Magam Biotechnology Research Institute
#311, Hujung-n, Kamsung-eup, Yongin-city, Kyonggi-do 449-910,
Republic of Korea

I. IDENTIFICATION OF THE MICROORGANISM	
Identification reference given by the DEPOSITOR: Saccharomyces cerevisiae HJ3501/MILK8 #36	Accession number given by the INTERNATIONAL DEPOSITORY AUTHORITY: KCTC J05828P
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 (Mark with a cross where applicable)	
III. RECEIPT AND ACCEPTANCE	
This International Depository 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 Depository Authority on _____ and a request to convert the original deposit to a deposit under the Budapest Treaty was received by it on _____	
V. INTERNATIONAL DEPOSITORY AUTHORITY	
Name: Korean Collection for Type Cultures	Signature(s) of person(s) having the power to represent the International Depository Authority of authorized official(s):  PARK Yong-Hu Director Date: January 17 2004
Address: Korea Research Institute of Bioscience and Biotechnology (KRIBB) #52 Oun-dong, Yusong-ku, Taejeon 305-380, Republic of Korea	

Form 1001 (BCTC) Form 1001

Ця сторінка не повинна прийматися відомствами: вона була надіслана факсом з роздільною здатністю 100 точок/дюйм і не придатна для читання навіть оком людини. У таких випадках оператори визнають увесь вміст сторінки зображенням, оскільки текст видобути неможливо.

[Кінець додатка 2 і стандарту]