

OMPI



SCIT/SDWG/10/3

ORIGINAL : anglais

DATE : 16 octobre 2008

F

ORGANISATION MONDIALE DE LA PROPRIÉTÉ INTELLECTUELLE
GENÈVE

COMITÉ PERMANENT DES TECHNIQUES DE L'INFORMATION

GROUPE DE TRAVAIL SUR LES NORMES ET LA DOCUMENTATION

Dixième session
Genève, 17 – 21 novembre 2008

RÉVISION DE LA NORME ST.22 DE L'OMPI (TÂCHE N° 37)

Document établi par le Secrétariat

1. Le Groupe de travail sur les normes et la documentation (SDWG) du Comité permanent des techniques de l'information (SCIT), à sa septième session tenue en mai-juin 2006, est convenu de créer une tâche aux fins de la révision de la norme ST.22 de l'OMPI (tâche n° 37), qui concerne des recommandations destinées à faciliter la reconnaissance optique des caractères (ROC) à l'intention des déposants qui déposent une demande de brevet sur papier. Le SDWG a aussi décidé de créer l'Équipe d'experts chargée de la norme ST.22 pour s'occuper de cette révision. En outre, le SDWG a demandé à l'équipe d'experts

a) d'examiner l'utilisation de caractères non latins, de sorte que la norme puisse s'appliquer ultérieurement à des langues n'utilisant pas des caractères latins;

b) d'examiner le taux d'exactitude de la reconnaissance optique des caractères (plus de 98,5% actuellement) afin d'envisager la possibilité d'indiquer un taux d'exactitude plus élevé;

c) de tenir compte des types de police de caractères et des tailles de police de caractères qui devraient être recommandés afin d'assurer une qualité optimale en termes de lisibilité, de présentation à l'écran et de reconnaissance optique des caractères

(voir les paragraphes 32 à 38 du document SCIT/SDWG/7/9).

2. Dans le prolongement de ce qui précède, l'Équipe d'experts chargée de la norme ST.22 a élaboré une proposition relative à la révision de la norme ST.22 de l'OMPI en vue de son examen par le SDWG à sa huitième session, qui s'est tenue en mars 2007. Le SDWG est toutefois convenu de différer l'adoption de la version révisée de la norme ST.22 de l'OMPI dans un souci d'harmonisation avec les travaux menés, à cette époque, dans le cadre du projet relatif au format unique des demandes pour les offices de la coopération trilatérale (Office des brevets et des marques des États-Unis, OEB et Office japonais des brevets), susceptibles d'avoir un effet sur cette norme. Le SDWG a aussi donné instruction à l'Équipe d'experts chargée de la norme ST.22 de prendre en considération les questions ci-après qui n'étaient pas encore traitées dans la version proposée de cette norme : dessins intégrés, photographies (échelle de gris et couleur), jeux de caractères, caractères non latins et formulaires présentés avec les demandes de brevet.

3. Le format unique des demandes a été publié sur le site Web des offices de la coopération trilatérale le 30 novembre 2007.

4. À sa neuvième session, tenue en février 2008, le SDWG a pris note de l'exposé présenté oralement par le responsable de l'Équipe d'experts chargée de la norme ST.22 sur le travail effectué par cette équipe d'experts en ce qui concerne la révision de la norme ST.22 de l'OMPI. Le responsable de l'équipe d'experts a indiqué au SDWG que le seul risque de conflit entre la proposition et le cadre publié avait trait à la numérotation des revendications dans les demandes de brevet. Le SDWG a demandé à l'Équipe d'experts chargée de la norme ST.22 de poursuivre les discussions afin de présenter la version de la norme ST.22 de l'OMPI applicable aux langues utilisant des caractères latins pour approbation à sa prochaine réunion et de commencer les travaux relatifs aux recommandations à formuler pour les langues qui ne s'écrivent pas en caractères latins.

5. Les résultats des discussions de l'Équipe d'experts chargée de la norme ST.22 sont présentés dans la révision proposée de la norme ST.22 de l'OMPI qui figure dans l'annexe du présent document en vue de son examen et de son approbation par le SDWG. Dans l'élaboration de cette proposition, l'Équipe d'experts chargée de la norme ST.22 a déjà tenu compte de toutes les observations reçues en ce qui concerne les caractères non latins, y compris les débats qui ont eu lieu pendant sa réunion informelle tenue au cours de la semaine pendant laquelle s'est déroulée la neuvième session du SDWG.

6. Il convient de noter que, à sa septième session, le SDWG est convenu que la norme ST.22 révisée de l'OMPI devrait être portée à l'attention des déposants, des offices, des fournisseurs commerciaux d'informations et des conseils en brevets (voir le paragraphe 39 du document SCIT/SDWG/7/9).

7. *Le SDWG est invité*

a) *à examiner et à adopter la proposition relative à la révision de la norme ST.22 de l'OMPI, qui figure dans l'annexe du présent document, et*

b) *à envisager, s'il le juge approprié, de demander au Bureau international de diffuser une circulaire pour mettre au courant les offices de propriété industrielle et les membres du SDWG de la révision de la norme ST.22 de l'OMPI et pour inviter les offices de propriété industrielle à publier un avis dans leur bulletin des brevets et sur leur site Web afin de porter cette norme révisée à l'attention des déposants, des fournisseurs commerciaux d'informations et des conseils en brevets relevant de leur zone géographique respective, ainsi que cela est indiqué dans le paragraphe 6 ci-dessus.*

[L'annexe suit]

ANNEXE

NORME ST.22

RECOMMANDATION RELATIVE AUX MODALITÉS D'ÉTABLISSEMENT DES DEMANDES DE BREVET
EN VUE DE FACILITER LA RECONNAISSANCE OPTIQUE DES CARACTÈRES (ROC)

Proposition élaborée par l'Équipe d'experts chargée de la norme ST.22

INTRODUCTION

1. La présente recommandation s'applique aux demandes de brevet soumises sur papier ou par la voie électronique (dépôt électronique) mais dont le corps de la demande est soumis dans un format image (p. ex., images PDF ou TIFF).
2. La présente recommandation a été élaborée en vue de faciliter l'établissement de demandes de brevet dactylographiées sous une forme qui permette par la suite de constituer un enregistrement numérisé électronique des informations figurant dans la demande de brevet au moyen d'un matériel de reconnaissance optique des caractères (ROC).
3. La présente recommandation tient compte de l'expérience acquise par différents offices qui utilisent du matériel de ROC. Elle a été rédigée en vue de limiter au maximum les erreurs lors de la lecture automatique du texte des demandes de brevet tout en permettant à la personne qui le souhaite de lire commodément le document. Il convient de noter que le présent document ne contient pas de recommandations détaillées pour le japonais et le coréen; le pourcentage du nombre des dépôts électronique en texte intégral par rapport au nombre total de dépôts par année dépasse effectivement 90% dans ces pays, ce qui signifie que cette norme n'est pas applicable concrètement pour le japonais et le coréen.
4. L'établissement d'un enregistrement numérisé d'une demande de brevet a pour principal objectif de permettre une publication facile de cette demande, selon un format établi, au moyen des techniques de photocomposition informatisée, et d'améliorer ainsi la présentation et la qualité des documents de brevet dans l'intérêt de tous les utilisateurs. L'objectif est aussi de créer une base de données déchiffrable par machine correspondant au texte complet des documents publiés afin de pouvoir ultérieurement tirer parti des possibilités qu'offre une recherche sur ordinateur fondée sur le texte complet.

DÉFINITION

5. Aux fins de la présente recommandation, l'expression "demande de brevet" désigne les demandes de brevet d'invention, de certificat d'auteur d'invention, de certificat d'utilité, de modèle d'utilité, de brevet ou de certificat d'addition, de certificat d'auteur d'invention additionnel et de certificat d'utilité additionnel.
6. Une formule mathématique ou chimique est réputée "complexe" lorsqu'elle ne peut pas être affichée comme une séquence linéaire de caractères, chaque caractère ayant un attribut facultatif en indice ou en exposant. Une formule est réputée particulièrement complexe lorsqu'elle contient des indices ou des exposants imbriqués ou les symboles mathématiques de la somme, de l'intégrale ou du produit.
7. Une matrice de caractère ou jeu de caractères est le plus petit rectangle aligné par rapport à l'axe qui comprend toutes les parties du caractère ou du jeu de caractères.
8. On entend par "cursive" une forme stylisée d'écriture dans lequel les lettres des mots sont liées les unes aux autres, un mot correspondant à une seule barre complexe. Les polices de caractères sont dites cursives si elles sont conçues pour ressembler à l'écriture manuscrite.

CRÉATION DE L'ORIGINAL

9. Les demandes de brevet seront souvent établies à l'aide de machines de traitement de texte. L'expérience montre que c'est avec les caractères définis dans la norme 1073/II de l'Organisation internationale de normalisation (ISO), dits ROC-B, que le matériel de ROC est le plus fiable.

SCIT/SDWG/10/3
Annexe, page 2

SUPPORT PAPIER POUR AUTANT QUE LA DEMANDE SOIT DÉPOSÉE SUR PAPIER

10. Afin de faciliter la numérisation, le support papier de la demande dactylographiée doit présenter les caractéristiques suivantes :

- a) papier fort, blanc et propre;
- b) le poids du papier doit être compris entre 70 et 120 g/m²;
- c) le papier doit être de préférence de format A4, c'est-à-dire de 210 mm x 297 mm ou de 8 ½ x 11 pouces (ce qui est la norme nord-américaine de facto);
- d) les feuilles ne doivent pas comporter de pliure, ni de trou, ni être roulées;
- e) le papier ne doit pas être absorbant afin d'éviter toute tâche d'encre (par exemple, avec une imprimante à jet d'encre).

RECOMMANDATIONS RELATIVES À LA PRÉSENTATION DES PAGES

11. Il convient d'éviter l'impression recto-verso. Si cela n'est pas possible, un papier suffisamment opaque devrait être utilisé en vue de permettre une numérisation propre recto-verso.
12. Les caractères doivent être noir intense, sur fond blanc.
13. Une marge minimale de 2 centimètres doit être ménagée en haut et en bas ainsi que sur le côté droit de chaque feuille, la marge de gauche devant être de 2,5 centimètres au minimum. Toutes les mentions concernant le déposant ou son mandataire doivent figurer dans la marge supérieure. On se reportera à l'appendice 1.
14. Il convient d'éviter toute numérotation de ligne. Si ce n'est pas possible, il convient d'utiliser pour ce faire des chiffres arabes, dans la marge de gauche, à 1 cm au moins de l'encadré, comme indiqué dans l'appendice 1. La taille de la police de caractères utilisée pour les chiffres doit être de 12 points au moins.
15. Les pages doivent être numérotées à l'aide de chiffres arabes, à l'exclusion de tous autres caractères de séparation. Les numéros de page doivent de préférence être centrés en haut ou en bas de la feuille, dans la marge, comme indiqué dans l'appendice 1.
16. Les textes de la description, de la ou des revendications et de l'abrégé doivent commencer chacun sur une nouvelle page. En outre, le premier mot imprimé sur la première page de chacune des trois parties précitées de la demande doit indiquer la partie dont il s'agit (dans la langue de la demande); le paragraphe des revendications doit être numéroté en continu. La présentation de la numérotation des revendications doit permettre de distinguer clairement entre le numéro de la revendication et le texte de la revendication pour chaque revendication. Il est recommandé d'utiliser ou bien des chiffres arabes suivis d'un point ou de mentionner le terme "revendication" - ou le mot équivalent dans la langue de la demande de brevet - suivi d'un espace et du chiffre arabe correspondant à la revendication, le texte suivant de la revendication étant renforcé à droite s'agissant du numéro de la revendication à raison d'au moins 1 centimètre dans les deux cas.
17. Les pages doivent comporter des paragraphes en colonne unique (paragraphes de texte ou paragraphes contenant une image intégrée).
18. Les pages contenant des paragraphes doivent être disposées verticalement (format à la française).
19. Il convient d'éviter toute disposition horizontale (format à l'italienne). Ce type de disposition ne peut être accepté que pour les pages contenant des dessins ou des tableaux intégrés, qui ne peuvent pas être disposés verticalement.
20. Sur chaque page, le texte doit être disposé d'une seule façon.
21. Les pages disposées horizontalement doivent faire l'objet d'une rotation à 90 degrés dans le sens inverse des aiguilles d'une montre afin d'être incorporées dans l'ensemble des pages disposées verticalement.
22. Il est recommandé d'éviter le recours à des notes de bas de page, à du texte dans la marge et à des en-têtes, sauf dans les cas indiqués au paragraphe 14 (numérotation de ligne), au paragraphe 15 (numérotation de page) et lors de l'indication de la référence du dossier du déposant dans le coin supérieur gauche de la marge.

RECOMMANDATIONS RELATIVES À LA PRÉSENTATION DES PARAGRAPHES

23. Il est recommandé que les tableaux, les formules chimiques ou mathématiques complexes, les images et les dessins soient séparés des paragraphes de texte. Il est conseillé de laisser de part et d'autre de ces éléments une marge supérieure et une marge inférieure vierge d'au moins 1 centimètre couvrant toute la largeur de la page.
24. Les images et les dessins devraient, au maximum, figurer dans la partie "dessins" et faire l'objet d'un renvoi dans les parties "description" et "revendications" de la demande de brevet.

25. Les images et les dessins devraient être en noir et blanc (les images en échelle de gris devraient être évitées, l'information contenue étant perdue lors de leur numérisation ou de leur conversion en noir et blanc). Les figures devraient contenir des lignes nettes suffisamment larges pour être bien représentées avec une résolution de 300 ppp.
26. Les paragraphes de texte manuscrits et les annotations doivent être évités. Si cela n'est pas possible, ces paragraphes manuscrits et annotations seront traités comme des dessins intégrés et devront respecter la recommandation visée au paragraphe 23.
27. Le texte dactylographié doit présenter un intervalle d'une ligne et demie.
28. Les paragraphes doivent être séparés les uns des autres par un espace correspondant à deux fois au moins l'intervalle entre chaque ligne.
29. Tous les caractères d'une ligne de paragraphe doivent être soigneusement alignés sur la ligne de base, sauf en ce qui concerne les caractères en indice et en exposant comme cela est indiqué au paragraphe 35.
30. Les paragraphes ne doivent pas être justifiés. Si cela n'est pas possible, l'espace entre deux mots doit être au moins aussi important que celui pour du texte non justifié. La justification de texte peut empêcher les systèmes de ROC d'identifier correctement les frontières de mot dans un paragraphe.
31. Lorsque cela est possible, il convient d'éviter les divisions de mot au moyen d'un trait d'union (par exemple, en fin de ligne ou de cellule). Cela ne s'applique pas aux langues comportant des noms composés (telles que l'allemand).

RECOMMANDATIONS RELATIVES AUX TABLEAUX

32. Seul un fond blanc doit être utilisé.
33. Les tableaux doivent être délimités par un cadre. L'épaisseur du trait de ce cadre doit être supérieur à 1,5 point et le trait doit être continu.

RECOMMANDATIONS RELATIVES AUX POLICES DE CARACTÈRES

34. La taille de la police de caractères recommandée est de 12 points au moins, 14 points étant préférable. D'une manière générale, la taille de la police doit être la même pour tous les caractères d'un même paragraphe.
35. La taille de la police des paragraphes de texte contenant des indices ou des exposants doit être de 12 points au moins (la police recommandée étant de 14 points - plus les caractères sont gros, mieux c'est). Veiller à ce que la matrice de caractère de l'indice ou de l'exposant présente une intersection suffisante avec celle des caractères ordinaires de la même ligne (cela empêche les techniques de ROC de mettre les indices ou les exposants sur des lignes différentes).
36. Les polices de caractères recommandées sont, dans l'ordre, les suivantes :
- police de caractères à espacement fixe : OCR-B, Courier New, Free Mono;
 - police de caractères avec empattement : ITC Officina Serif, Times New Roman, Free Times;
 - police de caractères sans empattement : Verdana, ITC Officina Sans, Arial, Helvetica, DéjàVu Sans.

Toutefois, les polices Arial, Helvetica, DéjàVu Sans, Free Times et Times New Roman ne sont pas recommandées pour les demandes contenant des formules chimiques ou mathématiques ou des acronymes composés de lettres et de chiffres. Pour les caractères chinois, c'est la police Song qui est recommandée.

37. Les caractères de la police doivent être bien formés, sans ombre. L'espace entre deux caractères doit être suffisamment large (il convient d'éviter tout espacement restreint).
38. Les polices de caractères étroites et cursives ne doivent pas être utilisées.
39. Il convient d'éviter autant que faire se peut de recourir à des caractères gras et à des caractères italiques.
40. Il convient d'éviter autant que faire se peut les caractères inhabituels (non normalisés/atypiques/irréguliers). Si cela n'est pas possible, on utilisera l'alphabet grec normalisé et la police des symboles (dans cet ordre de préférence). Les caractères introuvables dans la palette UNICODE ne doivent pas être utilisés : ces caractères sont identifiés comme des images intégrées par les appareils de ROC et rendent donc le texte reconnu difficile à lire. Chaque office doit définir et publier ces exigences en ce qui concerne le jeu de caractères qui peut être utilisé aux fins de l'établissement des demandes de brevet.
41. Il convient d'éviter de souligner le texte. Si cela n'est pas possible, il convient de s'assurer que le soulignement ne touche pas la matrice des caractères soulignés.

RECOMMANDATIONS RELATIVES AUX LANGUES AUTRES QUE LES LANGUES LATINES

42. La combinaison de langues asiatiques (c'est-à-dire reposant sur des idéogrammes) et européennes (c'est-à-dire utilisant les alphabets latin et cyrillique) posant un problème aux fins de la reconnaissance optique, il convient d'éviter d'y avoir recours dans les sections et dans les pages des demandes de brevet.

RECOMMANDATIONS RELATIVES À LA NUMÉRISATION

43. Les demandes de brevet doivent être numérisées en noir et blanc ou en échelle de gris.

44. La résolution recommandée pour la numérisation est 300ppp. La numérisation à des résolutions inférieures à 300ppp même en échelle de gris, peut nuire à la qualité des documents publiés par les offices car l'échange des documents entre les offices et le processus de publication impliquent souvent des conversions dans des pages en noir et blanc de 300ppp en format TIFF groupe 4.

45. Les documents numérisés doivent être convertis soit en format PDF soit en format TIFF.

CORRECTIONS

46. Pour les corrections apportées au texte d'une demande, chaque page corrigée devrait être réimprimée en entier. Les signes de preuve de correction, tels que ceux qui sont définis dans la norme internationale ISO 5776, ne sont pas acceptés. Les moyens de correction tels que le liquide blanc correcteur ou le ruban adhésif, tout comme les ratures et les bifures, ne sont pas acceptés. Il convient d'éviter d'envoyer à l'office les pages de remplacement par télécopie en utilisant la résolution courante : si les pages ne peuvent pas être envoyées physiquement, elles doivent être transmises par télécopie en utilisant une résolution supérieure ou égale à 300ppp ou par tout autre moyen de transfert par réseau accepté par l'office, à condition que chaque page numérisée ait une résolution supérieure ou égale à 300ppp.

RECOMMANDATIONS RELATIVES AUX OFFICES

47. Les offices de brevets doivent éviter de modifier les pages reçues avant toute opération de numérisation ou de reconnaissance optique. Ainsi, il arrive que des tampons soient apposés sur les documents, ce qui peut entraîner une superposition de caractères sur les pages et rendre le texte soumis par le déposant illisible par les techniques de ROC. Si l'office doit tamponner ou modifier les pages originales, il doit le faire dans la marge des documents, comme indiqué dans l'appendice 1.

48. À l'avenir, les offices de brevets doivent éviter de concevoir des formulaires papier aux fins de la communication entre les déposants et l'office. Compte tenu de l'expérience passée, il est préférable de concevoir et de mettre en place des systèmes de formulaires en ligne sécurisés à la création de systèmes de reconnaissance de formulaires papier. Toutefois, les recommandations ci-après sont formulées en vue de la conception de formulaires papier en vue de faciliter leur reconnaissance :

- Il conviendrait de ne pas utiliser de lignes de points de suspension dans les formulaires pour indiquer à l'utilisateur l'endroit où le texte doit figurer.
- Des couleurs de moindre intensité devraient être utilisées pour les matrices de caractère (gris clair).
- Il conviendrait de ne pas utiliser des couleurs de moindre intensité pour les champs destinés à recevoir un contenu en niveaux de gris ou coloré tel qu'une signature numérisée ou un dessin.

APPLICATION

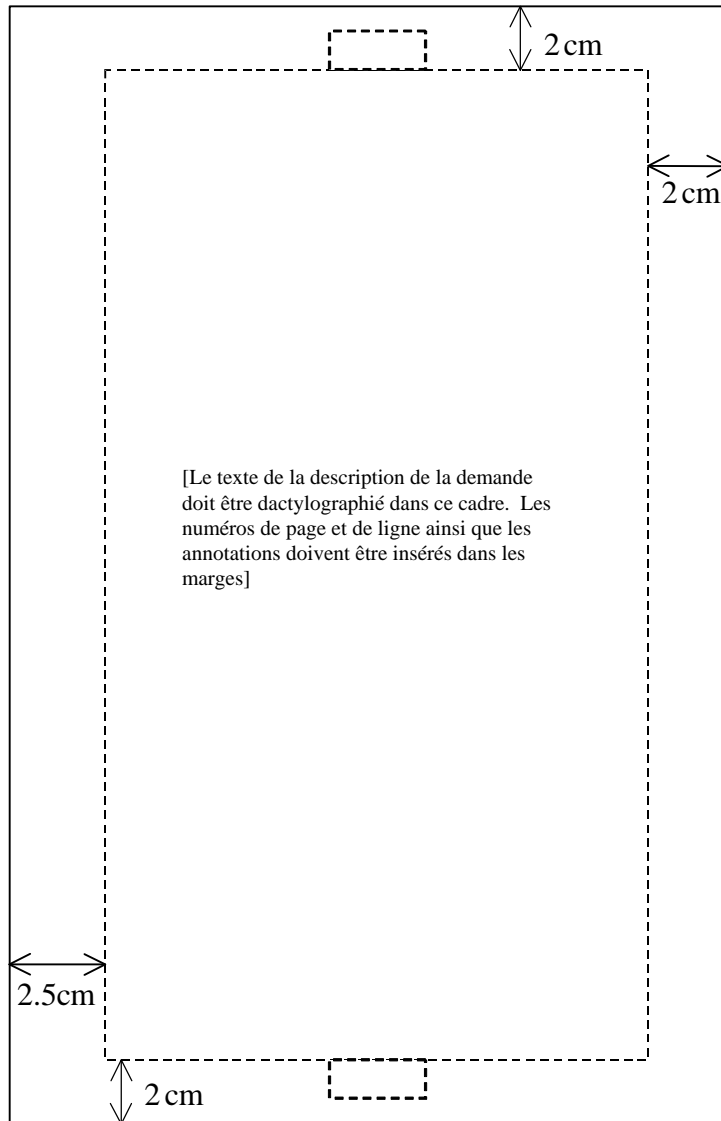
49. Il est recommandé que les offices ayant l'intention de commencer à accepter ou à demander le dépôt des demandes de brevet dactylographiées en format ROC publient des directives complètes dans leur bulletin officiel à intervalles réguliers ainsi que sur leur site Internet, en indiquant exactement le type ou les types de caractères admis ainsi que les dimensions exactes du papier.

Exemples

50. On trouvera dans l'appendice 2 de la présente recommandation des exemples de pratiques recommandées et de pratiques à éviter aux fins de la reconnaissance optique des caractères. Ces exemples illustrent ce qui doit être fait et ce qui doit être évité, et sont accompagnés d'une brève explication.

[Les appendices suivent]

APPENDICE 1



Dimension originale = A4

APPENDICE 2

EXEMPLES DE PRATIQUES RECOMMANDÉES ET DE PRATIQUES À ÉVITER

Le présent appendice contient des exemples de ce qu'il faut faire ou ne pas faire dans les pages de documents de brevet afin que les opérations de ROC donnent un résultat aussi précis que possible.

Exemples de pratiques recommandées

Exemple n° 1. Page de description correcte

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

15

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.

20

25

Exemple n° 2. Page de revendications correcte

WO 2006/111319

PCT/EP2006/003401

- 25 -

Claims

1. A projection exposure system, in particular for micro-lithography, comprising a light source (18) for producing light in the EUV region; a first
5 optical system (19, 20, 21, 22, 23, 24) for illuminating a mask (25) by the light of the light source (18); and a second optical system (26, 27, 28, 29, 30) for imaging the mask (25) on a structural member (32); wherein at least one polarization-optical element (1) for the EUV region is disposed on the beam path between the light source (18) and the structural member (32),
10 said at least one polarization-optical element comprising at least one reflective cone surface (3, 7, 12, 14), having a polarizing effect for the light produced by said light source.
2. Projection exposure system according to claim 1, wherein the polarization-optical element comprises
15 - at least one cone element (2) having an outer cone surface (3) being reflective for at least a given polarization component of the EUV light being incident under an angle of inclination α to the axis of rotational symmetry of the cone surface in a region of inclination angles α between 0° and a maximum angle of inclination such that the cone surface (3) has a polarizing effect, and
20 - having at least one further reflective component (4; 11, 13) for the EUV light to bundle the EUV light being reflected from the outer cone surface (3) of the cone element (2).
- 25 3. Projection exposure system according to claim 1 or 2, wherein the polarization-optical element comprises
- at least one further cone element (6) having an outer cone surface (3) being reflective for at least a given polarization component of the EUV

Toutes les recommandations ont été respectées : marges, police de caractères normalisée (Times New Roman), taille de la police correcte, pas de numéros de ligne, utilisation restreinte des caractères gras, pas de caractères italiques, pas de texte souligné, numérotation des revendications adéquate et bien séparée du texte des revendications.

Exemple n° 3. Page d'une description complexe correcte

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 (46)

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

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

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

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

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

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

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

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

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

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:

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

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.

Les paragraphes de texte ne contiennent pas de caractères mathématiques inhabituels. Les formules mathématiques sont correctement intégrées, permettant aux appareils de ROC de segmenter facilement les images intégrées. On peut voir en bleu un résultat possible de la segmentation.

SCIT/SDWG/10/3
Annexe, page 9
Appendice 2, page 4

EXEMPLES DE CE QU'IL NE FAUT PAS FAIRE

Exemple n° 1. Page de qualité médiocre, comportant de nombreuses lacunes

WO 2005/060413

PCT/US2004/033203

604115530.100802
LOGICAL VIEW REPORT

LOGICAL VIEW REPORT

LOGICAL VIEW REPORT

This is the high-level architecture for the POST Tools system. The logical view of the high-level architecture is the conceptual view, employing the SEI ABD decomposition representation of design elements, implementations, templates, requirements, decisions, and constraints.

ARCHITECTURE

The Architecture package consists of the functional subsystem decomposition in design elements and implementations, the templates corresponding to each subsystem, and the deployment and concurrency views of the architecture.

SUBSYSTEMS

The conceptual architecture subsystems are design elements and design element branches representing successive decompositions in a tree. These design elements are decomposed by functionality, quality scenarios, and constraints.

BROWSER SERVICES

The user's desktop PC browser implements the save, print and view report functionalities. This includes the capabilities provided by browser plug-ins.

DATABASE SERVICES

An SQL database query and result set, using Java and JDBC/ODBC.

HTTP PROTOCOL

The system employs a standard HTTP interface for client-server communication.

JAVA APPLET

This is a generic implementation class representing a Java applet. In our design the applet probably is implemented within the Silver Stream application server development and run-time environment.

JAVA BEANS

This implementation suggests any Java component in a plug-in style component framework, such as a Java Bean or an Enterprise Java Bean.

JAVA SERVLETS

This implementation suggests any Java Servlet managed by an application server or web server. Java Server Pages implementations might also be suitable alternatives.

FUNCTIONS

This class represents the top-level design element in our architecture, representing all functionality from Vision and stakeholder input, legacy system constraints, inter-project operation and dependencies, and other contributing factors. We decompose all ABD functionality from this design element.

TABLE LOOKUP

This is a generic implementation class suggesting that the system performs a table lookup to find a component given an index key.

Cet exemple n'est pas conforme au paragraphe 10 (la page a probablement été envoyée à l'office par télécopie, avec une résolution de 200 ppp – vous pouvez voir le “bruit informatique” – et des portions du texte apparaissent sur fond gris foncé). L'exemple n'est pas non plus conforme aux paragraphes 13 et 47 : un numéro de référence (604115530.100802) a été tamponné sur la page (le tampon devrait être dans la marge). La numérotation de la page est incorrecte (on devrait voir apparaître “9” et non “page 9” (voir le paragraphe 34)). Dernier point, la taille de la police est trop petite (voir le paragraphe 15). Dans l'idéal, l'office devrait refuser ces pages et exiger des pages de remplacement (cette page ne peut pas être lue correctement par les techniques de ROC).

Exemple n° 2. Page sans fond blanc

WO 2005/097403

- 13 -

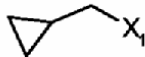
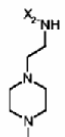
PCT/FR2005/050194

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

Cet exemple n'est pas conforme au paragraphe 12. Avant toute opération de ROC, la page doit être "filtrée" en vue d'essayer de supprimer le fond bruité. Si cette page est soumise telle quelle à un appareil de ROC, le texte obtenu sera illisible.

Exemple n° 3. Page comportant des caractères flous

#	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

Une petite partie de la page a été agrandie pour montrer les caractères; la couleur du texte original est probablement le gris, ce qui s'est traduit, après numérisation en blanc et noir à 300 ppp, par des caractères qui ne sont pas suffisamment intenses. Par conséquent, le texte résultant de la reconnaissance optique n'est pas précis (cet exemple n'est pas conforme au paragraphe 12).

Exemple n° 4. Page avec texte manuscrit

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

Comme on pouvait s'y attendre, le texte obtenu après ROC est illisible. Les offices doivent exiger que le texte soit dactylographié en vue d'obtenir une qualité minimale aux fins de la publication.

Exemple n° 5. Page dont la présentation n'est pas recommandée et comportant d'autres lacunes

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.

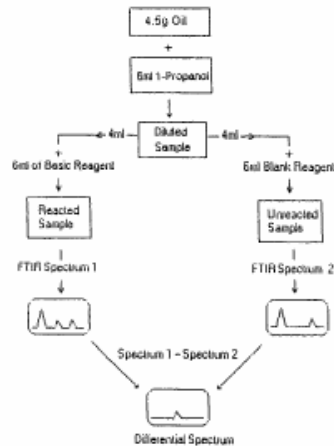


Figure 2. Analytical protocol for the determination of AN by FTIR spectroscopy.

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.

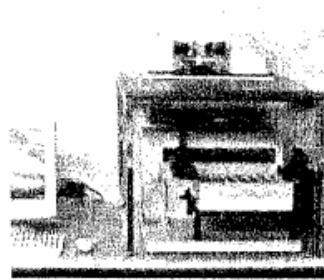


Figure 3. The COAT AN/BN Analyzer and its key components

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

Cette page n'est pas conforme aux recommandations suivantes : paragraphe 17 (formatage de colonne unique), paragraphe 39 (comporte des caractères italiques et des caractères gras), paragraphe 46 (corrections manuelles effectuées après l'impression). La justification à gauche et à droite du paragraphe n'est pas non plus recommandée (voir le paragraphe 30), même si, aux fins du présent exemple, il n'y aura aucune répercussion négative sur la ROC puisque les espaces entre les mots sont suffisants. Enfin l'exemple n'est pas non plus conforme au paragraphe 27 (intervalle d'une ligne et demie).

Exemple n° 6. Page dont les numéros de ligne sont trop petits

WO 2004/110497

PCT/US2004/013820

5 [0028] Figs. 9A-9B are plots showing the percent of mitomycin C released from liposomes comprised of HSPC/mPEG-DSPE/lipid-DTB-mitomycin C (Fig. 9A) and HSPC/cholesterol/mPEG-DSPE/lipid-DTB-mitomycin C (Fig. 9B) as a function of time of incubation in the presence of cysteine at concentrations of 150 μ M (closed symbols) and at 1.5 mM (open symbols);

10 [0029] Fig. 10 is a plot of growth rate of M109 cells, expressed as a percentage based on growth of M109 cells in the absence of drug and cysteine, as a function of mitomycin C amount, in nM, for free mitomycin c (open triangles), liposomes comprised of HSPC/mPEG-DSPE/lipid-DTB-mitomycin C (closed squares), and liposomes comprised of HSPC/cholesterol/mPEG-DSPE/lipid-DTB-mitomycin C (open circles);

15 [0030] Fig. 11A is a plot of growth rate of M109 cells, expressed as a percentage based on growth of M109 cells in the absence of drug or cysteine, as a function of mitomycin C concentration in nM. Shown are cells treated mitomycin C in free form (open triangles) and with mitomycin C in free form plus 1000 μ M cystein (closed triangles). Also shown are cells treated with the liposome formulation comprised of HSPC/PEG-DSPE/lipid-DTB-mitomycin C (open circles) and with the liposome formulation with additional cysteine added at concentrations of 150 μ M (open diamonds), 500 μ M (closed circles) and 1000 μ M (open squares);

20 [0031] Fig. 11B is a plot of growth rate of M109 cells, expressed as a percentage based on growth of M109 cells in the absence of drug or cysteine, as a function of mitomycin C concentration in nM. Shown are cells treated mitomycin C in free form (open triangles) and with mitomycin C in free form plus 1000 μ M cysteine (closed triangles). Also shown are cells treated with the liposome formulation comprised of 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);

25 [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;

30 [0033] Fig. 13A is a plot showing the concentration of mitomycin C in the blood of

Les numéros de ligne posent un problème aux appareils de ROC pour plusieurs raisons (voir le paragraphe 14) :

- ils peuvent ne pas se trouver dans l'alignement des lignes auxquelles ils correspondent, ce qui conduit à des erreurs de détection de la ligne de base;
- ils peuvent être trop petits, ce qui conduit à des erreurs de reconnaissance empêchant les procédures d'extraction XML de les séparer correctement du corps du texte dans la page;
- ils peuvent être mal placés dans la zone de texte ou se trouver dans la marge mais être trop près de la zone de texte et, par conséquent, ils figureront dans le flux de texte exporté à l'issue des opérations de ROC.

Dans le présent exemple, ils sont trop petits.

Les caractères en indice sont aussi trop petits (voir le paragraphe 35).

Exemple n° 7. Page comportant du texte disposé de plusieurs façons

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	
This invention		3		14							15	94.6	1.2	7.0	3.2	
This invention		5	5	10							15	94.6	1.2	6.2	0.8	
This invention	59.8		5			15	4.8	10.2				94.6	1.2	6.2	0.8	
This invention		10				10	4.8	10.2				94.6	1.2	6.2	0.8	
This invention						15	4.8	10.2				94.6	1.2	6.2	0.8	
Comparative example	60.9	10	10									94.6	1.2	6.6	0	
Comparative example	60.5			20								94.6	1.2	4.1	0	
Comparative example	60.9								5	10		94.6	1.2	0	6.3	
Comparative example	60.8	10	10				10.5	4.5				94.6	1.2	7.1	0.8	
Comparative example	60.8	10	10				10.5	4.5				94.6	1.2	7.1	0.8	

Le présent exemple n'est pas conforme au paragraphe 20

Les meilleurs appareils de ROC disponibles aujourd'hui ne peuvent néanmoins lire qu'une seule disposition de texte par page (une préprocédure consiste à détecter la disposition du texte principal de la page). Par conséquent, tous les mots qui ne sont pas disposés de la même manière que le texte principal sont ignorés. Bien entendu, il est possible d'avoir, sur une page, un tableau horizontal, ou même un texte principal horizontal avec des annotations verticales dans la marge (numéro de la page, numéro de la demande, etc.).

Exemple n° 8. Page avec des formules mathématiques intégrées et du texte

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, $\Delta \mathbf{L}$, 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,

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

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

- 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 $\frac{\partial}{\partial t} \left[\frac{1}{2} \mu_0 \mathbf{H} \cdot \mathbf{H} \right]$ term increases,

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

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

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

1.G. SPIN AND ORBITAL PARAMETERS

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

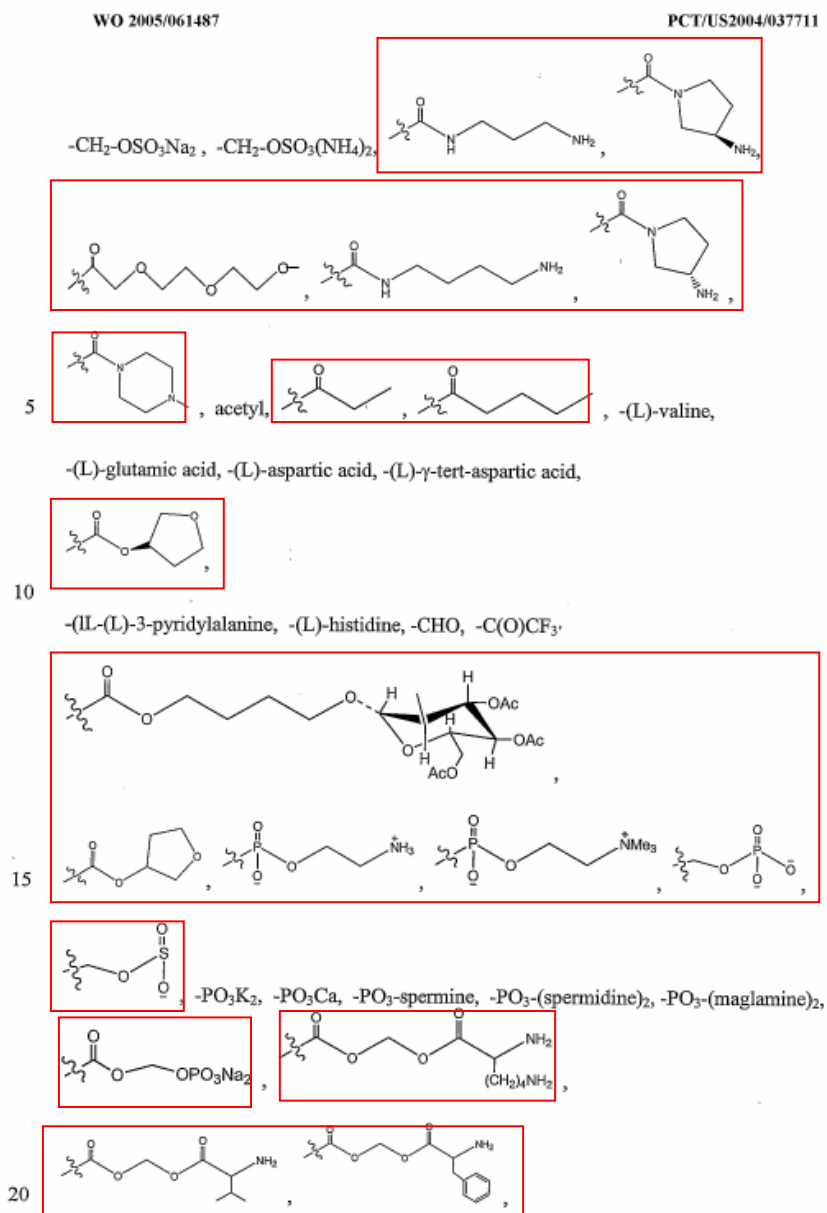
The spin function of the electron corresponds to the nonradiative $n = 1, \ell = 0$

Le présent exemple n'est pas conforme au paragraphes 23 L'appareil de ROC n'est pas en mesure de séparer correctement le texte des formules (voir le résultat de la segmentation manuelle de la formule qui s'affiche en rouge; les formules intégrées vont même jusqu'à s'interpénétrer).

D'une manière générale, on peut dire que, dans cet exemple, le texte et les formules sont trop denses pour pouvoir être bien reconnus; l'exemple n'est pas non plus conforme aux paragraphes 27 et 28.

Il est aussi fait usage de caractères inhabituels; les symboles grecs peuvent être utilisés même s'ils augmentent la difficulté de reconnaissance de la page (voir le paragraphe 40). Toutefois, il est vivement préférable d'éviter d'associer caractères italiques, caractères gras ou caractères soulignés à des caractères inhabituels (voir le paragraphe 39).

Exemple n° 9. Page avec des formules chimiques intégrées et du texte



Le présent exemple n'est pas conforme au paragraphe 23. On peut voir en rouge l'un des résultats attendus de la segmentation des dessins (effectuée manuellement). Cette segmentation ne peut pas être exécutée correctement par un appareil de ROC puisque les formules sont trop près du texte.

Exemple n° 10. Page comportant des caractères en indice trop petits

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;

R₂ is selected from alkoxy, cycloalkoxy, phenyl, 4- to 7-membered heterocycles, -O(CH₂)_nphenyl, -O(CH₂)_npyridyl, -E-(CR₂R₃)_n-Q, and Q, each of which is substituted with between 0 and 3 substituents selected from halogen, cyano, hydroxy, oxo, (CR₃R₄)-T, C₁-C₆alkyl, C₁-C₆alkoxy, C₁-C₆haloalkyl, C₁-C₆haloalkoxy, mono- and di-(C₁-C₆alkyl)amino, (C₁-C₆alkyl)((CR_AR_B)_j-T)amino, benzyl, S(O)_n(C₁-C₆alkyl), α,ω-C₁-C₄alkylene, α,ω-C₁-C₄alkyleneoxy, α,ω-C₁-C₄alkylenedioxy, -E-(CH₂)_n-Q, and Q;

T is CO₂H, CONH₂, C₁-C₆alkoxycarbonyl, mono- or di-(C₁-C₆alkyl)aminocarbonyl, SO₃H, SO₂NH₂ or SO₂(C₁-C₆alkyl);

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, NR₂, or a single covalent bond;

R₃ and R₄ are independently chosen from hydrogen, halogen, hydroxy, C₁-C₆alkyl, C₁-C₆alkenyl, (C₃-C₇cycloalkyl)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 R_x.

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-C₆alkyl, C₁-C₆alkenyl, C₁-C₆alkynyl, C₁-C₆haloalkyl, C₁-C₆aminoalkyl, C₁-C₆hydroxyalkyl, C₁-C₆carboxyalkyl, C₁-C₆alkoxy, C₁-C₆haloalkoxy, C₁-C₆alkylthio, C₁-C₆alkanoyl, C₁-C₆alkanoyloxy, C₃-C₇alkanone, C₁-C₆alkyl ether, mono- or di-(C₁-C₆alkyl)amino, C₀-C₄alkyl, -NHC(=O)(C₁-C₆alkyl), -N(C₁-C₆alkyl)C(=O)(C₁-C₆alkyl), -NHS(O)_n(C₁-C₆alkyl), -(C₁-C₆alkyl)C(=O)NH₂, -(C₁-C₆alkyl)C(=O)NH(C₁-C₆alkyl), -(C₁-C₆alkyl)C(=O)NH(C₁-C₆alkyl)(C₁-C₆alkyl), -S(O)_n(C₁-C₆alkyl), -S(O)_nNH(C₁-C₆alkyl), -S(O)_nN(C₁-C₆alkyl)(C₁-C₆alkyl) and Z; or

Il s'agit là d'un exemple typique de caractères en indice trop petits pour permettre une reconnaissance précise. Cela se produit fréquemment dans les demandes de brevets relevant du domaine de la chimie.

Exemple n° 11. Page comportant des tableaux mal formatés

WO 2005/063765

PCT/US2004/043492

Table D

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

Left-hand substituent	Left-hand ring	Aryl or heteroaryl	Ring substituent	Nitrogen feature	Right-hand substituent
CH3	morpholino	aryl	OCH2	NHM	alkyl
isopropyl	piperazine	thiophene	OCH2(CO)	NMM	alkoxy
CH3CH2O(CO)CH2			SO2	morpholino	alcohol
none			OCH2(CO)OCH2	piperazine	substituted amine
				piperidine	acid
				pyrazole	ester
				pyrrolidina	CH ₂ CH ₂ OCH ₃
					CH ₂ CH ₂ OH
					CH ₂ NH ₂
					CH ₂ NHCH ₂ CH ₂ CH ₃
					CH ₂ NHCH ₃
					CH ₂ NHCHCH ₃ CH ₃
					CH ₃
					CHCH 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.

Left-hand substituent	Left-hand ring	Aryl or heteroaryl	Ring substituent	Nitrogen feature	Right-hand substituent
CH3	morpholino	aryl	CH2	NHM	alkyl
isopropyl	piperazine	thiophene	CH2CH2	NMM	alkoxy
CH3CH2O(CO)CH2			CH2CH2CH2	morpholino	alcohol
none			CH2CH2CH2CH2	piperazine	substituted amine
				piperidine	acid
				pyrazole	ester
				pyrrolidina	CH ₂ CH ₂ OCH ₃
					CH ₂ CH ₂ OH
					CH ₂ NH ₂
					CH ₂ NHCH ₂ CH ₂ CH ₃
					CH ₂ NHCH ₃
					CH ₂ NHCHCH ₃ CH ₃
					CH ₃

Dans le présent exemple, le cadre du tableau manque (contrairement à ce que prévoit le paragraphe 33). Par conséquent, l'appareil de ROC essaiera de reconnaître le contenu des tableaux comme le texte d'un paragraphe. Dès lors, plusieurs autres problèmes se poseront :

- la taille de la police utilisée pour les caractères dans les tableaux sera trop petite (voir les paragraphes 34 et 35);
- la ligne de base du titre des colonnes ne sera pas la même partout (voir le paragraphe 29). Par conséquent, l'appareil va, à tort, détecter des indices ou des exposants;
- le flux de texte obtenu ne tiendra pas compte des colonnes suivantes :

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

CH3

Exemple n° 12. Page avec justifications

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

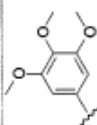
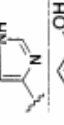
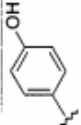
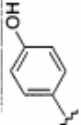
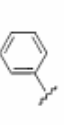
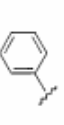
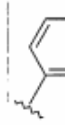
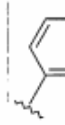
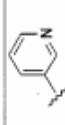
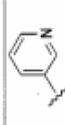


Dans le présent exemple, les paragraphes ont fait l'objet d'une justification à gauche et à droite. Si cela rend le texte plus agréable à regarder, les opérations de ROC peuvent parfois être plus difficiles parce que l'espace entre les mots est trop petit (voir le paragraphe 30). L'exemple n'est pas conforme au paragraphe 31 qui prévoit qu'il faut éviter, dans la mesure du possible, de couper un mot à la fin d'une ligne (l'appareil de ROC a parfois du mal à distinguer un trait d'union impératif d'un trait d'union facultatif, ce qui, en définitive, donne des mots contenant des traits d'union intempestifs).

Exemple n° 13. Tableau dont le cadre est imparfait

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, HC(1:3); H ₂ O(1:1)
108	B2	cb	C=O			2R-trans HC(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

Dans le présent exemple, le cadre du tableau de l'original reçu, avant numérisation, est de mauvaise qualité. Après numérisation, la procédure de ROC ne détecte pas correctement le tableau et une opération manuelle est nécessaire pour segmenter la page. Si une telle page n'est pas vérifiée par un opérateur, du point de vue de la qualité, le texte final contiendra des caractères intempestifs qui rendront moins efficace l'indexation du document par les moteurs de recherche.

SCIT/SDWG/10/3
Annexe, page 21
Appendice 2, page 16

Exemple n° 14. Caractères en indice et caractères en exposant incorrects

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, SR^a, and NR^aR^a, in which R^a and R^a 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, tetrahydrofuranyl,
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, SR^a, and NR^aR^a, in which R^a and R^a are
as defined above. Any nitrogen atom within such a heterocyclic ring

Le présent exemple contient les problèmes suivants (voir le paragraphe 35) :

- les caractères en indice ou en exposant sont trop petits;
- les caractères en indice sont situés trop au-dessous de la ligne de base;
- les caractères en exposant sont situés trop au-dessus de la ligne de base.

Par conséquent, les lignes 34 et 35 du texte sont reconnues par la procédure ROC comme suit :

"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"

Exemple n° 15. Exemple de demande comportant des caractères inhabituels

WO 2006/057705

PCT/

c = speed of sound in water;

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

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

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

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

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

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

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

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

Le présent exemple contient les problèmes suivants :

- caractères inhabituels : caractères grecs en italique et caractères avec un tilde;
- là encore, la taille des indices est trop petite.

La plupart des appareils de ROC ne reconnaîtront pas correctement les caractères inhabituels.

Exemple n° 16. Exemple de caractères étroits, avec un espacement restreint

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.

Le présent exemple n'est pas conforme aux paragraphes 37 et 38. Par conséquent, l'appareil de ROC ne distingue pas correctement les limites de mot et le résultat est un texte totalement inutilisable.

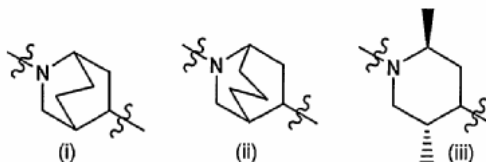
Exemple n° 17. Exemple de tampon apposé au mauvais endroit par l'office récepteur, avant la numérisation

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:



Le présent exemple n'est pas conforme au paragraphe 47. Par conséquent, les six premiers mots du texte de la page ne peuvent pas être lus par une technique de ROC. En outre, le tampon introduit des caractères supplémentaires caducs qui encombrant les moteurs d'indexation lorsque la qualité de la page n'a pas été vérifiée par un opérateur.

Exemple n° 18. Autre page comportant des formules mathématiques mal présentées

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,

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

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

5

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

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

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

10 and,

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

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

In the noiseless case,

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

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

There exists some functional relationship such that,

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

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

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

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

25

THE FREQUENCY INTERPOLATION FUNCTION

La présente page n'étant pas conforme à beaucoup de recommandations, le résultat de la ROC n'est pas utilisable :

- les formules mathématiques intégrées ne sont pas séparées des paragraphes de texte (voir le paragraphe 23);
- des caractères inhabituels sont présents dans les paragraphes de texte (voir le paragraphe 40);
- des caractères italiques sont associés à des caractères grecs (voir le paragraphe 39)

Aux fins de la présentation de cette page, il aurait fallu utiliser des espaces supplémentaires pour séparer les formules intégrées des paragraphes. Les lettres grecques n'auraient pas dû être mises en italique dans les formules, ni dans les paragraphes. Il convient d'éviter, dans la mesure du possible, d'utiliser des accents circonflexes (^) pour représenter des variables dans des paragraphes de texte, qui auraient pu être remplacés par des exposants; ainsi, "epsilon circonflexe" aurait pu être représenté comme suit : ε^\wedge ou ε^{hat} .

Exemple n° 19. Page comportant des caractères italiques soulignés non recommandés

WO 2006/038001

PCT/GB2005/003827

- 132 -

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

LCMS 399/401 [M+H]⁺, RT 1.88 min.

EXAMPLE 320

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

LCMS 413/415 [M+H]⁺, RT 3.20 min.

EXAMPLE 321

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

LCMS 413/415 [M+H]⁺, RT 3.19 min.

EXAMPLE 322

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

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

Il s'agit là d'un exemple fréquent de page posant un problème de ROC aux fins de la publication PCT. Cette page n'est pas conforme aux recommandations suivantes :

- Paragraphe 41 : le texte ne devrait pas être souligné. Le soulignement est déconseillé précisément pour les formules chimiques (les dictionnaires ne sont d'aucune aide en l'occurrence). Cela est problématique notamment pour tous les caractères touchant le soulignement :]y p... ne sont pas reconnus correctement.
- Paragraphe 39 : les caractères italiques ne sont pas recommandés. Ils sont notamment déconseillés aux fins du changement du type de police au sein d'un même mot (les appareils de ROC considèrent souvent que tous les caractères d'un même mot ont le même type de police). Par conséquent, tous les "1*H*" et "*N*-" sont reconnus de manière erronée.

Exemple n° 20. Page complètement illisible

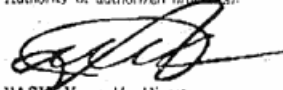
WO 2005/071074

PCT/KR2005/000214

BURBANK 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 : Mogen Biotechnology Research Institute
#391, Hujung-ro, Kousung-eup, Yongin-city, Kyonggi-do 449-910,
Republic of Korea

I. IDENTIFICATION OF THE MICROORGANISM	
Identification reference given by the DEPOSITOR: Saccharomyces cerevisiae WJ3501 / M61.K8 #36	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: KCTC J0582HP
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 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

Form IBS-IBTC Form 01

La présente page devrait être refusée par les offices; elle a été envoyée par télécopie avec une résolution de 100 ppp et ne peut même pas être lue par un être humain. Face à une telle situation, les opérateurs considèrent que l'ensemble du contenu de la page est une image puisque le texte ne peut pas en être extrait.

[Fin de l'appendice 2 et de la norme]

[Fin de l'appendice 2 et du document]