



СТАНДАРТ СТ.22 РЕКОМЕНДАЦИИ ПО ОФОРМЛЕНИЮ ЗАЯВОК НА ПАТЕНТЫ, НАПРАВЛЕННЫЕ НА ОБЛЕГЧЕНИЕ ОПТИЧЕСКОГО РАСПОЗНАВАНИЯ СИМВОЛОВ (ОРС)

Переработанные и одобренные Рабочей группой по стандартам и документации
ПКИТ на десятой сессии 21 ноября 2008г.

ВВЕДЕНИЕ

1. Данные Рекомендации относятся к заявкам на патенты, поданным или на бумаге, или в электронной форме (электронная подача), если текстовая часть заявки представлена в графическом формате (например, PDF или TIFF).
2. Данные Рекомендации были разработаны для того, чтобы помочь при подготовке заявки на патент в печатной форме, пригодной для последующего получения электронно-цифровой записи содержания заявки на патент с помощью оборудования для Оптического Распознавания Символов (ОРС или, на английском языке, OCR).
3. Данные Рекомендации были разработаны на основе опыта разных ведомств по использованию оборудования для ОРС. Они составлены с целью достижения наименьшего возможного процента ошибок на шаге автоматического чтения текста заявок на патент, и в то же время допускают персональное чтение документа при необходимости. Следует иметь в виду, что этот документ не дает детальных рекомендаций для японского и корейского языков, поскольку в этих странах процент заявок, полностью подаваемых в электронном виде, по отношению к общему числу поданных в год заявок, на самом деле превышает 90%, и, следовательно, для японского и корейского языков этот Стандарт практически не будет применяться.
4. Главная цель получения цифровой записи заявки на патент - возможность облегчения публикации такой заявки в печатном формате с использованием техники компьютерного набора, и, следовательно, такое преимущество для всех пользователей, как улучшение внешнего вида и качества патентных документов. Последующей целью является создание машиночитаемых баз данных полных текстов опубликованных документов для использования впоследствии преимуществ, предоставляемых полнотекстовым компьютерным поиском.

ОПРЕДЕЛЕНИЯ

5. Для целей данных Рекомендаций выражение «заявка на патент» означает заявки на: патент на изобретение, авторское свидетельство на изобретение, свидетельство о полезности, полезную модель, дополнительный патент или свидетельство, дополнительное авторское свидетельство на изобретение и дополнительное свидетельство о полезности.



6. Математическая или химическая формула называется «сложной», если она не может быть представлена линейной последовательностью знаков, и знаки которой могут иметь либо подстрочный, либо надстрочный индекс. Формула является очень сложной, если она содержит составные подстрочные или надстрочные индексы или если она содержит математические символы суммы, интеграла или произведения.

7. Ограничивающая рамка знака/ряда знаков - это наименьший привязанный к осям координат прямоугольник, который включает все части знака/ряда знаков.

8. Термин "курсив" относится к стилизованной форме рукописного почерка, когда буквы в словах соединены, изображая каждое слово одной сложной линией. Шрифты называются курсивными, если они имитируют рукописный стиль.

СОЗДАНИЕ ОРИГИНАЛА

9. Заявка на патент часто будет подготавливаться с использованием технических средств обработки текста. Как показал опыт, самым эффективным форматом из используемых, позволяющим надежно применять технику ОРС, является так называемый формат OCR-B, определенный в стандарте Международной организации по стандартизации (ИСО) 1073/II.

БУМАЖНЫЙ НОСИТЕЛЬ В СЛУЧАЕ ПОДАЧИ ЗАЯВКИ НА БУМАГЕ

10. Для облегчения сканирования бумажный носитель отпечатанной заявки должен иметь следующие характеристики:

- (a) Бумага должна быть прочной, белой и чистой.
- (b) Вес бумаги должен быть в интервале от 70, предпочтительно 80, до 120 г/м².
- (c) Размер бумаги должен быть предпочтительно A4, а именно 210 мм x 297 мм или 8 1/2 на 11 дюймов (де-факто Северо-Американский стандарт).
- (d) Листы не должны иметь заломов, дыр и/или быть скручены.
- (e) Бумага не должна быть впитывающей, чтобы избежать расплывания чернил (например, при использовании струйного принтера).

РЕКОМЕНДАЦИИ ПО РАЗМЕТКЕ СТРАНИЦЫ

11. Следует избегать двусторонней печати. Если это невозможно, должна использоваться бумага, достаточно непрозрачная для того, чтобы обеспечить хорошее качество сканирования и с одной, и с другой стороны листа.

12. Знаки должны быть целиком (однотонно) черными на белом фоне.

13. На верхней, нижней и правой сторонах каждого листа должно быть оставлено поле минимум 2 см, на левой стороне листа минимальное поле должно быть 2,5 см.



Любая справочная информация заявителя или его представителя должна приводиться в верхнем поле. Разметка страницы приведена в Приложении 1.

14. Следует избегать нумерации строк. Если требуется, номера строк должны печататься арабскими цифрами в левом поле страницы, размеченной, как показано в Приложении 1, на расстоянии минимум 1 см от границы текста. Размер шрифта номеров строк должен быть, по меньшей мере, 12 кегля / пунктов.

15. Нумерация страниц должна указываться арабскими цифрами без каких-либо других разграничивающих знаков. Номера страниц предпочтительно проставлять в центре верхнего или нижнего поля страницы, как это показано в Приложении 1.

16. Описание, формула и реферат должны быть напечатаны, начиная каждое с нового листа. Более того, первое слово, напечатанное на первом листе каждой из трех вышеупомянутых частей заявки, должно указывать соответствующую часть (на языке заявки). Пункты формулы должны быть последовательно пронумерованы. Формат нумерации пунктов формулы должен позволять чётко отделять номер пункта формулы от его текста. Рекомендуются форматы должны использовать либо арабские цифры, после которых ставится точка, либо слова «пункт формулы» - или их эквивалент на языке заявки на патент, - после которых через пробел ставится номер пункта формулы арабскими цифрами. Последующий текст каждого пункта формулы должен отделяться от номера пункта формулы отступом вправо не менее чем на 1 см в обоих случаях.

17. Страницы должны состоять из одного столбца абзацев (текстовые абзацы или абзацы, содержащие встроенные изображения).

18. Страницы, содержащие абзацы, должны иметь вертикальную (книжную) ориентацию.

19. Следует избегать горизонтальной («альбомной») ориентации страниц. Она допускается только для страниц, содержащих встроенные рисунки или таблицы, для которых не подходит вертикальная («книжная») ориентация.

20. На любой странице текст должен иметь только одно направление.

21. Горизонтально ориентированные страницы должны быть повернуты на 90 градусов против часовой стрелки при объединении в один комплект с вертикально ориентированными страницами.

22. Рекомендуется избегать использования сносок, текста и заголовков на полях, за исключением указанного в параграфе 14 (нумерация строк), параграфе 15 (нумерация страниц) и включения отсылок заявителя к данным подачи, проставляемым на полях в левом верхнем углу.



РЕКОМЕНДАЦИИ ПО ФОРМАТИРОВАНИЮ АБЗАЦЕВ

23. Рекомендуется, чтобы таблицы, сложные химические и математические формулы, чертежи и рисунки отделялись от абзацев текста. Желательно, чтобы такие встроенные объекты окружались сверху и снизу пустыми полями, по меньшей мере, в 1 см на всю ширину страницы.

24. Чертежи и рисунки максимально должны быть включены в раздел «Чертежи» а ссылки на них должны содержаться в разделах «Описание» и «Формула» заявки на патент.

25. Чертежи и рисунки должны быть черно-белыми (следует избегать изображений с серыми тонами, т.к. при их сканировании или преобразовании в черно-белые эта информация теряется). Рисунки должны быть сделаны ясными линиями, толщина которых должны быть достаточна для хорошего воспроизведения при разрешении 300 точек на дюйм.

26. Следует избегать рукописного текста в абзацах или примечаниях. Если необходимо, они должны рассматриваться как встроенные изображения и соответствовать рекомендациям, данным в параграфе 23.

27. Строки текста должны печататься через полтора интервала.

28. Абзацы должны разделяться пространством, по меньшей мере вдвое большим, чем расстояние между строками в абзаце.

29. Базисные линии всех знаков внутри строк абзаца должны быть тщательно выровнены, за исключением знаков подстрочных и надстрочных индексов, указанных в параграфе 35.

30. Следует избегать выравнивания текста абзацев по ширине. Если оно всё же применяется, расстояние между словами должно быть, по меньшей мере, такой же ширины, как в не выровненном тексте. Выровненный по ширине текст может мешать системам ОРС правильно идентифицировать границы слова в тексте.

31. Насколько возможно, следует избегать разделения слов с помощью тире (например, в конце строк или ячеек таблицы). Это не относится к языкам, которые используют составные существительные (например, к немецкому языку).

РЕКОМЕНДАЦИИ ПО ТАБЛИЦАМ

32. Должен использоваться только белый фон.

33. Таблицы должны иметь границы (рамки). Границы должны быть толще 1.5 пунктов и выполнены только сплошными линиями.



РЕКОМЕНДАЦИИ ПО ШРИФТАМ

34. Минимальный рекомендуемый размер шрифта должен быть 12 пунктов, предпочтительно – 14. Как общая рекомендация – все знаки текста должны иметь один и тот же размер шрифта.

35. Текстовые абзацы, содержащие подстрочные и надстрочные индексы, должны использовать, по меньшей мере, 12-й размер шрифта (рекомендуется 14 – чем больше, тем лучше). Следует гарантировать, чтобы ограничивающая рамка подстрочных или надстрочных знаков достаточно перекрывала ограничивающую рамку нормальных знаков на той же строке (это предотвращает при процедуре ОРС печать подстрочных/надстрочных знаков на других строках).

36. Рекомендуется использовать следующие шрифты (в приведенном порядке):

(a) Семейство шрифтов с одинаковой шириной знаков (Monospaced): OCR-B, Courier New, Free Mono.

(b) Семейство шрифтов с засечками (Serif): ITC Officina Serif, Times New Roman, Free Times.

(c) Семейство шрифтов без засечек (Sans Serif): Verdana, ITC Officina Sans, Arial, Helvetica, DejaVu Sans.

Однако шрифты Arial, Helvetica, DejaVu Sans, Free Times and Times New Roman не рекомендуются для заявок, содержащих химические и/или математические формулы, а также акронимы (слова, сформированные из первых букв словосочетаний, которые они заменяют), в которых смешаны буквы и цифры. Для китайских знаков рекомендуется шрифт Song.

37. Знаки шрифтов должны иметь четкую форму, без теней. Промежутки между знаками должны быть достаточно широкими (следует избегать узкой разрядки).

38. Не должны использоваться узкие и рукописные шрифты.

39. Насколько возможно следует избегать полужирных и наклонных шрифтов.

40. Насколько возможно следует избегать необычных (нестандартных/ нетипичных/ неправильных) знаков. Если необходимо, они должны быть стандартными буквами Греческого алфавита и символами шрифта (в порядке их предпочтения). Не должны использоваться знаки, которых нет среди UNICODE: такие знаки распознаются средствами ОРС как встроенные изображения и поэтому делают распознаваемый текст трудным для чтения. Каждое ведомство должно определить и опубликовать свои требования для тех знаков, которые могут использоваться при подготовке заявок на патенты.

41. Текст не должен подчеркиваться. Если это необходимо, следует обеспечить, чтобы подчеркивание не перекрывало ограничивающие рамки знаков.



РЕКОМЕНДАЦИИ ПО КОМБИНИРОВАНИЮ ЯЗЫКОВ

42. Смешивание азиатских (т.е. основанных на идеографическом письме) и европейских языков (т.е. основанных на латинском и кириллическом алфавитах) внутри разделов/ страниц заявок на патенты является проблемой для процедуры ОПС и его следует избегать, за исключением случаев, когда это необходимо.

РЕКОМЕНДАЦИИ ПО СКАНИРОВАНИЮ

43. Заявки на патент должны сканироваться или в черно-белой шкале, или в шкале уровней серого цвета.

44. Предпочтительным разрешением для сканирования является 300 точек на дюйм. Сканирование с разрешением ниже 300 точек на дюйм, даже в шкале уровней серого цвета, может привести к плохому качеству публикуемых ведомствами документов, т.к. обмен документами между ведомствами и процессы публикации часто включают преобразования к черно-белым страницам с разрешением 300 точек на дюйм в формате TIFF группы IV.

45. Отсканированные документы должны быть преобразованы либо в PDF формат, либо в TIFF формат.

ИСПРАВЛЕНИЯ

46. Исправления в тексте заявки должны делаться перепечаткой целой страницы. Не допускаются знаки подтверждения исправления – такие, как, например, указаны в Международном Стандарте ISO 5776. Не допускается делать исправления с помощью белой корректирующей жидкости, самоклеющихся полосок бумаги, стирания или зачеркивания. Заменяющие листы не должны посылаться в ведомство по факсу, использующему стандартное разрешение факсов: листы должны либо посылаться физически, либо по факсу, использующему разрешение не менее 300 точек на дюйм, либо любыми средствами сетевой передачи, поддерживаемыми ведомством, при условии, что каждый лист отсканирован с разрешением не менее 300 точек на дюйм.

РЕКОМЕНДАЦИИ ДЛЯ ВЕДОМСТВ

47. Патентные ведомства должны избегать изменения полученных листов до передачи их на сканирование и процедуру ОПС. Например, иногда в практику ведомств включена операция штампования полученных листов, что может загромождать символы листа, делая представленный заявителем текст нечитаемым при процедуре ОПС. Если печати/ изменения должны быть использованы на оригинальных листах, то ведомству необходимо принять меры к тому, чтобы такие изменения вносились только на поля документов, в соответствии с Приложением 1.

48. В будущем патентные ведомства должны избегать разработки бумажных бланков для взаимодействия между заявителями и ведомством. Как показывает имеющийся опыт, разработка и внедрение безопасных онлайн систем предпочтительнее, чем



построение систем распознавания бланков на бумаге. Тем не менее, для разработанных на бумаге бланков даются следующие рекомендации с точки зрения облегчения их распознавания:

- на бланках не должны использоваться строки из маленьких точек для указания пользователю, куда следует вводить текст.
- Для знакомест должны использоваться цвета, легко убираемые при сканировании (Drop-out colors) (светло - серый).
- Цвета, легко убираемые при сканировании (Drop-out colors), не должны использоваться для участков формы (полей бланка), которые, как предполагается, будут заполняться содержанием разных цветов или оттенков серого, таких, как отсканированная подпись или рисунок.

ВНЕДРЕНИЕ РЕКОМЕНДАЦИЙ

49. Рекомендуются, чтобы Ведомства, которые намерены начать прием или потребовать подачи заявок на патент, отпечатанных в формате, пригодном для ОРС, регулярно публиковали в своих Официальных Бюллетенях и на Веб-сайтах полное руководство для этого, точно определяя в нем разрешенные шрифты и допустимый размер бумаги.

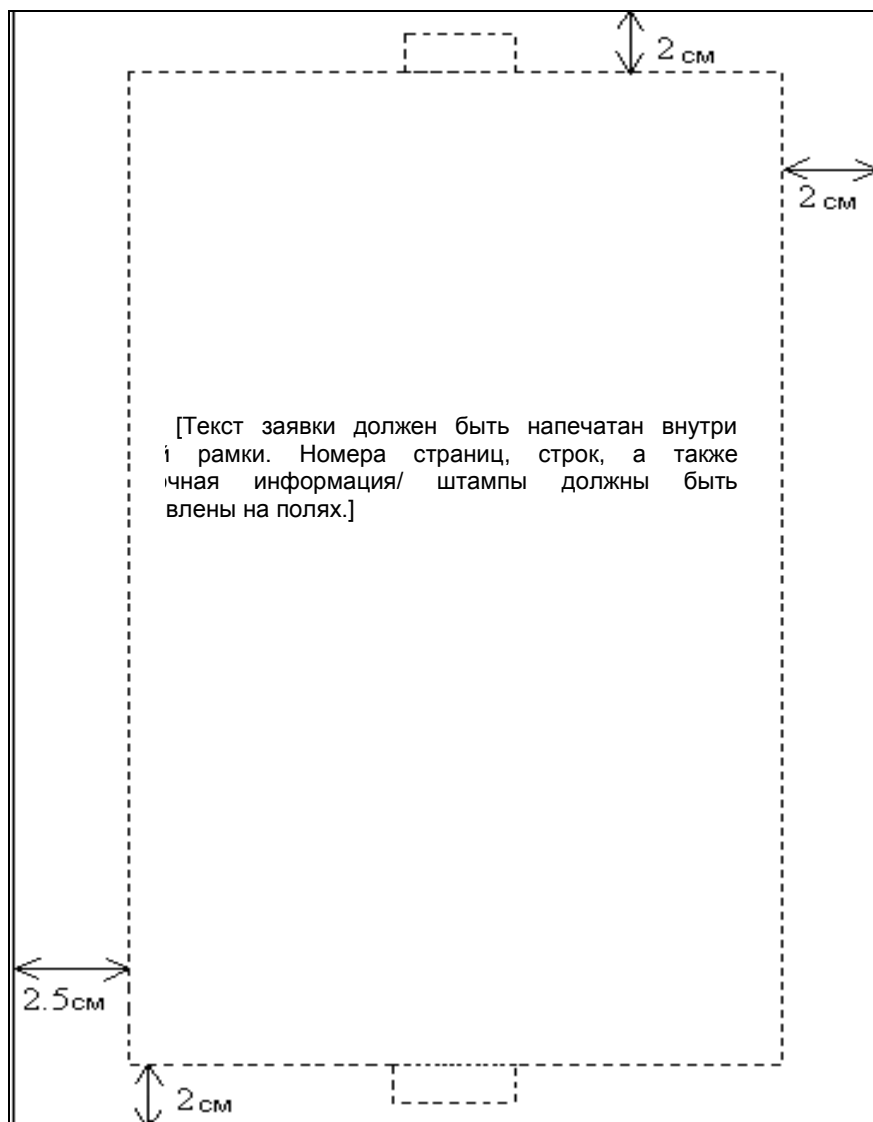
Примеры

50. Примеры хороших и плохих для ОРС страниц заявок приведены в Приложении 2 к данным Рекомендациям. Приведенные примеры показывают, что следует, а чего не следует делать, и сопровождаются краткими пояснениями.

[Приложения следуют]



ПРИЛОЖЕНИЕ 1



Размер оригинала = A4



ПРИЛОЖЕНИЕ 2

ПРИМЕРЫ ХОРОШИХ И ПЛОХИХ ОБРАЗЦОВ ПРЕДСТАВЛЕНИЯ

В этом Приложении Вы найдете примеры хорошего и плохого представления страниц патентных документов с точки зрения точности их воспроизведения, полученного после выполнения процедур OPC (OCR).

Примеры хороших образцов

Пример 1: страница описания хорошего качества

WO 2006/111319

PCT/EP2006/003401

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

5

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

10

For highest possible precision of the optical image to be obtained in complicated optical instruments such as a projection exposure system, the influence of the polarization of the light must be considered or, respectively, the polarization must be influenced specifically. For example, in particular
15 in case of great incidence angles, polarization effects occur in the mirror systems, which projection exposure systems in the EUV region are based on, for lack of suitable transparent materials. These polarization effects are in particular due to the varying reflectivity of the mirrors for s-polarized and p-polarized light and can give rise to imaging errors or other undesired
20 effects. Efforts have been made to measure possible polarization effects in the individual components of projection exposure systems.

25

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



Приложение 2

Пример 2: страница с формулой хорошего качества

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.

Соответствует всем рекомендациям: поля страницы, стандартный шрифт (Times New Roman), подходящий размер шрифта, без номеров строк, ограниченное использование полужирного шрифта, нет курсива, нет подчеркнутого текста, нумерация пунктов формулы соответствует требованиям и чётко отделена от текста формулы.



Приложение 2

Пример 3: страница со сложным описанием хорошего качества

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:

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

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

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

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

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:

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

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.

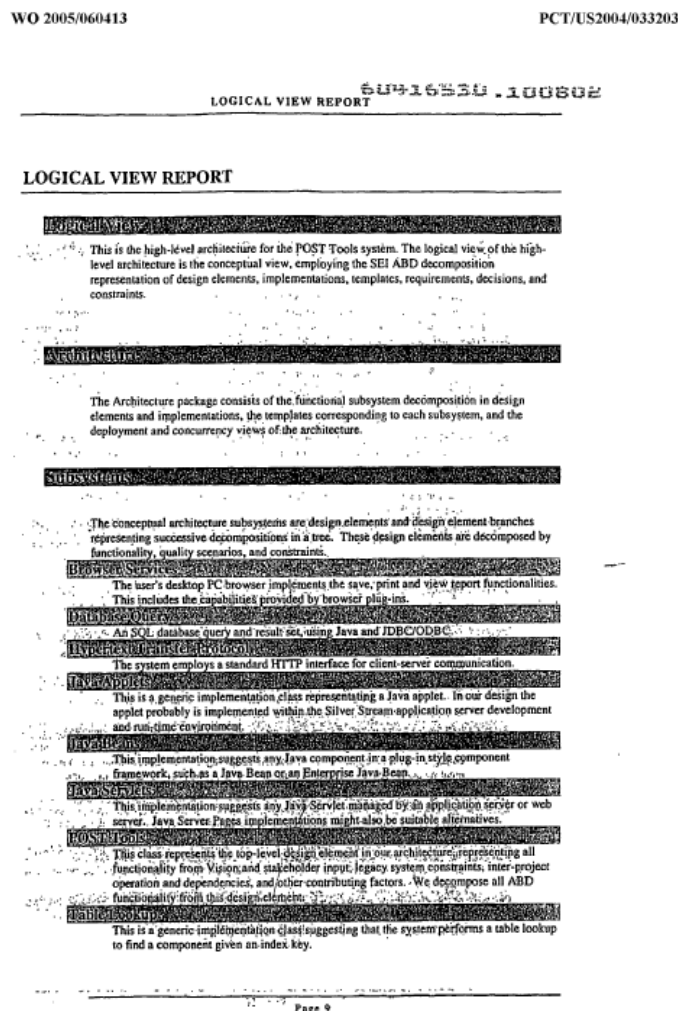
Абзацы текста не содержат нетипичных математических знаков. Математические формулы вставлены надлежащим образом, позволяя оборудованию для ОРС легко выделять встроенные изображения. Возможный результат их выделения (сегментации) показан голубой линией.



Приложение 2

Примеры плохих образцов

Пример 1: страница плохого качества со множеством дефектов



Данный пример не соответствует рекомендациям параграфа 10 (возможно, страница была передана в Ведомство по факсу с разрешением 200 точек на дюйм – видны помехи – и текст возникает на интенсивном сером фоне). Этот пример также не соответствует рекомендациям параграфов 13 и 47. Присвоенный номер (604115530.100802) проштампован внутри страницы (он должен быть на полях). Нумерация страницы - неверная (должно быть "9", а не "страница 9", см. параграф 34). Наконец, размер шрифта слишком мал (параграф 15). В идеале такие страницы не должны приниматься Ведомством и им должны быть затребованы заменяющие страницы (эта страница не может быть правильно преобразована при OCR).



Приложение 2

Пример 2: страница с небелым фоном

WO 2005/097403

- 13 -

PCT/FR2005/050194

REVENDICATIONS

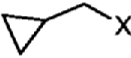
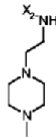
1. Dispositif d'usinage (D) du type de celui associant une machine-outil d'usinage (100) à un dispositif porte-pièce (200) équipé d'un axe de mise en mouvement de rotation transversal (A) par rapport à l'axe de plongée (Z), CARACTÉRISÉ PAR LE FAIT QUE le dispositif porte-pièce (200) est constitué par 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

Данный пример не согласуется с параграфом 12. Для этой страницы необходима процедура фильтрации, чтобы попытаться удалить фоновые помехи перед передачей её на процедуру OPC. Если произвести OPC данной страницы, как она есть, то полученный текст будет нечитаемым.



Приложение 2

Пример 3: страница с блеклыми знаками

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

Beispiele 26-40

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

5 beschrieben. Das für die Darstellung des Amids ei

Небольшая область данной страницы увеличена, чтобы показать знаки: вероятно, цвет оригинального текста – серый, получающийся после черно-белого сканирования знаков, имеющих не сплошные линии, с разрешением 300 точек на дюйм. В результате качество преобразованного OPC текста получается низким (данный пример не согласуется с параграфом 12).

Пример 4: страница с рукописным 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

По всей вероятности, текст, полученный после OPC данной страницы, будет нечитаемым. Для гарантии минимального качества публикации Ведомство должно требовать машинописный текст.



Приложение 2

Пример 5: страница с nereкомендуемым расположением текста и другими недостатками

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.

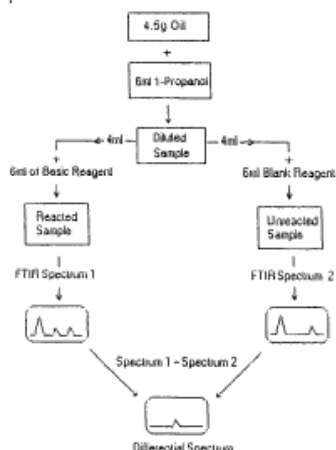


Figure 2. Analytical protocol for the determination of AN by FTIR spectroscopy.

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.

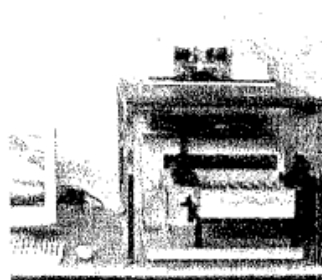


Figure 3. The COAT AN/BN Analyzer and its key components

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

Данная страница не соответствует следующим рекомендациям: параграф 17 (форматирование в один столбец), параграф 39 (используются наклонные и полужирные шрифты), параграф 46 (имеет исправления, выполненные от руки после печати). Выравнивание текста по ширине (и слева, и справа) также не рекомендуется (параграф 30), хотя в данном случае это не привело бы к отрицательным последствиям после ОПС, поскольку слова всё же достаточно разделены белыми промежутками. И, наконец, данный пример не согласуется с параграфом 27 (должен быть полуторный интервал между строками).



Приложение 2

Пример 6: страница со слишком маленькими номерами строк

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);

[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);

[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);

[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);

[0032] Fig. 12 is a plot showing the percent increase in cytotoxicity (as determined by $(IC_{50, \text{no cysteine}}/IC_{50, \text{cysteine}}) \times 100$) of free mitomycin C (closed squares), mitomycin C associated with liposomes comprised of HSPC/cholesterol/mPEG-DSPE/lipid-DTB-mitomycin C (closed circles), and liposomes comprised of HSPC/mPEG-DSPE/lipid-DTB-mitomycin C (open triangles) to M109 cells *in vitro* at various concentrations of cysteine;

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

Номера строк вызывают проблемы для оборудования OPC по нескольким причинам (параграф 14):

- Они могут быть не на одном уровне со строками, которым соответствуют, что приводит к неправильному определению пронумерованной строки;
- Они могут быть слишком малы, что приведет к ошибкам распознавания, и, для защиты XML процедур извлечения, корректно будут удалены из данной страницы;
- Они могут быть ошибочно помещены внутрь текстовой зоны страницы или на полях, но слишком близко к тексту, и, как результат, возникать внутри текстового потока, создаваемого при операциях OPC.

В данном примере они слишком малы.

В данном примере также слишком малы подстрочные символы (параграф 35).



Приложение 2

Пример 7: страница, содержащая несколько направлений текста

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	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

Данный пример не соответствует рекомендациям параграфа 20.

Одно из ограничений лучших из существующих сейчас видов оборудования для ОРС состоит в том, что они могут читать только одно направление текста на странице (предварительная обработка страницы заключается в определении главной ориентации текста страницы). В результате все слова, не совпадающие с этим главным направлением, игнорируются. Несомненно, допустимо иметь на странице «альбомно» ориентированные таблицы или даже основной «альбомно» ориентированный текст с «книжным» (вертикальным) расположением примечаний на полях (номер страницы, номер заявки и т.д.).



Приложение 2

Пример 8: страница с комбинацией встроенных математических формул и текста

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 $\Phi_0 = \frac{\hbar}{2e}$,

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,

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

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

- 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 $\left[\frac{g}{2} \right]$ is 1.001 159 652 137. The experimental value [23] of $\left[\frac{g}{2} \right]$ 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 $n=1$, $\ell=0$

Данный пример не согласуется с параграфом 23 **Ошибка! Источник ссылки не найден.** Оборудование для ОРС не может правильно разделить текст и формулы (посмотрите результат ручного выделения (сегментации) формул красной линией: встроенные формулы даже пересекаются).

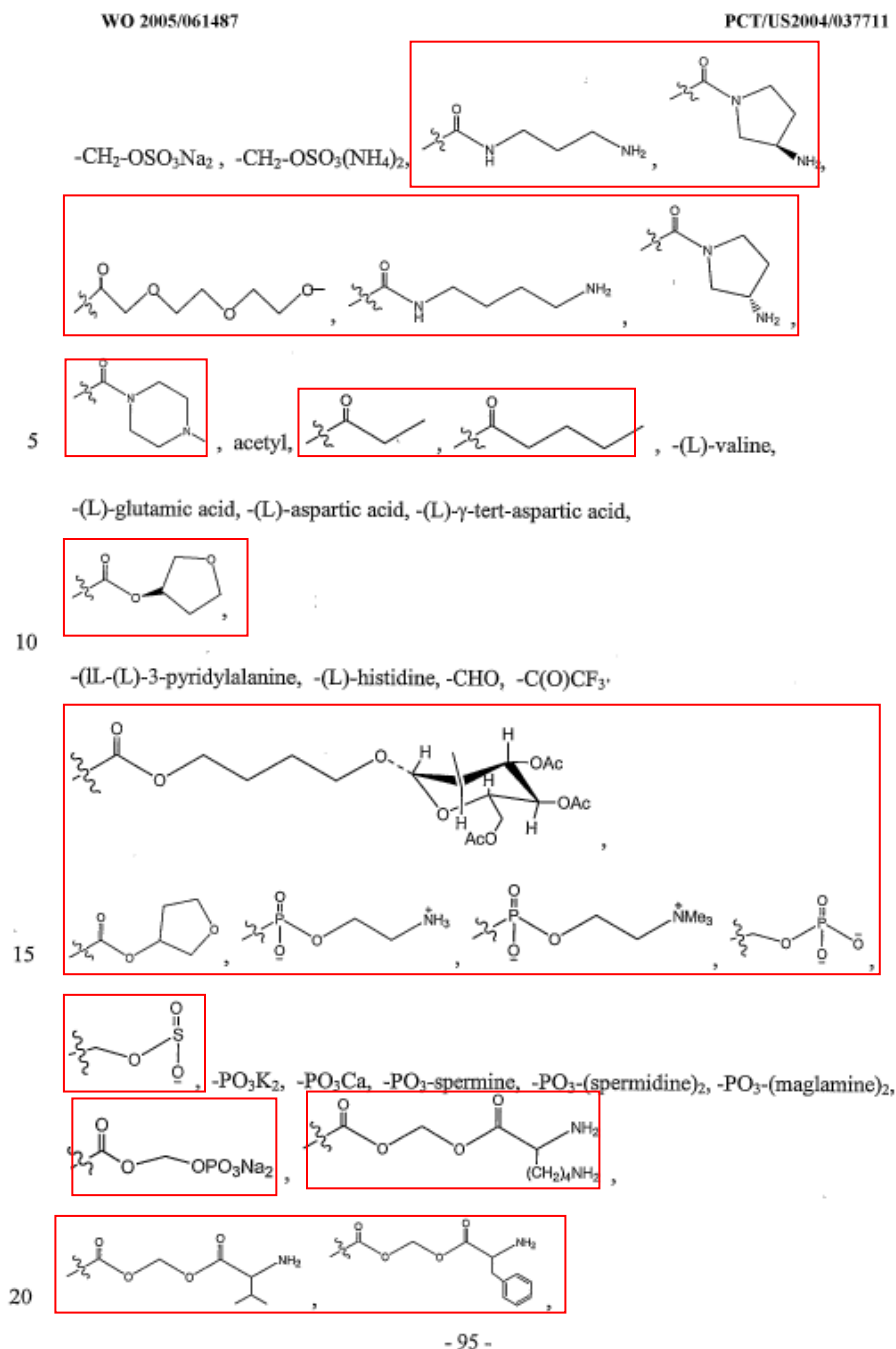
Как общий комментарий, - в данном примере текст и формулы слишком плотно расположены для хорошего распознавания; данный пример также не соответствует рекомендациям параграфов 27 and 28.

Кроме того, в данном примере используются нетипичные знаки: Греческие символы могут использоваться, даже если они увеличивают трудности распознавания на странице (см. параграф 40). Однако категорически не рекомендуется комбинировать наклонные, полужирные и подчеркнутые шрифты с нетипичными знаками (параграф 39).



Приложение 2

Пример 9: страница с комбинированием встроенных химических формул и текста



Данный пример не соответствует рекомендациям параграфа 23. В красных рамках Вы можете увидеть ожидаемый Вами результат сегментации рисунков (сделанной вручную). Но такая сегментация не может быть правильно выполнена оборудованием для ОРС, поскольку формулы слишком близко расположены к окружающему тексту.



Приложение 2

Пример 10: страница со слишком маленькими подстрочными символами

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, $(C_2$ - C_7 cycloalkyl) C_0 - 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, $-E-(CR_4R_5)_m-Q$, and Q , each of which is substituted with between 0 and 3 substituents selected from halogen, cyano, hydroxy, oxo, $(CR_4R_5)_2-T$, C_1 - C_4 alkyl, C_1 - C_4 alkoxy, C_1 - C_6 haloalkyl, C_1 - C_6 haloalkoxy, mono- and di- $(C_1$ - C_4 alkyl)amino, $(C_1$ - C_4 alkyl) $((CR_4R_5)_2-T)$ amino, benzyl, $S(O)_n(C_1$ - C_4 alkyl), α,ω - C_{1-4} alkylene, α,ω - C_{1-4} alkyleneoxy, α,ω - C_{1-4} alkylenedioxy, $-E-(CH_2)_n-Q$, and Q ;

T is CO_2H , $CONH_2$, C_1 - C_4 alkoxycarbonyl, mono- or di- $(C_1$ - C_4 alkyl)aminocarbonyl, SO_2H , 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_4 and R_5 are independently chosen from hydrogen, halogen, hydroxy, C_1 - C_6 alkyl, C_1 - C_6 alkenyl, $(C_2$ - C_7 cycloalkyl) C_0 - 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_{21} .

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_4 alkyl, C_1 - C_4 alkenyl, C_1 - C_4 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_2 - C_6 alkanone, C_1 - C_4 alkyl ether, mono- or di- $(C_1$ - C_4 alkyl)amino C_0 - C_4 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

Это – типичный пример, когда подстрочные символы слишком малы для точного их распознавания. Такой случай часто встречается в патентах в области химии.



Приложение 2

Пример 11: страница с плохо сформатированной таблицей

WO 2005/063765

PCT/US2004/043492

Table D

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

5

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

В данном примере отсутствуют границы таблицы (не согласуется с параграфом 33). В результате оборудование для ОРС будет пытаться распознать содержание таблицы как абзачный текст. Это ведет к нескольким другим проблемам:

Размер шрифта знаков в таблице слишком мал (параграфы 34 and 35).

Линии строк заголовков столбцов смещены друг относительно друга (параграф 29). В результате оборудование для ОРС будет ошибочно распознавать их как подстрочные и надстрочные символы.

Получаемый при ОРС текстовый поток не будет учитывать столбцы:

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



Приложение 2

Пример 12: страница, выровненная и слева, и справа (по ширине)

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

В данном примере к тексту применено и левое, и правое выравнивание. Хотя это делает текст более эстетично выглядящим, иногда это вызывает трудности при операциях ОРС, когда расстояния между словами становятся слишком малы (параграф 30). Этот пример не согласуется и с параграфом 31, который устанавливает, что разделения слова в конце строки следует избегать насколько возможно (оборудование для ОРС иногда испытывает трудности при различении жестких и мягких переносов, в результате чего на выходе получают нежелательные переносы в словах).



Приложение 2

Пример 13: таблица с плохими рамками

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

В данном примере рамки (границы) таблицы полученного оригинала перед сканированием уже имели плохое качество. После сканирования процедура OPC не в состоянии корректно различить таблицу, и необходима последующая ручная сегментация страницы. Если такая страница не проверяется оператором на качество, выходной текст будет содержать нежелательные «прерывистые» участки, что сделает менее эффективной индексацию документа поисковой машиной.



Приложение 2

Пример 14: неудачные подстрочные и надстрочные символы

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, tetrahydrofuranyl, tetrahydrothiophenyl, piperidinyl, piperazinyl, azepane, azocane, morpholinyl, isochroman 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

В данном примере существуют следующие проблемы (параграф 35):

- подстрочные и надстрочные символы слишком малы
- подстрочные символы расположены слишком низко относительно основной строки
- надстрочные символы расположены слишком высоко относительно основной строки

В результате строки 34 и 35 текста распознаются при процедуре OPC следующим образом:

"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"



Приложение 2

Пример 15: пример с нетипичными знаками

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 , beam pair 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 $p+1$, 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

В данном примере существуют следующие проблемы:

- Нетипичные знаки: наклонные (курсивные) греческие буквы, и даже буквы с тильдой.
- Подстрочные символы здесь снова слишком малы

В большинстве устройств для OPC все нетипичные знаки не будут правильно распознаваться.



Приложение 2

Пример 16: пример с узкими шрифтами и узкой разрядкой

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.

Данный пример не соответствует рекомендациям параграфов 37 и 38. Как следствие, оборудование для ОРС не сможет правильно различить границы слов, и результат ОРС будет совершенно непригоден для использования.



Приложение 2

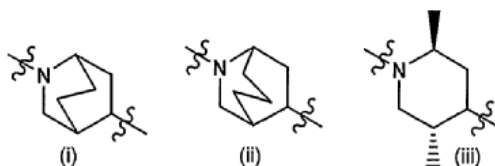
Пример 17: неудачное проставление штампа получающим ведомством перед сканированием

~~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:



Данный пример не согласуется с параграфом 47. В результате первые шесть слов текста на странице не могут быть прочитаны при помощи процедуры OPC. Более того, штамп привносит в знаки очень сильные искажения, что испортит работу индексирующих устройств, если качество этой страницы не будет проверено оператором.



Приложение 2

Пример 18: ещё одна страница с плохо расположенными математическими формулами

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,

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

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

In the noiseless case,

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

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

There exists some functional relationship such that,

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

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

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

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

25

THE FREQUENCY INTERPOLATION FUNCTION

Поскольку данная страница не соответствует многим рекомендациям, результат ОПС не пригоден для использования:

- встроенные математические формулы не отделены от текста (параграф 23);
- использованы нетипичные знаки в тексте (параграф 40);
- курсив скомбинирован со знаками греческого алфавита (параграф 39).

Рекомендуемый путь для форматирования данной страницы – использовать дополнительные промежутки для отделения встроенных формул от абзацев текста. Буквы греческого алфавита не должны быть наклонными (курсивными) как в формулах, так и в тексте. Следует, насколько возможно, избегать использования циркумфлекса (диакритического знака ^ над гласной) для указания переменных в тексте: вместо них могут использоваться надстрочные символы: $\varepsilon^{\wedge}$ ("epsilon circumflex") может быть представлена как $\varepsilon^{\text{circumflex}}$.



Приложение 2

Пример 19: страница с наклонными и подчеркнутыми знаками

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.

Это – часто встречающаяся в PCT публикациях проблема OPC. Данная страница не соответствует следующим рекомендациям:

- Параграф 41: текст не должен быть подчеркнут. Подчеркивание особенно не рекомендуется для химических формул (в этих случаях при распознавании не могут помочь словари). Особенно это вызывает проблемы со всеми теми знаками, которые пересекаются с подчеркивающей линией: l) y p - они неверно распознаются.
- Параграф 39: не рекомендуется курсив. Особенно не рекомендуется менять шрифт внутри слова (оборудование для OPC(OCR) часто предполагает, что все знаки слова имеют одинаковый шрифт). В результате все "1*H*" и "-*N*-" плохо распознаются.



Приложение 2

Пример 20: Полностью нечитаемая страница

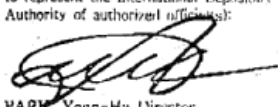
WO 2005/071074

PCT/KR2005/000214

BUENOS AIRES TREATY ON THE INTERNATIONAL RECOGNITION OF THE DEPOSIT
OF MICROORGANISMS FOR THE PURPOSE OF PATENT DEPOSIT

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

TO : **Mugam Biotechnology Research Institute**
#341, Hujung-ro, Nonsung-eup, Yongin-city, Kyonggi-do 449-910,
Republic of Korea

I. IDENTIFICATION OF THE MICROORGANISM	
Identification reference given by the DEPOSITOR: <i>Saccharomyces cerevisiae</i> 013501 / MBLK8 #36	Accession number given by the INTERNATIONAL DEPOSITARY AUTHORITY: KCTC 10582HP
II. SCIENTIFIC DESCRIPTION AND/OR PROPOSED TAXONOMIC DESIGNATION	
The microorganism identified under I above was accompanied by: ☒ a scientific description ☐ 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-Hu Director Date: January 17 2004

Form 1001 (01/01/04) Form 1001

Такая страница не должна приниматься ведомствами: она была послана по факсу с разрешением в 100 точек на дюйм и не читается даже глазом человека. В таких случаях операторы объявляют, что всё содержание страницы, - как изображение, так и текст, - не может быть распознано.

[Конец Приложения 2 и Стандарта]