

Содержание

1. Предмет заявки
2. Описание
Выделенная заявка
История изобретения
Описание материалов, использованных при экспертизе заявки
Краткое описание изобретения
Краткое описание прилагаемых иллюстраций
3. Подробное описание изобретения
Типовая конфигурация сети
Типовой клиент/серверная архитектура
Типовая поисковая система
Типы временных данных
Дата создания документа
Обновление контента
Анализ запросов
Ссылочный критерий
Ссылочный текст
Трафик
Поведение пользователя
Информация о домене
История ранжирования
Заданные/выработанные пользователям данные
Уникальные слова, биграммы, фразы в ссылочном тексте
Связь независимых одноранговых узлов
Разделы документа
4. Примерная обработка
5. Заключение

Документальный поиск, основанный на временных данных.

Краткий обзор

Система, идентифицирующая документ и получающая временные данные, связанные с документом. Система может оценивать документ, по крайней мере, частично, по одному или нескольким типам временных данных.

Создатели:

Ачари, Анураг; (*Кэмпбелл, штат Калифорния*) ; **Катт, Маттс;** (*Маутин Вью, штат Калифорния*) ; **Дин, Джеффри;** (*Пало – Альто, штат Калифорния*) ; **Хагр, Пол;** (*Сан – Франциско, штат Калифорния*) ; **Хенцингер, Моника;** (*Лозанна*) ; **Хельцль, Урс;** (*Пало – Альто, штат Калифорния*) ; **Лоуренс, Стив;** (*Маутин Вью, штат Калифорния*) ; **Пфлегер, Карл;** (*Маутин Вью, штат Калифорния*) ; **Серчиноглу, Олкан;** (*Маутин Вью, штат Калифорния*) ; **Тонг, Саймон;** (*Маутин Вью, штат Калифорния*).

ПРЕДМЕТ ЗАЯВКИ.

1. Метод оценки документа включает в себя следующие компоненты: идентификация документа, получение одного или более типов временных данных документа, установление его ценности на основе этих временных данных.
2. Метод, описанный в пункте 1, где временные данные документа включают информацию, касающуюся момента создания документа. А проведение оценки подразумевает: установление даты создания документа и определение ценности документа по этой дате.
3. Метод, описанный в пункте 2, где документ включает в себя множество других документов, а под оценкой документа понимается: определение точного возраста каждого из множества документов (отсчет ведется с момента его создания), определение среднего времени существования документа на основе возраста каждого из множества документов, и, по крайней мере, частичную оценку документов на основе

4. разницы между возрастaми каждого из множества документов и средним временем существования документа.

5. Метод, описанный в пункте 2, где установление ценности документа включает, по меньшей мере, частичную оценку документа, за истекший период, начиная со времени создания документа.

6. Метод, описанный в пункте 2, где время начала работы над документом ведется с момента, когда поисковая система впервые находит документ или ссылку на него, или с момента, когда документ достигает заранее установленного числа страниц.

7. Метод, описанный в пункте 1, где один или более типов временных данных о документе включают информацию об изменениях контента документа с течением времени. Установление ценности включает: определение частоты, с которой изменяется весь контент документа или его часть.

8. Метод, описанный в пункте 6, где частота, с которой изменяется контент документа, зависит: от среднего времени, за которое происходят изменения, от количества изменений за определенный период времени, и сравнительной оценки изменений документа за текущий и предыдущий период времени.

9. Метод, описанный в пункте 6, где установление ценности также включает: определение объема изменяющегося с течением времени контента, оценку документа на основе частоты изменения объема контента с течением времени.

10. Метод, описанный в пункте 8, где объем изменяющегося контента зависит: от количества новых страниц, относящихся к данному документу, от соотношения количества новых страниц к общему количеству страниц в документе. Учитывается процент контента, который был изменен в течение определенного времени.

11. Метод описанный в пункте 8, где определение объема изменяющегося контента включает: «вес» различных частей контента документа, определяемых по их важности; установление объема, на который изменился контент, как функцию от величины поступившего контента.

11. Метод, описанный в пункте 6, где документ включает в себя множество других документов, а оценка документа включает: определение даты

последнего изменения контента, выведение средней даты изменения с учетом точных дат изменения контента. Оценка документа осуществляется на основе разницы между датой последнего изменения контента и средней датой.

12. Метод, описанный в пункте 1, где один или более типов временных данных о документе включают: информацию, относящуюся к способу изменения контента документа со временем; установление ценности документа включает: определение объема изменяющегося с течением времени контента и оценку этого изменения.

13. Метод, описанный в пункте 12, где объем изменяющегося контента документа, частично, зависит: от количества новых страниц, относящихся к документу, отношения количества новых страниц к общему числу страниц документа, и проценту измененного контента с течением времени.

14. Метод, описанный в пункте 12, где определение объема изменяющегося контента включает: те же параметры, которые описаны в пункте 10.

15. Метод, описанный в пункте 1, где один или более типов временных данных о документе включают информацию о том, насколько часто документ включается в результаты поиска, а установление ценности включает: определение размера документа, когда он был включен в результаты поиска; оценку этого документа, на основе его размера в момент включения в результаты поиска.

16. Метод, описанный в пункте 15, где оценка документа подразумевает присвоение документу более высокой ценности, если в определенный промежуток времени он включается в результаты поиска чаще других.

17. Метод, описанный в пункте 1, где один или более временных данных о документе включают информацию о том, насколько поисковые термины входят в поисковый образ запроса; а установление ценности подразумевает: определение, связан ли документ с поисковыми запросами, которые приводят к результатам, меняющимся в течение определенного промежутка времени и, по крайней мере, частичную оценку на основе указанных запросов.

18. Метод, описанный в пункте 1, где один или более типов временных данных о документе включают информацию о запросах, которые сами в

течение определенного промежутка времени остаются неизменны, но результаты поиска по ним в это время изменяются; и где установление ценности включает: определение документа соответствующего запросам, влекущим изменения в результатах поиска и оценку документа на этой основе.

19. Метод, описанный в пункте 1, где один или более типов временных данных документа включают информацию об устаревших документах, а определение ценности документа включает определение актуальности документа, и оценку документа на этой основе.

20. Метод, описанный в пункте 19, где оценка документа включает: установление факта удовлетворительности данному поисковому запросу документа, определенного как устаревшего, и оценку документа на этой основе.

21. Метод, описанный в пункте 20, где определение степени актуальности устаревшего документа поисковому запросу зависит от частоты выбора данного документа наряду с новыми по этому запросу.

22. Метод, описанный в пункте 1, где один или более типов временных данных о документе включают информацию, относящуюся к динамике простановки ссылок за определенный промежуток времени, а установление ценности документа включает: определение динамики простановки ссылок документа, оценка документа на этой основе.

23. Метод, описанный в пункте 22, где ссылочная динамика зависит от постановки или снятия одной или более ссылок, относящихся к документу.

24. Метод, описанный в пункте 23, где постановка одной и более ссылок связана с датой появления, по меньшей мере, одной новой ссылки на документ, частоте постановки ссылок за определенный период времени и общим количеством ссылок. А снятие одной или более ссылок связано с датой снятия существующей ссылки, частотой их исчезновения и общим количеством ссылок снятых за определенный промежуток времени.

25. Метод, описанный в пункте 22, где динамика постановки ссылок, относящихся к документу, подразумевает контроль, по меньшей мере, за одной из временных характеристик ссылок, учет количества проставленных и снятых ссылок за определенный промежуток времени, а также тенденцию к

появлению большего числа новых ссылок по сравнению с числом снятых ссылок.

26. Метод, описанный в пункте 1, где один или более типов временных данных о документе включает информацию о «новизне» ссылок, а определение ценности документа подразумевает установление новизны ссылок, установление «веса» ссылок по основе их новизны и оценку документа с учетом веса новых.

27. Метод, описанный в пункте 26, где новизна ссылки, относящейся к документу, зависит, по крайней мере, частично от даты постановки ссылки и даты ее изменения; даты появления ссылочного текста и даты его изменения; даты появления ссылающегося документа и его изменения.

28. Метод, описанный в пункте 26, где вес, передаваемый ссылке, зависит от степени надежности, авторитетности и новизны документа, содержащего ссылку.

29. Метод, описанный в пункте 26, где оценка документа включает: определение возраста каждой ссылки, указывающей на документ, проведение возрастного распределения ссылок с учетом времени их существования, и, по крайней мере, частичную оценку документа на основе возрастного распределения ссылок.

30. Метод, описанный в пункте 1, где один или более типов временных данных о документе включает информацию, относящуюся к способу изменения ссылочного текста в течение долгого времени. А определение ценности документа включает: идентификацию изменений ссылочного текста соответствующего ссылке на документ, и, по крайней мере, частичную оценку документа на этой основе.

31. Метод, описанный в пункте 1, где один или более типов временных данных о документе включает информацию, касающуюся различий между документами и ссылочным текстом, соответствующему ссылкам на эти документы. А установление ценности документа включает: определение, было ли изменение контента документа, с условием, что контент отличается от ссылочного текста, соответствующего одной или более ссылкам на документ и, по меньшей мере, частичную оценку документа на этой основе.

32. Метод, описанный в пункте 1, где один или более типов временных данных о документе включает информацию, связанную с новизной ссылочного

текста, и где определение ценности документа включает определение новизны ссылочного текста, соответствующего одной или более ссылкам на документ и, по меньшей мере, частичную оценку документа на этой основе.

33.Метод, описанный в пункте 32, где новизна ссылочного текста, к которому относится ссылка на документ, зависит, по крайней мере, частично от даты появления ссылочного текста и даты его изменения, даты появления документа и даты его изменения.

34. Метод, описанный в пункте 1, где один или более типов временных данных включает информацию о трафике документов, а определение ценности подразумевает определение свойств этого трафика и, по меньшей мере, частично, оценку документа на основе свойств трафика документа.

35.Метод, описанный в пункте 34, где определение свойств трафика документа включает анализ модели трафика документа с целью выявления изменений происходящих с ней за определенный промежуток времени.

36.Метод, описанный в пункте 1, где один или более типов временных данных включает информацию о поведении пользователя при использовании документами, а определение ценности включает: наблюдение за поведением пользователя при использовании документами и, по крайней мере, частичную оценку документа на основе этого поведения пользователя.

37.Метод, описанный в пункте 36, где поведение пользователя соотносится с тем, что хотя бы единожды документ попадал в результаты поиска и с количеством обращений пользователей к документу.

38.Метод, описанный в пункте 1, где один или более типов временных данных включает информацию о доменах, на которых расположены документы. А установление ценности включает: анализ информации о доменах, на которых расположены документы и, по крайней мере, частичную оценку документа на основе результатов анализа.

39.Метод, описанный в пункте 38, где оценка документа включает: определение факта является ли домен, на котором расположен документ, приобретенным законно и оценку документа на этом основании.

40. Метод, описанный в пункте 38, где информация о домене связана со сроком окончания регистрации домена, регистрационной записью DNS и с именем сервера на котором находится домен.

41. Метод, описанный в пункте 1, где одни или более типов временных данных включает информацию о динамике изменения документов в предыдущее ранжирование. А установление ценности включает определение динамики изменения документов в предыдущее ранжирование и, по крайней мере, частичную оценку документа на этой основе

42. Метод, описанный в пункте 41, где оценка документа включает: определение количества или степени изменений в ранжировании документа за определенный промежуток времени и, по крайней мере, частичную оценку документа на этой основе.

43. Метод, описанный в пункте 41, где динамика изменения документов в предыдущее ранжирование зависит: от числа запросов, по которым документ включался в результаты поиска за определенный промежуток времени; частоты, с которой документ включался в результаты поиска за данный промежуток времени; сезонных колебаний; пакетных данных; изменения за определенный промежуток времени оценки пары запрос-URL.

44. Метод, описанный в пункте 41, где динамика изменения документов в предыдущем ранжировании включает контроль за амплитудой колебаний позиций документа в процессе ранжирования за определенный период времени.

45. Метод, описанный в пункте 1, где одни или более типов временных данных включает информацию о созданных или сохраненных пользователем данных, а определение ценности включает: определение того, указывают ли созданные или сохраненные пользователем данные, что документ представляет интерес для пользователя и, по крайней мере, частичную оценку документа на этой основе.

46. Метод, описанный в пункте 45, где созданные или сохраненные пользователем данные относятся, по меньшей мере, к одному из списков Избранного, закладкам, кэшированным файлам, которые связаны между собой одним или множеством пользователей.

47. Метод, описанный в пункте 45, где оценка документа включает: анализ созданных или сохраненных пользователем данных за определенный промежуток времени для определения, по меньшей мере, одного из направлений, в соответствии с которым следует добавлять или удалять данные из документа, определение частоты, с которой добавлялись или удалялись данные, созданные или сохраненных пользователем и, по крайней мере, частичную оценку документа на основе результатов анализа.

48. Метод, описанный в пункте 1, где одни или более типов временных данных включает информацию об увеличении параметров ссылочного текста, и где установление ценности включает: определение возросших параметров ссылочного текста, к которому относятся одна или более ссылок на документ и, по крайней мере, частичную оценку документа на этой основе.

49. Метод, описанный в пункте 1, где одни или более типов временных данных включает информацию о ссылочных связях независимых друг от друга ресурсах одного уровня, и где установление ценности документа включает: определение увеличения числа независимых друг от друга ресурсах одного уровня, которые включают документ и, по крайней мере, частичную оценку документа на основе количества указанных ресурсов.

50. Метод, описанный в пункте 1, где одни или более типов временных данных включает информацию обо всех тематиках документа, и где установление ценности включает: отбор подтемы, относящейся к документу, контроль за ее изменением в течение определенного промежутка времени и, по крайней мере, частичную оценку документа на основе этих изменений.

51. Метод, описанный в пункте 1, кроме того, также включает: формулировку поискового запроса, при условии, что идентифицированный документ определяется как релевантный этому запросу; оценку релевантности, которая зависит от степени релевантности документа поисковому запросу, а установление ценности документа, по крайней мере, частично зависит от типов временных данных и оценке релевантности.

52. Система оценки документа включает: способы идентификации документа, способы получения множества типов временных данных о документе и способы установления ценности документа, по крайней мере, частично зависящие от множества типов временных данных.

53. Система оценки документа включает: временной компонент, который видоизменяется в зависимости от получения одного или более типов временных данных о документе, и ранжирующий компонент, который видоизменяется в зависимости от установленной ценности документа, которая, в свою очередь, по крайней мере, частично основывается на типах временных данных.

54. Метод ранжирования документа, на который ссылаются, включает: определение возраста ссылочных связей этого документа и ранжирование

документа, на который ссылаются по убывающей функции возраста ссылочных связей.

55.Метод, описанный в пункте 54, где данные о ссылочных связях включают, по меньшей мере, одну ссылку.

56.Метод, описанный в пункте 54, где данные о ссылочных связях включают ссылочный текст.

57. Метод, описанный в пункте 54, где данные о ссылочных связях включают ранжирование. По крайней мере, частично оно зависит от ссылок и ссылочного текста, соответствующих одному или более ссылающихся документов и связанных с документом на который ссылаются.

58. Метод, описанный в пункте 57, кроме того, также включает: определение долговечности ссылочного окружения; получение показателя обновления контента ссылающегося документа, обеспечивающего данные о ссылочном окружении; регулирование процесса ранжирования ссылающегося документа на основе срока службы ссылочного окружения и показателя обновления контента ссылающегося документа.

59.Метод, описанный в пункте 58, где регулирование процесса ранжирования включает понижающее ранжирование, если срок службы ссылочного окружения короткий, и повышающее ранжирование в случае долгого срока службы ссылочного окружения.

60.Метод, описанный в пункте 59, где регулирование процесса ранжирования включает: понижающее ранжирование, если, по меньшей мере, одна часть контента ссылающегося документа учитывается как устаревшая за определенный промежуток времени и повышающее ранжирование, если часть контента ссылающегося документа обновлена за определенный промежуток времени.

61.Метод, описанный в пункте 54, включает выявление линкопомойки для ссылающегося документа, обеспечивающего данные о ссылочном окружении; и регулирование процесса ранжирования документа на основе данных о линкопомойке.

62. Метод, описанный в пункте 61, показатель линкопомойки вычисляется как функция степени, когда одна или более ссылок, обеспеченных ссылающимся документом, меняются за определенный промежуток времени.

63. Метод, описанный в пункте 62, где регулирование процесса ранжирования включает понижающее ранжирование, если линкопомойка превышает допустимый предел.

ОПИСАНИЕ

Выделенная заявка.

1. По данной заявке заявлен приоритет 119 раздела 35-ого Свода законов США на основе предварительной заявки № 60/507,617, поданной 30 октября 2003 года, содержание которой оформлено надлежащим образом.

История изобретения.

2. Область изобретения.

3. Настоящее изобретение напрямую связано с системой выборки информации, а более точно, с системами и методами обработки результатов поиска, которые зависят, по крайней мере, частично, от временных данных, содержащихся в релевантных документах.

4. Описание материалов, использованных при экспертизе заявки.

5. Система World Wide Web содержит большой объем информации. Поисковые системы помогают пользователям размещать необходимые порции данной информации в веб-каталогах. Как правило, в ответ на запрос пользователя, поисковая система предоставляет ссылки на документ, релевантный запросу.

6. Поисковые системы анализируют предпочтения пользователей, основываясь на поисковых запросах. Цель любой поисковой системы -

предоставить ссылки, которые ведут к максимально релевантным результатам по заданному запросу. Как правило, поисковая система выполняет поставленную задачу, подбирая веб-документы к поисковому запросу. Веб-документы, содержащие искомую терминологию, используемую пользователем в запросе, предоставляются пользователю.

7. В идеале, поисковая система в ответ на запрос пользователя, должна предоставить максимально релевантные результаты. Одна из категорий поисковых систем определяет релевантные документы на основе сравнения терминов, используемых в поисковом запросе со словами, содержащимися в документе. Другая категория поисковых систем определяет релевантные документы с использованием других или дополнительных факторов. Например, поисковая система может руководствоваться информацией, которую несут в себе ссылки на или с документов для определения релевантности найденной информации.

8. Обе категории поисковых систем направлены на предоставление высококачественных результатов по поисковому запросу. Существует несколько факторов, которые могут оказать влияние на качество результатов, предоставляемых поисковыми системами. Например, некоторые создатели сайтов используют спамерские приемы для создания искусственной популярности. Устаревшие документы (т.е. документы, которые не обновлялись в течение длительного периода времени и содержат устаревшую информацию) могут иметь преимущества перед «свежими» документами (т.е. документами, которые недавно были обновлены и содержат более новую информацию). В ряде случаев, более высокие позиции устаревших документов ухудшают результаты поиска.

9. Существует необходимость повышения качества результатов, предоставляемых поисковыми системами.

Краткое описание изобретения.

10. Системы и методы, используемые в данном изобретении, осуществляют оценку документов на основе временных данных документов. Оценка документов используется для улучшения результатов поиска, которые были сформированы по поисковому запросу.

11. В соответствии с одним из аспектов, удовлетворяющим принципам работы изобретения, предлагается метод оценки документа. Метод включает в

себя идентификацию документа и привлечение данных, относящихся к документу. В дальнейшем метод будет включать, по крайней мере, частичное установление ценности документа, на основе одного или более типов временных данных.

12. В соответствии с другим аспектом, предлагается иной метод оценки документа. Метод включает в себя определение возраста ссылочного окружения, относящейся к документу на который ссылаются, и оценку документа, основываясь на возрасте ссылочной информации.

Краткое описание прилагаемых иллюстраций.

13. Сопровождающие иллюстрации, включенные в данную спецификацию, наглядно представляют изобретение и, наряду с описанием, объясняют его.

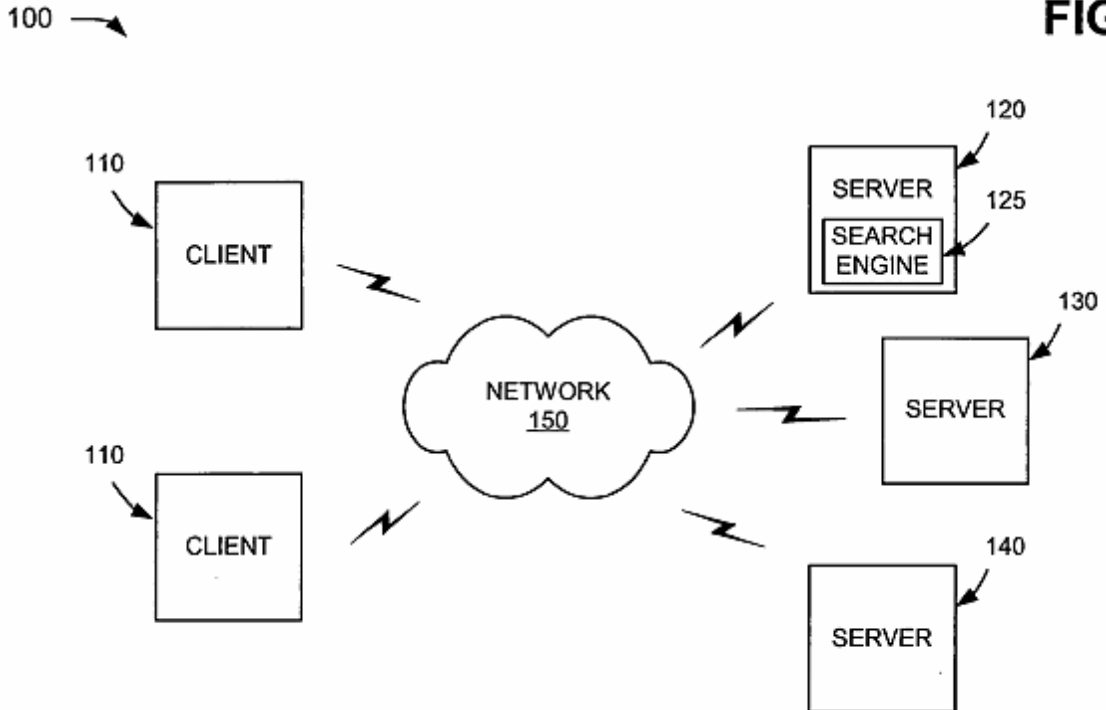
Описание терминов, употребляемых в иллюстрациях.

410 – идентификация документа
420 – получение исторических данных, связанных с документом
430 – оценка документа, основанная, по крайней мере, частично на временных данных

110 – клиент
120,130,140 – сервер
125 – поисковая система

260 – устройство ввода
270 – устройство вывода
280 – интерфейс
230 – оперативная память
240 – постоянное запоминающее устройство
250 – запоминающее устройство

FIG. 1



[0014] Рис.1 – диаграмма типовой сети, в которой системы и методы, используемые в изобретении, находят свое применение.

110-140 →

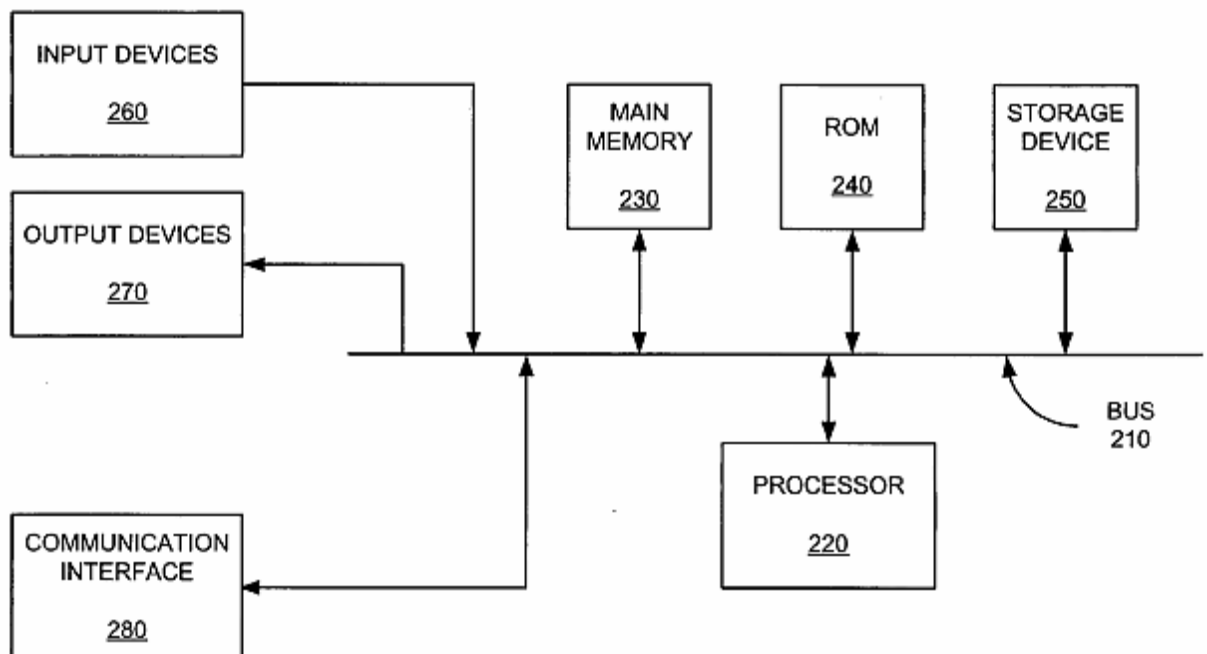
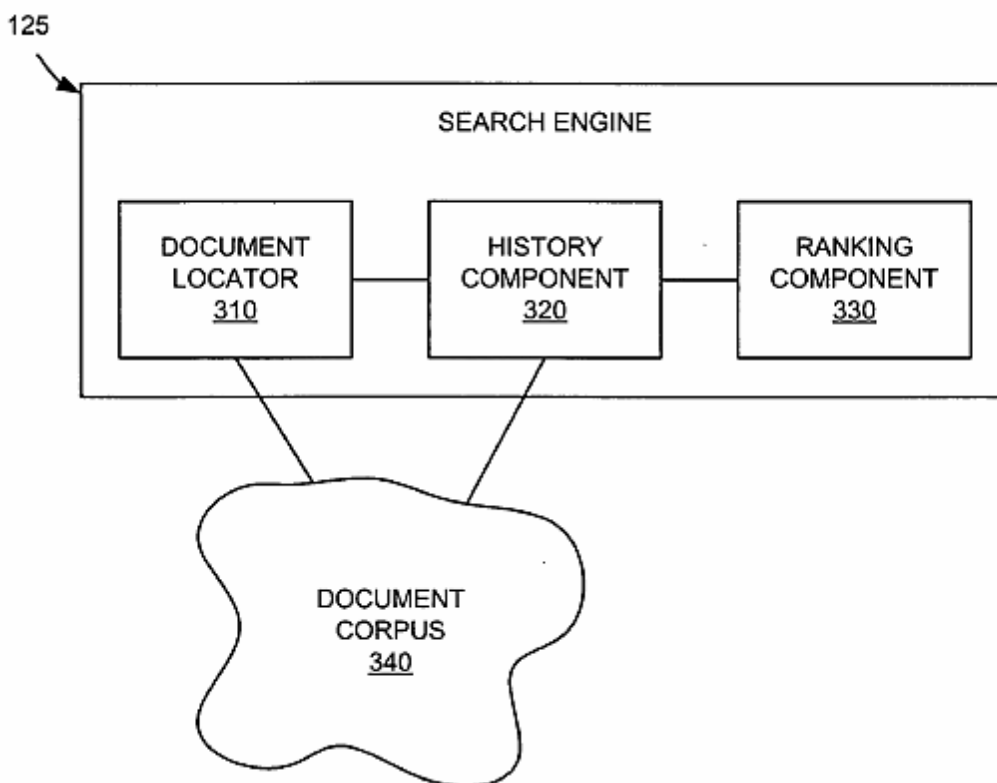


FIG. 2

[0015] Рис.2 – типовая диаграмма клиента/сервера рис.1, удовлетворяющая принципам изобретения.



[0016] Рис.3 – блок-схема примерного процесса оценки документа, удовлетворяющая принципам изобретения.

ПОДРОБНОЕ ОПИСАНИЕ ИЗОБРЕТЕНИЯ

18. Детальное описание изобретения связано с прилагаемыми иллюстрациями. Ссылки на номера различных иллюстраций могут указывать на одинаковые или похожие элементы. Предлагаемое детальное описание не ограничивает возможности изобретения.

19. Системы и методы, используемые в данном изобретении, способствуют оценке документа, используя, например, историческую информацию, касающуюся определенных документов. В системах и методах используется оценка для предоставления высококачественных результатов поиска.

20. «Документ» - термин, употребляемый в широком значении. Документ может включать в себя электронную почту, веб-сайт, файл, комбинацию файлов, один или более файлов, содержащих ссылки на другие файлы, блог, интернет-рекламу и др. В Интернете обычный документ - веб-страница. Веб-страницы часто включают в себя текстовую информацию, а также мета-информацию, рисунки, гиперссылки и др. Страница может относиться к

документу или части документа. Поэтому слова «страница» и «документ» иногда взаимозаменяемы. В других случаях, страница относится к части документа, например, к суб-документу. Страница может относиться более чем к одному документу.

21. В предлагаемом ниже описании документы представлены как:
- документы, содержащие ссылки на другие документы и/или с других;
 - документов. Например, когда документ содержит ссылку на другой документ;
 - документ, ссылка считается прямой. Когда ссылка указывает на документ, ссылка считается обратной. При использовании термина «ссылка», подразумевается как прямая, так и обратная ссылка.

Типовая конфигурация сети.

22. Рис.1- типовая диаграмма сети 100, в которой могут быть применимы методы и системы, используемые в данном изобретении. Сеть 100 включает различных клиентов 110, связанных с разнообразными серверами 120-140 через сеть 150. Сеть 150 включает локальную сеть, глобальную, телефонную, Интранет, Интернет, устройство памяти, другую сеть или комбинацию сетей. Два клиента 110 и три сервера 120-140 связаны с сетью 150 из-за простоты и удобства пользования. На самом деле, клиент может выполнять функции сервера, в этом случае, становится понятно, какими конкретно функциями наделен сам клиент.

23. Клиенты 110 представляют собой сообщество. Такое сообщество определяется как своего рода инструмент коммуникации, напоминающий по своим функциям беспроводной телефон. Серверы 120-140 представляют собой сообщества, которые осуществляют сбор, поиск, обработку информации, а также приводят в надлежащий вид документы, которые должны удовлетворять принципам изобретения. Клиенты 110 и серверы 120-140 связываются с сетью 150 через проводные, беспроводные и оптические средства связи.

24. При введении в эксплуатацию изобретения, серверы 120 связаны с поисковой системой 125, которую используют клиенты 110. Сервер 120 обходит, индексирует документы, сохраняет информацию, касающуюся документов, в специальном «хранилище» документов, которые были обойдены. Серверы 130 и 140 сохраняют документы, которые обходит сервер 120. Пока серверы 120-140 представлены как сообщества, существует возможность для одного или более серверов 120-140 выполнять одну или более функций других серверов 120-140. Например, два или более сервера 120-140 вводятся в эксплуатацию как один единый сервер. Существует возможность для одного

или более серверов 120-140 начать действовать как два или более отдельных инструмента.

Типовой клиент/серверная архитектура.

25. Рис.2 - типовая диаграмма сообщества клиентов или серверов (далее сообщество клиентов/серверов), которая соответствует одному или более клиентам 110 и серверам 120-140 в соответствии с применением принципов, соответствующих изобретению. Модуль клиентов/серверов может включать канал передачи информации 210, процессор 220, основную оперативную память 230, память только для чтения (ROM) 240, блок памяти 250, один или более устройств для ввода данных 270, коммуникационный интерфейс 280. Канал передачи информации 210 включает один или более проводников, которые обеспечивают коммуникацию между компонентами сообщества клиентов/серверов.

26. Процессор 220 включает один или более процессоров обычного типа или микропроцессоров, которые обрабатывают и выполняют указания.

Основная оперативная память 230 может включать в себя память с прямой выборкой (RAM) или другой тип динамического блока памяти, который хранит информацию и инструкцию по применению процессора 220. ROM 240 может включать ROM обычного типа или другой тип статического блока памяти, который хранит статическую информацию и указания по применению процессора 220. Блок памяти 250 включает магнитный носитель информации и соответствующий накопитель информации.

27. Устройство(а) для ввода данных 260 включает один или более механизмов обычного типа, которые позволяют оператору вводить информацию в сообщество клиентов/серверов, например, клавиатура, мышь, биометрические механизмы и др. Устройство(а) для ввода данных включает один или более механизмов обычного типа, которые передают информацию оператору, а также на дисплей, принтер, спикер и т.д. Коммуникационный интерфейс 280 включает в себя механизм-трансивер, который позволяет сообществу клиентов/серверов вступать в контакт с различными устройствами или системами. Например, коммуникационный интерфейс 280 включает механизмы, позволяющие вступать в контакт с другими устройствами или системами через сеть, например, сеть 150.

28. Как будет описано ниже, сообщество клиентов/серверов, удовлетворяющее принципам изобретения, выполняет определенные

поисковые операции. Сообщества клиентов/серверов выполняют эти операции в соответствии с инструкциями программного обеспечения процессора 220, которые содержатся в памяти 230. Вычислительная среда может быть определена как логическое или физическое устройство памяти.

29. Указания программного обеспечения могут читаться в памяти 230 с другой вычислительной среды, например, с устройства хранения данных 250 или с другого устройства через коммуникационный интерфейс 280. Указания программного обеспечения, содержащиеся в памяти 230, могут вызвать процессы в процессоре 220, которые будут описаны ниже. Реализация принципов изобретения не ограничивается специфической комбинацией схем и программного обеспечения.

Типовая поисковая система.

30. Рис.3 – типовая функциональная структурная схема поисковой системы 125, удовлетворяющей принципам изобретения. Поисковая система 125 включает адрес документа, временной компонент 320 и оценочный компонент 330. Как показано на рис.3, один или более адресов документа 310 и исторический компонент 320 включают информацию, относящуюся к документам, которые были обойдены, проиндексированы и сохранены, например, в базе данных, доступных с помощью поисковой системы 125. Временные данные, как будет описано чуть ниже, связаны с каждым документом в своде документов 340. Такие временные данные сохраняются в своде документов 340 либо где-то еще.

31. База документов 310 идентифицирует блок документов, контент которых удовлетворяет поисковому запросу. База документа 310 первоначально включает документы из свода документов 340, сравнивая термины в поисковых запросах пользователей с информацией, содержащейся в своде документов. В целом, процессы индексации документов и поиска уже проиндексированных документов, которые содержат поисковые термины, хорошо известны. Поэтому функциональное назначение базы документов 310 подробно не описывается.

32. Временной компонент 320 собирает временные данные, касающиеся документов в своде документов 340. При реализации принципов изобретения, временные данные включают информацию, касающуюся даты создания документа, изменений в документе, анализа запросов, ссылочного текста (например, текст, который содержит гиперссылку, обычно ярко выделенную в документе), информацию о трафике, поведении пользователей, а также информацию, собранную пользователями. Различные типы временных данных

дополнительно описаны ниже. Помимо этого, временные данные включают дополнительную информацию.

33. Компонент ранжирования 330 определяет показатели ранжирования (одного или более документов в своде документов 340). Компонент ранжирования 330 определяет показатели ранжирования как в зависимости, так и вне от поискового запроса. Когда документы ассоциируются с определенным поисковым запросом (например, документ идентифицирован как релевантный определенному поисковому запросу), поисковая система 125 сортирует документы на основе показателей ранжирования и возвращает отсортированные документы пользователю, который сделал соответствующий запрос. С учетом аспектов изобретения, показатели ранжирования - показатели, которые способствуют качественной и количественной оценке документов. При реализации принципов изобретения показатели ранжирования базируются, по крайней мере, частично на исторической информации, полученной от исторического компонента 320.

ТИПЫ ВРЕМЕННЫХ ДАННЫХ

Дата создания документа.

34. При реализации принципов изобретения дата создания документа необходима при формировании показателей, относящихся к документу. Термин «дата» используется достаточно широко. Как описано выше, существует несколько приемов для определения даты создания документа. Некоторые приемы «нетрадиционны», в том смысле, что на них могут влиять, те, кто желает улучшить показатели, относящиеся к документу. Другие приемы являются традиционными. Любые из этих приемов, комбинации этих приемов, а также другие приемы могут использоваться для определения даты создания документа.

35. Определение даты создания документа возможно, начиная с даты, когда поисковая система 125 впервые узнает или проиндексирует документ. Поисковая система 125 исследует документ посредством его обхождения, через представление документов поисковой системе 125 из внешних источников или посредством комбинации двух приемов. И наоборот, определение даты создания документа возможно, начиная с даты, когда поисковая система 125 впервые обнаруживает ссылку на заданный документ.

36. В соответствии с другим подходом, дата регистрации документа под доменным именем может рассматриваться как дата создания документа. Еще один подход определяет дату создания документа с момента появления ссылки на данный документ с другого документа, например, это может быть новая статья, группа новостей или комбинация одного или более таких документов. В соответствии с подходом, который будет рассматриваться в дальнейшем, дата, когда документ стал включать в себя необходимое количество страниц, считается датой создания документа. Существуют другие подходы к определению даты создания документа, но здесь они рассматриваться не будут.

37. Поисковая система 125 использует дату создания документа для оценки документа. Например, справедливо предположить, что недавно созданный документ не может содержать достаточное количество ссылок с других документов. Оценка документа основывается на количестве ссылок с или на документ. Соответственно, документ, который создан недавно, будет оценен ниже, чем документ, который содержит достаточное количество ссылок (например, обратных). Учитывая дату создания документа, оценка документа будет изменяться либо в положительном направлении, либо в отрицательном.

38. Приведем в пример документ, который был создан вчера и имеет десять обратных ссылок. Такой документ будет оценен поисковой системой выше, нежели документ десятилетней давности, который содержит около ста обратных ссылок. Объяснить это явление достаточно просто: темпы ссылочного роста первого документа значительно выше, нежели темпы последнего. Резкий рост количества обратных ссылок является не только фактором, которым руководствуется поисковая система 125 при оценке документов, но может служить сигналом к попытке «заспамить» поисковую систему 125. В такой ситуации поисковая система 125 понизит критерии оценки документа с целью минимизации возможности спама.

39. В соответствии с реализацией принципов данного изобретения, поисковая система 125 использует дату создания документа для определения

правил, которые действуют в документах при создании ссылок. Эти правила в дальнейшем используются для оценки документа, например, больший «вес» получают те документы, на которые ссылки ставятся наиболее часто.

40. При одном из подходов поисковая система 125 изменяет оценку документа на основе ссылок следующим образом:

$$H=L/\log(F+2),$$

41. где H -оценка временной заданности ссылки;

L - оценка ссылки, присвоенной документу, которая может быть выявлена при использовании любых известных приемов оценки ссылок (например, прием оценки ссылки, описанный в американском патенте № 6,285,999);

F- время, истекшее с момента создания документа (или какой-то промежуток времени).

42. По некоторым запросам, более старые документы являются предпочтительнее новых. В результате, имеет смысл оценивать документы, основываясь на времени их создания. Другими словами, поисковая система 125 определяет возраст документов при выдаче результатов (например, учитывая дату создания документа), определяет средний возраст документов, изменяет оценку документа (положительно или отрицательно) на основе разницы между возрастом определенного документа и средним возрастом документов.

43. В итоге, поисковая система 125 по крайней мере, частично оценивает документ, основываясь, на информации, относящейся к дате создания документа.

Обновление контента.

44. В соответствии с реализацией принципов данного изобретения, информация, относящаяся к способу изменения контента документа с течением времени используется для изменения оценки данного документа. Например, документ, контент которого неоднократно редактировался, будет оцениваться по-другому, нежели документ, который оставался неизменен в течение длительного периода времени. Документ, содержащий большой объем изменяющегося со временем контента, будет оцениваться по-другому, нежели документ, содержащий маленький объем изменяющегося со временем контента.

45. В соответствии с одним из подходов, поисковая система 125 оценивает обновленный контент следующим образом:

46. $U=f(UF, UA)$,

47. где f – функция, например, от суммы или взвешенной суммы;

UF- оценка частоты, которая показывает, как часто документ (или страница) обновляется,

UA- количественная оценка, которая показывает, насколько значительно документ (или страница) был изменен в течение времени.

UF определяется несколькими способами с учетом среднего времени между обновлениями, количества обновлений за определенный период времени и т.д.

48. UA определяется как влияние одного или более факторов, например, количества «новых» или уникальных страниц, относящихся к документу в течение определенного периода времени. Еще один фактор включает отношение количества новых или уникальных страниц, относящихся к документу в течение определенного периода времени, к общему количеству страниц, относящихся к документу. Другой фактор учитывает, сколько раз документ (или страница) изменялся в течение определенного периода времени (например, в течение последних дней).

49. В соответствии с традиционным подходом, UA может быть определен как функция, вычисляемая в зависимости от различий веса частей контента документа. Например, контент на первый взгляд не представляет никакого интереса после произошедшего обновления. Это могут быть комментарии, реклама, навигационные элементы, газетные материалы. С другой стороны, контент может представлять большой интерес после обновления, например, заголовков или ссылочный текст могут получить больший вес, чем остальной контент.

50. UF и UA оказывают влияние на оценку документа. Например, уровень изменений за текущий период времени сравнивается с уровнем изменений за другой период времени с целью определения различных тенденций. Документы, в которых наблюдаются значительные изменения, оцениваются выше, чем документы, которые более стабильны. Количество изменений оказывает влияние на оценку. Например, документы, которые подвержены значительным изменениям, превышающим пороговый уровень, оцениваются выше, чем документы, где изменения стабильны и не превышают порогового уровня.

51. Иногда ресурсов хранения информации оказывается недостаточно для хранения документов для контроля с целью обнаружения изменений контента. В этом случае поисковая система 125 сохраняет формулировку документов и контролирует эти формулировки на предмет изменений. Например, поисковая система 125 сохраняет «подписи» документов вместо самих документов с целью отслеживания изменений контента документа. В этом случае поисковая система 125 сохраняет вектор терминов для документа (страницы) и наблюдает за происходящими изменениями. В соответствии с другим подходом, поисковая система 125 сохраняет и отслеживает относительно небольшие части документа, которые являются важными или наиболее часто встречаются (за исключением стоп слов).

52. В соответствии с еще одним подходом, поисковая система 125 сохраняет краткое содержание документа и следит за изменением информации. В соответствии с дальнейшим подходом, поисковая система 125 формирует похожую «ненужную» информацию (которая используется для отслеживания дубликатов документа), относящуюся к документу, и следит за изменениями. Изменения такой «ненужной» информации могут соответствовать крупным изменениям самого документа. При использовании других подходов прибегают к разнообразным приемам отслеживания изменений в документе. При наличии ресурсов хранения информации, документы хранятся полностью и используются для определения изменений.

53. По некоторым запросам документы, контент которых не был изменен, являются более предпочтительными, чем документы, контент которых был изменен недавно. В результате, более выгодно оценивать документ, основываясь на времени произошедших изменений. Другими словами, поисковая система 125 определяет дату, когда контент каждого документа, находящегося в результатах поиска, был изменен в последний раз, определяет среднюю дату изменений документа и модифицирует оценку документа (отрицательно или положительно) на основе разницы между датой изменения документа и средней даты.

54. В заключение, поисковая система 125 устанавливает (или изменяет) оценку документа на основе временных данных. Для очень объемных документов, которые включают контент, принадлежащий нескольким лицам или организациям, оценка соответствует каждому суб-документу.

Анализ запросов.

55. В соответствии с реализацией принципов данного изобретения, один или более факторов, касающихся запроса, используются для установления (или изменения) оценки документа. Например, один фактор, касающийся запроса, связан с частотой выборки документа в течение времени, когда документ включен в результаты поиска. В этом случае, поисковая система 125 должна оценивать документы, которые выбирают пользователи, выше, чем остальные документы.

56. Другой фактор, касающийся запроса, связан с частотой появления определенного поискового термина в запросах. Определенный набор поисковых терминов увеличивается в запросах с течением времени. Например, запросы, относящиеся к «горячим» темам, которые набирают популярность в течение определенного промежутка времени, возможно, будут появляться чаще. В этом случае, поисковая система 125 оценивает документы, относящиеся к этим поисковым запросам, выше, чем документы, которые к этим запросам не относятся.

57. Еще один фактор, касающийся запроса, относится к изменениям с течением времени в результатах поиска по похожим запросам. Значительное увеличение результатов поиска по похожим запросам должно указывать на наличие «горячей» темы или свежих новостей и провоцировать поисковую систему 125 увеличивать оценку документов, относящихся к данным запросам.

58. Еще один фактор, касающийся запроса, относится к запросам, которые остаются относительно постоянными с течением времени, но приводят к результатам, которые периодически изменяются. Например, запрос «мировой чемпион» относится к таким запросам. Запрос остается неизменным, а результаты с течением времени постоянно меняются. Подобные изменения отслеживаются и соответственно используются для оценки документа.

59. Еще один фактор, касающийся запроса, относится к устаревшим документам, которые выдаются в результатах поиска. «Несвежесть» документа устанавливается на по таким факторам, как дата создания документа, ссылочный рост, трафик, изменения контента, увеличение прямых/обратных ссылок и т.д. Поисковая система 125 узнает, какие запросы являются наиболее важными, анализируя наиболее востребованные документы в результатах поиска. Особенно поисковая система 125 обращает внимание, насколько предпочтительнее являются документы, в которых произошли изменения и которые занимают более низкую позицию, по сравнению с устаревшими документами. Более того, если в течение времени определенный документ удовлетворяет популярным запросам, (например, «Мировой чемпион») и

существуют специфические запросы (например, «янки в Нью-Йорке»), то фактор, касающийся запроса, будет уменьшать ценность документа, который является более устаревшим.

60. Иногда устаревший документ предпочтительнее, чем относительно новые документы. В результате, поисковая система 125 учитывает частоту, с которой избирается документ в течение времени, устанавливая ценность документа. Например, если по определенному запросу пользователи предпочитают выбирать устаревшие документы, которые занимают невысокие позиции, то поисковая система 125 будет оценивать такой устаревший документ.

61. Еще один фактор, касающийся запроса, учитывает частоту, с которой документ появляется в результатах поиска по различным запросам. Например, если определенный документ постоянно появляется в результатах выдачи, абсолютно не соответствуя запросу, то поисковая система 125 посчитает такой документ спамом и оценит его очень низко.

62. В итоге, поисковая система 125 изменяет оценку документа, основываясь, по крайней мере, частично, на одном или более факторов, касающихся запросов.

Ссылочный критерий.

63. В соответствии с реализацией принципов изобретения. Один или более ссылочных факторов используются для изменения оценки документа. В соответствии с одним подходом, ссылочные факторы относятся к дате появления новых ссылок на документ и к исчезновению существующих ссылок. Дата появления ссылки рассматривается как дата, начиная с которой поисковая система 125 находит ссылку (например, дата, когда была обнаружена ссылка на документ или дата, когда она была в последний раз обновлена). Дата исчезновения ссылки является первой датой, когда документ, содержащий ссылку, может ее потерять или исчезнуть сам.

64. Такие даты определяются поисковой системой 125 во время процесса обхождения или индексирования. Используя в качестве такой даты адресную ссылку, поисковая система 125 отслеживает поведение ссылок, например, когда ссылки появились и когда исчезли, частотность, с которой ссылки появляются и исчезают, количество появляющихся и исчезающих ссылок за определенный промежуток времени, поиск тенденции появления новых ссылок и исчезновения старых и т.д.

65. С учетом поведения ссылок, поисковая система 125 оценивает документ. Например, тенденция к уменьшению количества новых ссылок (основанная на сравнении количества новых ссылок в предыдущий и текущий периоды) в течение времени подскажет поисковой системе 125, что документ является «устаревшим». В этом случае, поисковая система 125 понизит оценку документа. И наоборот, тенденция к росту подскажет поисковой системе 125, что документ «свежий» (например, контент документа был недавно создан или обновлен). В зависимости от ситуации и подхода, такой документ считается более релевантным.

66. Анализируя поведение ссылок (увеличение/уменьшение обратных ссылок на документ) в течение определенного времени, поисковая система 125 извлекает ценную информацию по поводу того, насколько документ является «свежим». Например, если подобный анализ свидетельствует об отсутствии обновлений в течение длительного периода времени, уменьшении значимости документа, то такой документ расценивается как устаревший.

67. В соответствии с одним из подходов, анализ зависит от количества новых ссылок на документ. Например, поисковая система 125 отслеживает количество новых ссылок на документ за последние n дней и сравнивает с количеством ссылок с момента появления документа. И наоборот, поисковая система 125 определяет возраст старых ссылок и сравнивает их с самой первой ссылкой на документ.

68. Для иллюстрации рассмотрим $u=10$ и два документа (в данном случае веб-сайты), которые были найдены 100 дней назад. 10% ссылок на первый сайт было найдено раньше, чем 10 дней назад, в то время как, на второй сайт за то же время было найдено 0% ссылок. Они все были найдены раньше. В этом случае метрический показатель для сайта А – 0,1, а для сайта В – 0. Метрические данные могут получить другой вид, если подвергнуть детальному анализу размещение ссылочных дат. Например, создать модули, определяющие зависимость типа сайта от того, как расставлены ссылки (сайт, более не обновляемый, популярный, непопулярный или замененный сайт).

69. По другим расчетам, анализ может зависеть от веса ссылок. В данном случае найти вес ссылки поможет ее функция, важность которой зависит от актуальности ссылки. Актуальность ссылки определяется по дате появления/изменения ссылки, основного текста, связанного с ссылкой, либо документа, содержащего ссылки. Дата появления/изменения документа, содержащего ссылки, является лучшим показателем актуальности ссылки,

Данное утверждение основано на теории о том, что хорошие ссылки остаются неизменными даже при изменении контекста. Чтобы не менять даты всех ссылок вслед за незначительными изменениями документа, определяется важность внесенных изменений (изменения в большом разделе документа или в нескольких небольших разделах). И в соответствии с этим меняют или не меняют даты ссылок.

70. Определять вес ссылок можно по-разному. Например, на основе важности документа, содержащего ссылки (правительственным документам присваивается высокая важность); надежности документа, содержащего ссылки (надежность документа определяется в соответствии с Патентом США № 6.285.999); даты документа, содержащего ссылки, используя дополнительные способы определения даты (часто изменяемый документ внезапно теряет ссылку, например Yahoo home page).

71. Поисковая система 125 может высоко или низко оценить документ, в котором есть ссылки, выполняющие функцию суммы веса ссылок, указывающих на этот документ. Этот способ может применяться рекурсивно (т.е. в обратном порядке). Например, предположим, что документу S 2 года. Документ S можно считать актуальным, если $n\%$ ссылок на нем актуальны или документ, содержащий прямые ссылки на S, является актуальным. Последнее можно определить по дате создания документа.

72. В соответствии с другим способом, анализ может зависеть от срока давности размещения ссылок, указывающих на документ. Т.е. даты создания ссылок можно определить и применить для нахождения даты размещения. Можно предположить, что даты устаревшего и актуального документа будут сильно отличаться друг от друга. Поисковая система 125 сможет оценить документ лишь по дате размещения ссылок.

73. По дате появления ссылок можно выявить «спам», документ с большим количеством ссылок для привлечения большего внимания поисковой системы. Обычный, «законный» документ привлекает обратные ссылки медленно. Много обратных ссылок может сигнализировать о возникновении нового сайта (например, CDC веб-сайт на первом этапе своего существования может создавать много ссылок) или о попытках спама в поисковых системах (преследуя более высокую позицию в результатах поиска), обмениваясь ссылками, покупая их или получая их с документов, не имеющих право делать ссылки. Примерами таких документов являются гостевые книги, форумы, страницы «для всех», позволяющие любому добавить ссылку на документ.

74. В соответствии со следующим способом, анализ может зависеть от даты исчезновения ссылки. Исчезновение большинства ссылок может означать, что документ, на который они указывают, устарел (больше не модернизируется или заменен другим документом). Например, поисковая система 125 может отслеживать дату исчезновения или уменьшения ссылок на документ в специальном временном окне. Поисковая система 125 может перестать учитывать ссылки с устаревшего документа.

75. В данном способе анализ зависит не только от возраста ссылок на документ, но и от динамики их проставления. Например, поисковая система 125 может по-разному определить вес документа, где каждый день разные ссылки, несмотря на их актуальность (например, ниже) и документа, который постоянно обновляется и ссылается на данный целевой документ. В данном случае, поисковая система 125 может оценить документ, основанный на оценке документов с ссылками на все документы, данные о которых есть во временном окне.

76. Поисковая система 125 может производить (изменять) оценку документа, по одному имеющемуся или более ссылочным факторам.

Ссылочный текст.

77. Информация, связанная с тем как изменяется ссылочный текст в течение времени, используется для оценки документа. Например, изменения в ссылочном тексте, связанные с ссылками на документ, указывают на то, что в сам документ были внесены изменения.

78. Если в контенте документа происходят изменения, такие, что он начинает значительно отличаться от ссылочного текста, связанного с обратными ссылками, то и домен, относящийся к документу, значительно или полностью изменяется. Это происходит, когда домен устаревает. Ссылочный текст часто рассматривают как часть документа, на который указывают связанные с ним ссылки. Домен может выдать в результатах поиска несвязанную с запросом ссылку, что не желательно.

79. Одним из решений данной проблемы является определение даты, когда домен изменил местоположение. Это можно сделать, определив дату, когда текст документа либо ссылочный текст был сильно изменен. Все ссылки и/или ссылочный текст, предшествующие этой дате, могут быть проигнорированы.

80. Актуальность ссылочного текста используется как один из принципов оценки документов. Она может быть определена по датам появления/изменения ссылочного текста, ссылок, связанных с ссылочным текстом, либо документа, на который указывают ссылки. Дата появления/изменения документа, на который указывают ссылки, является хорошим индикатором актуальности ссылочного текста. Это утверждение основывается на теории о том, что хороший ссылочный текст остается неизменным, когда модернизируется документ. Чтобы не изменять дату ссылочного текста из-за незначительных изменений части документа, измененные разделы документа тестируются на важность внесенных изменений и только после этого дата ссылочного текста может быть изменена или нет.

81. Поисковая система производит (изменяет) оценку документа, по информации о том, как изменяется ссылочный текст с течением времени.

Трафик.

82. Информация о трафике документа может использоваться для оценки документа. Например, поисковая система 125 прослеживает использование документа пользователями за определенный период времени. Сильное снижение трафика может означать то, что документ устарел (не модернизируется или заменен на другой).

83. Поисковая система 125 сравнивает среднестатистический трафик документа за последние j количество дней (где $j=30$) с лучшим трафиком в течение месяца, учитывая сезонные изменения, или в течение последних k дней (где $k=365$). Дополнительно поисковая система 125 отмечает повторяющую модель трафика или изменения модели в течение времени. Это позволяет определять периоды, когда документ пользовался большей или меньшей популярностью. Сверяя эти модели, поисковая система 125 точно оценивает документ вне зависимости от внешних факторов.

84. Дополнительно, поисковая система 125 рассматривает временные характеристики «рекламного трафика» определенных документов. Например, поисковая система 125 отслеживает следующие факторы: (1) позиции рекламы, представляемой документом; (2) качество рекламируемого товара (документ, реклама которого ссылается на известные поисковой системе документы, обладает высоким трафиком); (3) насколько реклама повлияла на посещаемость документа. Поисковая система использует временные характеристики рекламного трафика для оценки документа.

85. Поисковая система 125 оценивает документ на основании информации о временно трафике документа.

Поведение пользователя.

86. Для согласованного внедрения принципов изобретения установление (или изменения) оценки документа может быть произведено на основе информации о поведении одного или множества пользователей, запрашивающих документ в течение определенного промежутка времени. Например, поисковая система 125 отмечает количество раз включения документа в результаты поиска и/или сколько времени один или более пользователей проводят на странице документа. Как минимум, по этой информации поисковая система 125 может частично оценить документ

87. Выполнение запроса определенного документа требует времени, определение релевантности документа запросу займет некоторое время у пользователя. Это один из способов определить, актуален контент или нет.

88. Поисковая система способна оценить документ, основываясь и на поведении пользователя по отношению к документу.

Информация о домене.

89. В соответствии с реализацией соответствующих принципов изобретения, информация о домене, связанная с документом, может использоваться для оценки документа. Например, поисковая система 125 следит за информацией о том, как документ функционирует в компьютерной сети, и использует ее для оценки документа.

90. Пользователи, пытающиеся обмануть поисковую систему, часто используют дорвеи для увеличения трафика своего сайта. Информация о законности доменов используется поисковыми системами для оценки документа.

91. Отличить законный домен от незаконного позволяют определенные сигналы. Например, домены обновляют в течение 10 лет. Законные (важные) домены обычно оплачиваются за несколько лет вперед. Тогда как домены-дорвеи (незаконные) используются чуть больше года. Этот фактор может быть использован для определения законности домена, следовательно, и документа.

92. По записи доменного имени сервера (DNS) также можно определить законен или нет домен. Запись DNS содержит информацию о том, кто регистрировал домен, административный и технический адреса, адреса именных серверов. Изучив эту информацию можно определить незаконный домен. Например, поисковая система 125 отслеживает информацию о верном адресе, о том, как часто изменяется контактная информация домена и т.д. Список неточной контактной информации, именных серверов, IP адреса идентифицируется, хранится и используется для определения законности домена.

93. Для определения законности домена используется информация о возрасте и связанная с именованием сервером. У «хорошего» именованного сервера может быть различные домены от различных регистраторов и с различной историей хостинга. У «плохого», главным образом, порнографические домены и домены-дорвеи с коммерческими словами (общепринятый спам), все домены от одного регистратора. Новизна именованного сервера еще не признак незаконности домена, но в совокупности с другими факторами, описанными выше, может им быть.

94. Поисковая система 125 оценивает (изменяет оценку) документ по информации о законности домена, связанного с документом.

История ранжирования.

95. Информация о предыдущих позициях документа используется для оценки документа. Например, поисковая система 125 отслеживает позиции документа в разный период времени, в соответствии с поступающими поисковыми запросами.

96. По количеству и качеству изменений позиций документа можно определить будущую оценку документа. Для каждого набора поисковых результатов, документ может быть взвешен в соответствии с его позицией в N результатах поиска. Например, для N= 30 функция выглядит так $[(N+1)-\text{SLOT}]/N$, доп.4. в этом случае лучший результат может быть оценен на 1.0, приближаясь к 0 для N-ого результата поиска.

97. Набор запросов (например, коммерческие запросы) может повторяться и документы, выбранные более чем в M% результатах поиска, могут ослабнуть или процентно возрасти; для ранжирования это сигнал к установлению ценности документа. Например, поисковая система 125 определяет, насколько высока средняя оценка лучших результатов и есть ли значительные изменения

в топ результатах от месяца к месяцу. Для коммерческих запросов вероятность спама выше, поэтому поисковая система тщательнее обрабатывает их.

98. В дополнении к вышесказанному, поисковая система контролирует один или несколько факторов, таких как число запросов, позицию (понижающуюся, повышающуюся), документ, выбранный как поисковый результат; сезонность, данные, использующие полосу канала лишь эпизодически, и другие характеристики выбранного документа; изменения оценки пары URL- запрос.

99. Поисковая система 125 следит за числом документов (например, URL), не основываясь на критерии запроса. Например, поисковая система 125 следит за средним баллом топ результатов поиска и оценивает их релевантность. Учитывает число ссылок на определенный запрос. Заметив их увеличение, поисковая система будет оценивать данные результаты выше.

100. Поисковая система следит за позициями документов, выявляя внезапные улучшения позиции документов. Подобные изменения могут означать либо обычное явление, либо попытку обмануть поисковую систему, например, покупая или продавая ссылки. Поисковая система принимает необходимые меры по борьбе с попытками спама, например, используя ограничения для улучшения определенных позиций. Либо позволяет оставаться документу на определенной позиции в течение определенного времени. Для дальнейшей статистики, чтобы отличать обычный документ от спамового, поисковая система 125 полагается на упоминания о документе в новостных статьях, дискуссионных группах, тогда как (в теории) спамовые документы не упоминаются, например, в новостях. Эти методы могут быть использованы для снижения количества попыток обмануть поисковую систему.

101. Для документов, заслуживающих доверия, таких как правительственные, web-каталоги и документы, имевшие стабильно высокие позиции, поисковая система 125 будет делать исключения. Если у подобного документа будут замечены резкие улучшения позиции и количества ссылок, он не будет рассматриваться как спам.

102. Резкое снижение позиций может быть расценено поисковой системой 125 как потеря популярности данного документа. Например, документ значительно сдал свои позиции, из этого следует, что он стал менее интересным пользователям, поисковая система оценивает его уже ниже.

103. Поисковая система может оценивать документ, основываясь лишь на данных о ранее занимаемых им позициях.

Заданные/Выработанные пользователями данные.

104. Заданные/выработанные пользователем данные также являются критерием оценки документа. Например, по «закладкам», favourites можно судить о том, насколько определенный документ интересен пользователям. Поисковая система получает подобную информацию прямым (с помощью вспомогательного браузера) или косвенным (с помощью браузера) путем. Анализируя количество закладок/favourites, определяет важность документа.

105. Поисковая система учитывает динамику появлений документа в списках закладок или favourites пользователей. Если документ часто фигурирует в подобных списках, это говорит о его важности и наоборот. В соответствии с этим поисковая система дает свою оценку документу.

106. Поисковая система 125 использует для оценки документа и другую информацию, указывающую на предпочтения пользователей. Например, данные буферных файлов или cookie-файлов.

107. Одним из критериев оценки для поисковой системы служат заданные/выработанные пользователем данные.

Уникальные слова, биграммы, фразы в ссылочном тексте.

108. На основании уникальных слов, биграмм, фраз в ссылочном тексте осуществляется оценка соответствующего документа. Поисковая система 125 учитывает данные ссылочных диаграмм и изменения в них при оценке документа, выявлении спама и др. целей. Обычно данные реальных диаграмм объективны и отражают действительность. Искусственно созданные веб-диаграммы, обычно определяющие спам, основаны на согласованных решениях, причиной которых является высокая плотность уникальных слов, фраз, биграмм, насколько можно ожидать релевантных.

109. Причиной подобной плотности может стать большое количество идентичных ссылок со многих документов, иногда созданных специально. Поисковая система 125 следит за ссылками и учитывает их при оценке документа. Если поисковая система рассматривает данную ссылку как спам, пострадает оценка соответствующего документа.

110. Одним из критериев оценки документа для поисковой системы является информация об уникальных словах, диграфах, фразах соответствующего текста.

Связь независимых одноранговых узлов.

111. Информация о независимых одноранговых узлах используется для оценки документа.

112. Резкое увеличение числа независимых одноранговых узлов с исходящими и входящими ссылками на документ может означать попытку спама и иметь отношение к искусственным веб-диаграммам. Необычно согласованный или несвязный ссылочный текст может усилить подозрения в спаме. Подобная информация может привести к снижению веса ссылки.

113. Одним из критериев оценки документа для поисковой системы 125 является информация о связи независимых одноранговых узлов.

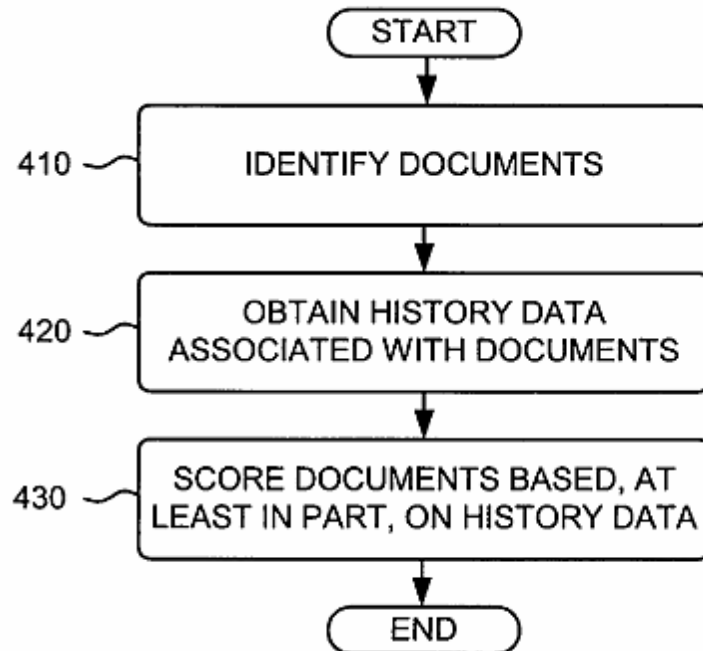
Разделы документа

114. Для оценки документа поисковая система использует информацию о разделах документа. Например, поисковая система может провести выборку разделов (по категориальным признакам, анализу URL, анализу контента, по выделению блоков одной тематики, резюмированию, по использованию уникальных, но редко встречаемых, слов и др.). Поисковая система 125 может контролировать раздел(ы) документа определенный промежуток времени и использовать эту информацию с целью оценки.

115. Значительные изменения в разделах документа могут означать, что сменился владелец документа и прежняя информация о документе (оценка, ссылочный текст) больше неверна. Исчезновение одного из разделов документа может также повлиять на оценку документа.

116. Одним из критериев оценки документа для поисковой системы являются изменения в одном или нескольких разделах документа.

ПРИМЕРНАЯ ОБРАБОТКА



117. Рис.4 это графическое представление примерной обработки данных по оценке документов, на основе реализации соответствующих принципов изобретения. Начать обработку можно с сервера 120, идентифицирующего документы (акт 410). Документы могут состоять из одного или нескольких документов, относящихся к запросам поиска, например, документы релевантные запросу. Или наоборот, документы включают один или несколько документов из единой базы данных, документы, на которые не влияет запрос поиска.

118. Поисковая система может получить хронологические данные, имеющие отношение к идентификации документов (акт 420). Выше было отмечено, что хронологические данные могут иметь разный вид. Например, в хронологические данные могут входить документы, касающиеся даты создания, изменения контента, анализа запроса, ссылочного текста, трафика, поведения пользователя и т.д. Поисковая система 125 может получить один из видов перечисленной информации, либо их комбинацию.

119. Поисковая система 125 может оценить документ на основе хронологических данных (акт 430). Если идентифицированный документ связан с запросом, поисковая система 125 может оценить важность документов, основанных на их релевантности запросу. Чтобы получить общую оценку

документа, поисковая система объединяет оценки хронологии и релевантности. Но возможно также оценить документ, не учитывая оценку релевантности.

120. После идентификации документов, связанных с запросом, поисковая система 125 выдает результаты поиска на основании оцененных документов. Например, поисковая система 125 сортирует документы по их оценке. Затем делает ссылки на данный документ, которые могут состоять из заголовка текста и его фрагмента. Поисковая система 125 может сделать ссылки на документы, представляющие особую значимость для пользователя.

ЗАКЛЮЧЕНИЕ

121. С помощью систем и методов, согласующихся с принципами изобретения, на основе хронологических данных проводится оценка документов и формирование высоких результатов поиска.

122. Все примеры использования данного изобретения сопровождаются иллюстрациями и описанием, но не являются его единственно возможными реализациями. Доступны различные модификации и вариации использования изобретения. Например, определенный набор действий может быть по-разному использован в соответствии с принципами реализации изобретения.

123. Как было описано, сервер 120 представляет исчерпывающую информацию, связанную с обработкой данных (РИС.4). В соответствии с другими принципами реализации изобретения, информация о последовательности действий может быть представлена другими серверами, например, 130 и/или 140.

124. Следует отметить, что все представленные аспекты изобретения могут быть использованы по-разному в зависимости от программных средств сопровождения, встроенного программного обеспечения и аппаратного обеспечения.