



**ФЕДЕРАЛЬНАЯ СЛУЖБА
ПО ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ**

(12) ЗАЯВКА НА ИЗОБРЕТЕНИЕ

(21)(22) Заявка: 2012144649/08, 20.04.2011

Приоритет(ы):

(30) Конвенционный приоритет:
21.04.2010 US 12/764,676

(43) Дата публикации заявки: 27.04.2014 Бюл. № 12

(85) Дата начала рассмотрения заявки РСТ на
национальной фазе: 19.10.2012(86) Заявка РСТ:
US 2011/033306 (20.04.2011)(87) Публикация заявки РСТ:
WO 2011/133705 (27.10.2011)

Адрес для переписки:

129090, Москва, ул. Б. Спасская, 25, стр.3, ООО
"Юридическая фирма Городиский и Партнеры"

(71) Заявитель(и):

МАЙКРОСОФТ КОРПОРЕЙШН (US)

(72) Автор(ы):

**ФАКСМЭН Ариэль (US),
НГУЙЕН Хоа (US),
СИЛВА Джулиана Фрейр де Лима Э. (US),
ПАПАЗИЗОС Стелиос (US),
АГРАВАЛ Ракеш (US),
ЧЭНЬ Чжиминь (US),
КОЛАДЖОВАННИ Лоуренс Уилльям (US),
СИКЧИ Пракаш (US)****(54) СИНТЕЗ ПРОДУКТА ИЗ МНОЖЕСТВЕННЫХ ИСТОЧНИКОВ****(57) Формула изобретения**

1. Способ автоматического синтеза информации о продукте из множественных источников данных в интерактивный каталог, содержащий:

получение исторической информации, соответствующей существующему продукту, представленному в интерактивном каталоге, из множества источников исторических данных, в котором историческая информация содержит множество исторических пар атрибут-значение, и каждая историческая пара атрибут-значение содержит историческое имя атрибута и соответствующее историческое значение атрибута;

определение соответствия между атрибутами источника и каталога, содержащее определение соответствия между первым историческим именем атрибута, включенным в первую историческую схему продукта первого источника исторических данных, и первым атрибутом каталога, включенным в схему каталога интерактивного каталога, причем схема каталога содержит множество пар атрибут-значение каталога, а каждая пара атрибут-значение каталога содержит другое имя атрибута каталога и соответствующее значение атрибута каталога;

определение извлечения пар атрибут-значение на основе соответствия между атрибутами источника и каталога, содержащее для каждой исторической пары атрибут-значение, включенной в историческую информацию, определение существующей ассоциативной связи между существующим именем атрибута каталога и, по меньшей мере, частью первого исторического значения атрибута; и

сохранение существующей ассоциативной связи между существующим именем атрибута каталога и, по меньшей мере, частью первого исторического значения атрибута в схеме каталога,

причем получение поступающей и исторической информации из множества источников поступающих и исторических данных содержит получение неструктурированных данных и структурированных данных в множестве различных схем из множества источников поступающих и исторических данных.

2. Способ по п. 1, дополнительно содержащий:

получение поступающей информации, соответствующей, по меньшей мере, одному из: существующему продукту или новому продукту из множества источников поступающих данных, причем множество источников поступающих данных содержит, по меньшей мере, одно из: по меньшей мере, части множества источников исторических данных или, по меньшей мере, одного нового источника данных, причем поступающая информация содержит множество поступающих пар атрибут-значение, и каждая поступающая пара атрибут-значение содержит поступающее имя атрибута и соответствующее поступающее значение атрибута;

определение соответствия между атрибутами источников, содержащее определение соответствия между первым поступающим именем атрибута, включенным в первую поступающую схему продукта первого источника поступающих данных, и вторым поступающим именем атрибута, включенным во вторую поступающую схему продукта второго источника поступающих данных;

кластеризацию, на основе соответствия между атрибутами источников, поступающих пар атрибут-значение, так что каждый кластер соответствует другому продукту;

извлечение пар атрибут-значение из поступающей информации, содержащее:

получение конкретного поступающего значения атрибута, соответствующего конкретному кластеру, соответствующему конкретному существующему продукту или конкретному новому продукту интерактивного каталога;

определение, на основе известного извлечения пары атрибут-значение, новой ассоциативной связи между конкретным поступающим значением атрибута и одним из: конкретным существующим именем атрибута каталога или конкретным новым именем атрибута каталога; и

добавление новой ассоциативной связи к схеме каталога;

слияние поступающей информации, содержащее определение по новой ассоциативной связи и, по меньшей мере, одной другой ассоциативной связи, включенной в схему каталога, типичного значения атрибута каталога, соответствующего одному из: конкретному существующему или конкретному новому имени атрибута каталога; и

обновление, с помощью типичного значения атрибута каталога, записи в интерактивном каталоге, соответствующей конкретному существующему или конкретному новому продукту.

3. Способ по п. 2, в котором получение поступающей информации из множества источников поступающих данных содержит получение поступающей информации посредством, по меньшей мере, одного из: считывания веб-канала или извлечения объекта из веб-сайта из, по меньшей мере, одного из:

мнения согласно отзывам, рейтинга согласно отзывам, индикатора полезности согласно отзывам, часто задаваемых вопросов (FAQ), характеристик, соответствующих продуктам, вопроса и ответа, родственного продукта, видео или сигнала ценового прогноза.

4. Способ по п. 3, в котором получение поступающей информации дополнительно содержит получение поступающей информации, по меньшей мере, из одного из: предложения о продаже, данных, активизируемых щелчком мышью, изображения,

ранга продукта, технического описания, рекомендации, иерархии продукта, карты балльных оценок продуктов, индекса мнений или руководства по использованию продукта.

5. Способ по п. 2, дополнительно содержащий сегментацию множества поступающих имен атрибутов.

6. Способ по п. 2, дополнительно содержащий нормализацию поступающих имен атрибутов и поступающих значений атрибутов на основе схемы каталога и таксономии каталога, соответствующих интерактивному каталогу.

7. Способ по п. 2, дополнительно содержащий классификацию поступающей информации в одну или более категорий таксономии каталога, причем схема каталога содержит отличающуюся схему категории для каждой одной или более категорий.

8. Способ по п. 7, в котором кластеризация поступающих пар атрибут-значение содержит кластеризацию поступающих пар атрибут-значение дополнительно на основе одной или более категорий таксономии каталога.

9. Способ по п. 2, дополнительно содержащий очистку поступающей информации, содержащую исправление ошибок кодирования, преобразование букв в прописные и проверку правописания.

10. Способ по п. 2, дополнительно содержащий получение обновленной поступающей информации, по меньшей мере, из одного из: множества поступающих источников данных, по меньшей мере, два раза в день и получения обновленной поступающей информации из всех из множества источников данных, по меньшей мере, ежемесячно.

11. Система для автоматического синтеза информации о продукте в интерактивный каталог, содержащая одно или более вычислительных устройств, имеющих процессор, память, и компьютерно-исполняемые команды, хранимые в памяти и исполняемые процессором, содержит:

интерактивный каталог, содержащий:

таксономию каталога, содержащую множество категорий продуктов,

схему каталога, содержащую схему категорий для каждой категории продуктов, причем каждая схема категорий содержит набор пар атрибут-значение каталога, и каждая пара атрибут-значение каталога содержит имя атрибута каталога и соответствующее значение атрибута каталога, и

множество записей, соответствующих существующим продуктам, причем каждая из множества записей основана на таксономии каталога и содержит множество соответствующих пар атрибут-значение каталога, включенных в схему каталога;

компонент сбора данных с осуществляющим связь подключением к сети и выполненный с возможностью получения из множества источников поступающих электронных данных в сети, поступающей информации о продукте, соответствующей одному или более существующим продуктам, и поступающей информации о продукте, соответствующей одному или более новым продуктам, причем поступающая информация о продукте содержит поступающие схемы продуктов, каждая из которых имеет набор поступающих пар атрибут-значение, при этом каждая поступающая пара атрибут-значение имеет поступающее имя атрибута и соответствующее поступающее значение атрибута, и причем поступающая информация о продукте включает в себя мнение согласно отзывам, рейтинг согласно отзывам, индикатор полезности согласно отзывам, вопросы FAQ, характеристики, соответствующие продуктам, вопрос и ответ, родственный продукт, видео или сигнала ценового прогноза;

компонент вычисления, выполненный с возможностью осуществления операций над поступающей информацией о продукте и содержащий:

компонент связей объектов, выполненный с возможностью определения соответствия между атрибутами источников, содержащего определение соответствия между первым

поступающим именем атрибута, включенным в первую таксономию и первую схему, соответствующую первому источнику поступающих электронных данных, и вторым поступающим именем атрибута, включенным во вторую таксономию, и вторую схему, соответствующую второму источнику электронных данных, причем первый и второй источники поступающих электронных данных выбраны из множества источников поступающих электронных данных;

компонент кластеризации, выполненный с возможностью осуществления операций над поступающей информацией о продуктах и определения набора кластеров продуктов на основе поступающих имен атрибутов, поступающих значений атрибутов и соответствия между атрибутами источников, так что каждый кластер продукта соответствует отличающемуся продукту; и

компонент извлечения, выполненный с возможностью осуществления операций над набором кластеров продуктов для определения извлеченного соответствия между значением атрибута кластера и одним из: существующим именем атрибута каталога в схеме каталога или новым именем атрибута каталога для схемы каталога;

компонент слияния, выполненный с возможностью определения, на основе одного или более извлеченных соответствий, единственного типичного значения атрибута каталога для конкретного имени атрибута каталога, включенного в схему каталога; и

издательский компонент с подключением к интерактивному каталогу и выполненный с возможностью включения единственного типичного значения атрибута каталога в запись в интерактивном каталоге.

12. Система по п. 11, дополнительно содержащая:

компонент нормализации, выполненный с возможностью осуществления операций над поступающей информацией о продуктах и определения соответствия между первым особым именем атрибута, включенным в схему не каталога в составе поступающей информации о продукте, и первым особым именем атрибута каталога, включенным в схему каталога; и

компонент классификации, выполненный с возможностью осуществления операций над поступающей информации о продукте и обеспечения ассоциативной связи между конкретным новым продуктом или конкретным существующим продуктом, представленным в поступающей информации о продукте, по меньшей мере, с одной категорией в таксономии каталога, причем

компонент связей объектов дополнительно выполнен с возможностью определения соответствия между атрибутами источника и каталога, содержащего определение соответствия между каждым особым именем атрибута первой схемы не каталога в составе поступающей информации о продукте и соответствующим именем атрибута каталога, включенным в схему каталога, а

компонент кластеризации определяет набор кластеров продуктов дополнительно на основе категорий таксономии каталога.

13. Система по п. 11, в которой компонент сбора данных включает в себя экстрактор объектов и считывателя веб-каналов.

14. Система по п. 13, дополнительно содержащая объявленный API (интерфейс прикладных программ), предоставляющий доступ, по меньшей мере, к одному из: экстрактору объектов или считывателю веб-каналов.

15. Система по п. 11, дополнительно содержащая компонент ввода данных человеком, соединенный, по меньшей мере, с одним из: компонентом сбора данных, компонентом вычислений или интерактивным каталогом, причем компонент ввода данных человеком включает в себя, по меньшей мере, одно из: модель или правило.