

СТАНОВИЩЕ

ОТНОСНО: проучване за новост на " Система за извличане на релевантна информация, приоритизацията и верификацията ѝ за решаване на правни казуси"

При-142/28.09.2020

Към искането за проучване са приложени описание, чертеж и претенции.

Формулирани са една независима и две зависима претенции.

Съгласно патентните 1-3 претенции проучването се отнася до система за извличане на релевантна информация, приоритизацията и верификацията ѝ за решаване на правни казуси, включваща блок с нормативна база, блок съдебна практика, блок за задания, блок за професионални юридически публикации, блок за масмедийни юридически публикации, приоритизиращ блок, верификационен блок, интегратор за верифицирани резултати, подредени по приоритет. Блокове нормативна база, съдебна практика, професионални юридически публикации, масмедийни юридически публикации, приоритизиращ и верификационен са с електронна база данни и вградени процесори.

Задача на изобретението е да създаде система, радикално преодоляваща недостатъците на известните технически решения, да е с ниска себестойност и с ниски експлоатационни разходи, осигуряваща улеснено систематизиране на информация за юридически квалифициран персонал за достигане на правно обосновано юридическо решение.

При извършеното проучване се установи наличие на известни от предшестващото състояние на техниката документи, отразяващи нивото на техниката, които са посочени в приложения доклад от проучване и които не са противопоставими по новост на обекта на проучване, а именно:

- **TW202001781 (A)** публикуван **01.01.2020 г.**, който разкрива система за анализ на големи данни, прогнозиране и визуализация на правни документи, при въвеждане на информация, изобретението ще анализира статистически всички съответни закони, наредби и свързани решения, след което ще покаже резултата в проста таблица, като ще спомогне чрез прогнозиране на съдебни насоки и резултати да се спести време и пари.

- **CN101950310 (A)** публикуван **19.01.2011 г.**, който разкрива система за нормализиране на присъдите, включваща система база данни състояща се от бази данни за присъди, за алгоритми, за дела, за съдебни тълкувания, за закони и наредби и нормиране на присъдите.

- **CN1707487 (A)** публикуван **14.12.2005 г.**, който разкрива метод на изкуствен интелект при анализ на различни видове правна информация, след като потребителя въведе относителна информация за факти чрез специален хардуер и софтуер, разсъждаващият компютър изисква необходимата информация в библиотеката на знанията, завършва анализа и преценката в правилата за препратки, които програмата определя и връща резултата в специална фирма на потребителя

НОВОСТ

В нито един от намерените при проучването документи не е разкрито техническо решение за система за извличане на релевантна информация, приоритизацията и верификацията ѝ за решаване на правни казуси, което да съдържа всички технически признаци на обекта на проучването.

Експертната група счита, че обектът на проучването **отговаря на изискванията за новост, съгласно чл. 8 от Закона за патентите и регистрацията на полезните модели (ЗПРПМ).**

05.11.2020 г.

МЛАДШИ ЕКСПЕРТ:

(Д. Петрова - Иванова)

ДЪРЖАВЕН ЕКСПЕРТ:





ПАТЕНТНО ВЕДОМСТВО НА РЕПУБЛИКА БЪЛГАРИЯ

ДОКЛАД ОТ ПРОУЧВАНЕ

ПРОУЧВАНЕ НОВОСТ ВХ. № При-142/28.09.2020 " Система за извличане на релевантна информация, приоритизацията и верификацията ѝ за решаване на правни казуси"

I. Уточнен класификационен индекс по МПК 8

G06Q10/10; G06Q50/18

II. ОБЛАСТ НА ПРОУЧВАНЕ:

Минимален обхват на проучване

Вид класификация:

МПК - 8

Класификационен индекс:

G06Q10/10; G06Q50/18

Бази данни, в които е извършено проучването

IPAS, esp@cenet; epoquenet

III. Релевантни документи

Категория	Цитиране на документите с посочване на релевантните части	Релевантни към претенция
A	TW202001781 (A) публ. 01.01.2020 г. LIN WEI LING [TW]	1-3
A	CN101950310 (A) публ. 19.01.2011 г. BEIJING BOYA YINGJIE TECHNOLOGY CO LTD	1-3
A	CN1707487 (A) публ. 14.12.2005 г. YUAN YUYU [CN]	1-3

Категория на документа

X – документът е от особено значение; претендираното изобретение няма новост или изобретателска стъпка, когато документът се разглежда самостоятелно
A – документ, отразяващ нивото на техниката, който не е от особено значение

Y – документът е от особено значение; претендираното изобретение няма изобретателска стъпка, когато документът се комбинира с един или повече други документи и тази комбинация е очевидна за специалист в областта
D – документ, цитиран в описанието

Дата: 05.11.2020

Име: Д. Петрова- Иванова

Подпис:

Дирекция "Експертиза и закрила на изобретенията и промишления дизайн"





(19)中華民國智慧財產局

(12)發明說明書公開本

(11)公開編號：TW 202001781 A

(43)公開日：中華民國 109 (2020) 年 01 月 01 日

(21)申請案號：107120071

(22)申請日：中華民國 107 (2018) 年 06 月 11 日

(51)Int. Cl. : G06Q50/18 (2012.01)

(71)申請人：林威伶(中華民國) LIN, WEI LING (TW)

新竹縣竹北市自強南路 416 號

(72)發明人：林威伶 LIN, WEI LING (TW)

申請實體審查：有 申請專利範圍項數：4 項 圖式數：4 共 12 頁

(54)名稱

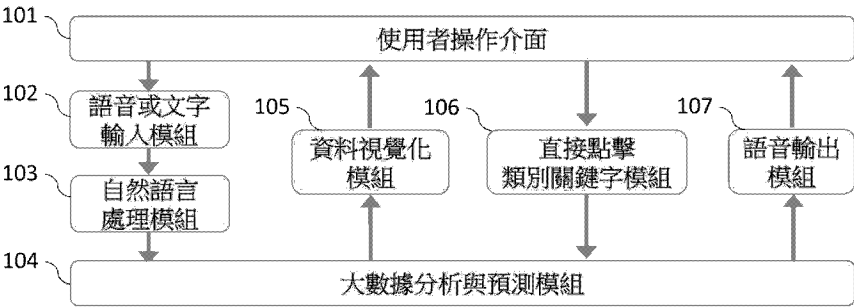
法律文件的大數據分析、預測、資料視覺化系統

(57)摘要

法律用語的艱澀難懂足以使一般民眾望而生怯，並阻礙法治教育的推廣普及，使法律無法「平民化」。而在民眾遇到法律糾紛時，法官與律師需要去搜尋法律判決書資料庫，了解過去遇到同類問題時，前人的經驗是如何處理的。但人類無法在有限的時間內，閱讀完大量的判決書，目前法律實務的作法是使用有限的關鍵字，來搜尋判決書，關鍵字的使用需要專業與技巧，更需要經驗的累積，這對一般民眾而言，望塵莫及。本發明提出法律文件的大數據分析、預測、資料視覺化系統。利用人工智慧大量搜尋與學習法律、法規、行政命令等，甚至千萬多筆的判決書資料，當非法律專業背景的使用者輸入所需資料，本發明將統計分析所有類似案件的相關法規與審判結果，以簡單的圖像表格顯示出來。本發明目的不是要取代法界人士的工作，而是讓一般民眾透過查找所有相關法令、類似案件及訴訟結果，來更了解台灣與大陸海峽兩岸的法律與法院實務判決，預測可能的訴訟方向與結果，讓民眾決定是否興訟，節省大量法律資源。此發明可供律師事務所、法院、甚至所有民眾使用，把法律資料視覺化，加速全民法律的擴散。並配合現代智慧型手機的普及與 APP 應用的蓬勃發展，發展對應的應用程式平台。

Legal language is difficult for the general public to understand and make popular education unreachable. When legal disputes occur, the judges and attorneys need to search the database of legal judgments to obtain the experience of dispute decision. However, human beings cannot read a large number of judgments within a limited period of time. It is also impossible for the general public to use limited keywords to do legal research. Since the search of verdicts and keywords requires professional skills and experience. This invention proposes a big data analysis, prediction, and graph display system for legal documents. We use artificial intelligence to search and study a large number of laws, regulations, orders, and even more than ten million judgments, and etc. When the general public users input information, the invention will statistically analyze all relevant laws, regulations and related judgments. This invention will show the result in a simple image table. This invention is not to replace the work of the legal profession, but to help the general public to know more about the laws, regulations and legal practice in Taiwan or Mainland China. The general public can save time and money through predicting possible litigation directions and results. People can decide whether to initiate litigation more easily. This invention can be used by law firms, courts, and even all citizens. Visualization of the legal data makes laws easy to understand and accelerates the proliferation of legal knowledge. This invention also provides the corresponding application platform in modern mobile device.

指定代表圖：



【圖1】

符號簡單說明：

101~107 . . . 係為本發明系統之功能圖

【發明說明書】

【中文發明名稱】

法律文件的大數據分析、預測、資料視覺化系統

【英文發明名稱】

BIG DATA ANALYSIS, PREDICTION, AND DATA VISUALIZATION SYSTEM
FOR LEGAL DOCUMENTS.

【技術領域】

- 【0001】 法律
- 【0002】 人工智慧
- 【0003】 資料視覺化

【先前技術】

- 【0004】 公開編號：200743969，統計資料的圖形顯示方法
- 【0005】 申請案號：104141828，資料視覺化方法與資料視覺化裝置
- 【0006】 申請案號：101103701，量刑決策支援系統
- 【0007】 申請案號：105142810，法學資料搜尋系統、搜尋處理方法及電
腦可讀取紀錄媒體
- 【0008】 申請案號：106211023，一種具有類神經模糊網路的法學知識學
習資料庫系統

【發明內容】

【0009】 法律用語的艱澀難懂足以使一般民眾望而生怯，並阻礙法治教育的推廣普及，使法律無法「平民化」。而在民眾遇到法律糾紛時，法官與律師需要去搜尋法律判決書資料庫，了解過去遇到同類問題時，前人的經驗是如何處理的。但人類無法在有限的時間內，閱讀完大量的判決書，目前法律實務的作法是使用有限的關鍵字，來搜尋判決書，關鍵字的使用需要專業與技巧，更需要經驗的累積，這對一般民眾而言，望塵莫及。

【0010】 本發明提出法律文件的大數據分析、預測、資料視覺化系統。利用人工智慧大量搜尋與學習法律、法規、行政命令等，甚至千萬多筆的判決書資料，當非法律專業背景的使用者輸入所需資料，本發明將統計分析所有類似案件的相關法規與審判結果，以簡單的圖像表格顯示出來。

【0011】 本發明目的不是要取代法界人士的工作，而是讓一般民眾透過查找所有相關法令、類似案件及訴訟結果，來更了解台灣與大陸海峽兩岸的法律與法院實務判決，預測可能的訴訟方向與結果，讓民眾決定是否興訟，節省大量法律資源。

【0012】 本發明可供律師事務所、法院、甚至所有民眾都可使用，把法律資料視覺化，加速全民法律的擴散。並配合現代智慧型手機的普及與APP應用的蓬勃發展，發展對應的應用程式平台。

【圖式簡單說明】

【0013】 第一圖係為本發明系統之功能圖

【0014】 第二圖係為本發明系統之範例圖表

【0015】 第三圖係為本發明系統之範例圖表

【0016】 第四圖係為本發明系統之範例圖表

【實施方式】

【0017】 這裡將說明一些典型實施例，其中的說明在本質上係當作說明之用，而非架構於限制本發明。

【0018】 本發明提出使用者操作介面，不識字的使用者可以選用麥克風唸出欲請求的資訊，或用文字輸入來做「以文找文」的查詢。例如唸出「故意殺人罪在苗栗地方法院會判多久的有期徒刑?」

【0019】 當使用者以文字或語音方式輸入請求資訊後，自然語言處理模組將分析出關鍵字及意圖資料，以了解使用者的需求，並輸出給一大數據分析與預測模組。例如分析出「苗栗地方法院」的關鍵字、「判多久」的意圖資料。

【0020】 該大數據分析與預測模組進行全文檢索，搜尋歷年訴訟審判資料後，統計所有類似案件的審判結果，並進行分類。例如以「苗栗地方法院」的關鍵字搜尋後，統計並分類出「殺人罪 10%」、「毒品罪 21%」、「偷竊罪 37%」。即可輸出到圖表顯示模組：分析結果圖表(如圖二)、與類別關鍵字「殺人」、「毒品」、「偷竊」。

【0021】 該使用者操作介面顯示上述的分析結果圖表(如圖二)，使用者可以直接點擊任意的類別關鍵字，此類別關鍵字將輸出給該大數據分析與預測模組，再次進行全文檢索，統計所有類似案件的審判

結果，並進行分類。例如直接點擊「殺人罪」的關鍵字後，該大數據分析與預測模組統計並分類出「無罪3%」、「0-2年 7%」、「2-7年 21%」…。即可輸出到圖表顯示模組：分析結果圖表(如圖三)、與類別關鍵字「無罪」、「0-2年 7%」、「2-7年 21%」…。

【0022】該使用者操作介面顯示上述的分析結果圖表(如圖三)，使用者可以看到在二個類別關鍵字下，7-10年有期徒刑的機率最大。或可以回到上一個類別關鍵字重新點擊並重新全文檢索。

【0023】該使用者操作介面也可顯示任意的分析結果圖表(如圖四)，使用者可直接點擊機率最大的「7-10年有期徒刑」類別關鍵字，進一步顯示出該分類的案件內容供使用者查詢。

【0024】該使用者操作介面也可以讓不識字的使用者聆聽使用者操作介面的資訊。

【符號說明】

- 101 一使用者操作介面，包含至少讓使用者輸入的介面、顯示的介面、與點擊的介面。
- 102 一語音或文字輸入模組，讓使用者以文字或語音方式輸入欲查詢的資料；
- 103 一自然語言處理模組，用以將使用者的一請求資訊，分析成至少一可能意圖語法資料，每一可能意圖語法資料包括至少一關鍵字及一意圖資料；
- 104 一大數據分析與預測模組，利用自然語言處理模組輸出的一關鍵字及一意圖資料，本大數據分析與預測模組將搜尋與統計歷年訴

訟審判資料後，輸出包括至少一分析結果，一預測結果，或一建議的類別關鍵字；

- 105 一資料視覺化模組，把該大數據分析與預測模組的輸出，包括至少分析資料、預測資料、與建議的類別關鍵字，以圖像表格方式顯示到該使用者操作介面。
- 106 一直接點擊類別關鍵字模組，讓使用者可以直接點擊使用者操作介面建議的類別關鍵字，輸出給該大數據分析與預測模組。
- 107 一語音輸出模組，讓不識字的使用者可以聆聽使用者操作介面的資訊。



202001781

【發明摘要】

【中文發明名稱】

法律文件的大數據分析、預測、資料視覺化系統

【英文發明名稱】

BIG DATA ANALYSIS, PREDICTION, AND DATA VISUALIZATION SYSTEM FOR
LEGAL DOCUMENTS.

【中文】

法律用語的艱澀難懂足以使一般民眾望而生怯，並阻礙法治教育的推廣普及，使法律無法「平民化」。而在民眾遇到法律糾紛時，法官與律師需要去搜尋法律判決書資料庫，了解過去遇到同類問題時，前人的經驗是如何處理的。但人類無法在有限的時間內，閱讀完大量的判決書，目前法律實務的作法是使用有限的關鍵字，來搜尋判決書，關鍵字的使用需要專業與技巧，更需要經驗的累積，這對一般民眾而言，望塵莫及。

本發明提出法律文件的大數據分析、預測、資料視覺化系統。利用人工智慧大量搜尋與學習法律、法規、行政命令等，甚至千萬多筆的判決書資料，當非法律專業背景的使用者輸入所需資料，本發明將統計分析所有類似案件的相關法規與審判結果，以簡單的圖像表格顯示出來。

本發明目的不是要取代法界人士的工作，而是讓一般民眾透過查找所有相關法令、類似案件及訴訟結果，來更了解台灣與大陸海峽兩岸的法律與法院實務判決，預測可能的訴訟方向與結果，讓民眾決定是否興訟，節省大量法律資源。

此發明可供律師事務所、法院、甚至所有民眾使用，把法律資料視覺化，加速全民法律的擴散。並配合現代智慧型手機的普及與APP應用的蓬勃發展，發展對應的應用程式平台。

【英文】

Legal language is difficult for the general public to understand and make popular education unreachable. When legal disputes occur, the judges and attorneys need to search the database of legal judgments to obtain the experience of dispute decision. However, human beings cannot read a large number of judgments within a limited period of time. It is also impossible for the general public to use limited keywords to do legal research. Since the search of verdicts and keywords requires professional skills and experience.

This invention proposes a big data analysis, prediction, and graph display system for legal documents. We use artificial intelligence to search and study a large number of laws, regulations, orders, and even more than ten million judgments, and etc. When the general public users input information, the invention will statistically analyze all relevant laws, regulations and related judgments. This invention will show the result in a simple image table.

This invention is not to replace the work of the legal profession, but to help the general public to know more about the laws, regulations and legal practice in Taiwan or Mainland China. The general public can save time and money through predicting possible litigation directions and results. People can decide whether to initiate litigation more easily.

This invention can be used by law firms, courts, and even all citizens. Visualization of the legal data makes laws easy to understand and accelerates the proliferation of legal knowledge. This invention also provides the corresponding application platform in modern mobile device.

【指定代表圖】

第一圖

【代表圖之符號簡單說明】

101~107：係為本發明系統之功能圖。

【發明申請專利範圍】

【第1項】

一種法律文件的大數據分析、預測、資料視覺化系統，包含

一使用者操作介面，以圖像表格方式顯示分析資料、預測資料，以及讓使用者可以直接點擊建議的類別關鍵字。

一大數據分析與預測模組，搜尋與統計歷年訴訟審判資料後，輸出包括但不限於一分析結果，一預測結果，或一建議的類別關鍵字；

一資料視覺化模組，把該大數據分析與預測模組的輸出，回傳到該使用者操作介面。

一直接點擊類別關鍵字模組，讓使用者可以在操作介面上直接點擊建議的類別關鍵字，輸出給該大數據分析與預測模組。

【第2項】

一種實現如申請專利範圍第1項之法律文件的大數據分析、預測、資料視覺化系統，

該使用者操作介面，讓使用者可以選用語音或文字輸入，及以圖像表格方式顯示分析資料、預測資料，以及直接點擊建議的類別關鍵字；

該大數據分析與預測模組，利用自然語言處理模組輸出的一關鍵字及一意圖資料，本大數據分析與預測模組將搜尋與統計歷年訴訟審判資料後，輸出包括但不限於一分析結果，一預測結果，或一建議的類別關鍵字；

更包括

一語音或文字輸入模組，讓使用者以文字或語音方式輸入欲查詢的資料；

一自然語言處理模組，用以將使用者的一請求資訊，分析成至少一可能意圖語法資料，每一可能意圖語法資料包括至少一關鍵字及一意圖資料；

【第3項】

一種實現如申請專利範圍第1項之法律文件的大數據分析、預測、資料視覺化系統，更包括

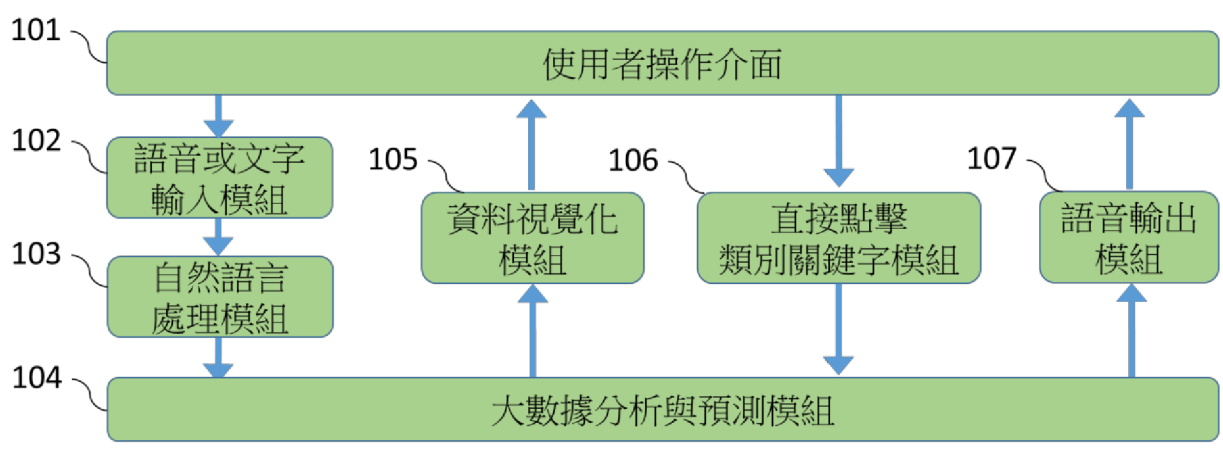
一語音輸出模組，讓不識字的使用者可以聆聽使用者操作介面的資訊。

【第4項】

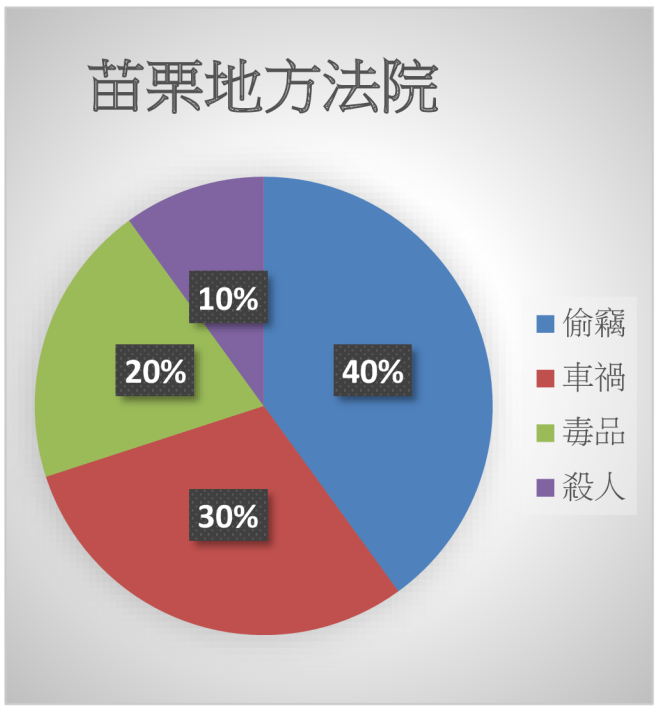
一種實現如申請專利範圍第1項之法律文件的大數據分析、預測、資料視覺化系統，更包括

一種行動電子裝置 APP 系統，可以將上述操作也實現在行動電子裝置，平板電腦，或個人電腦上。

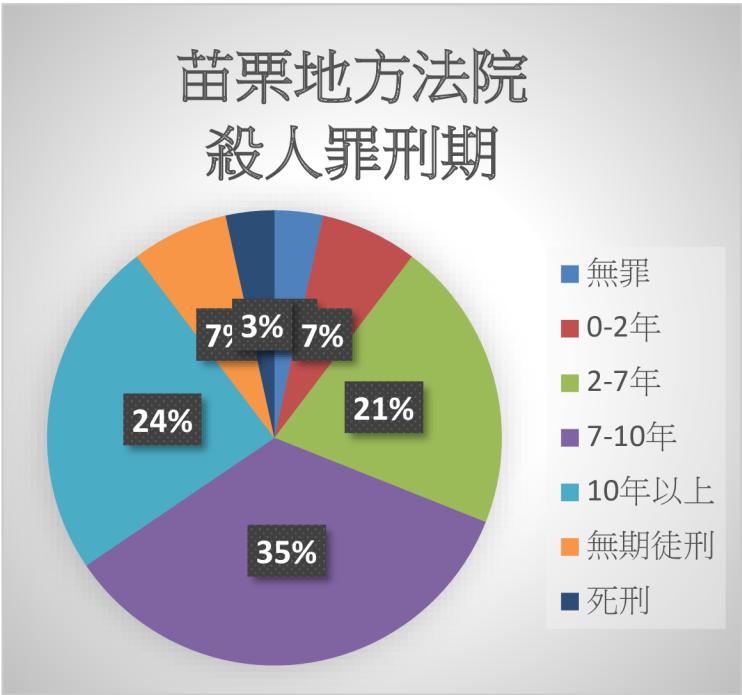
【發明圖式】



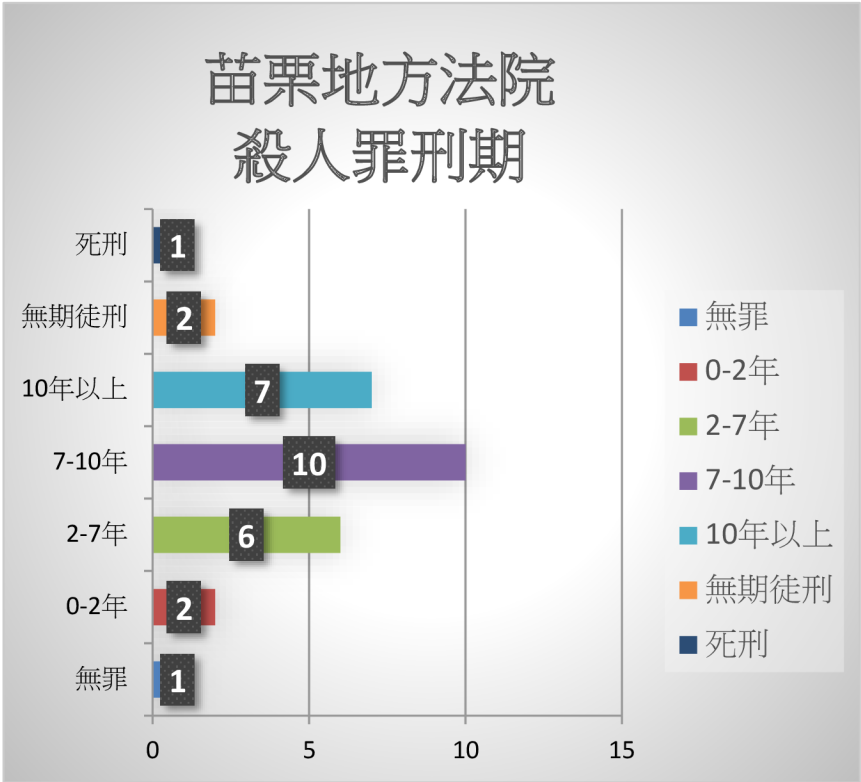
【圖1】



【圖2】



【圖3】



【圖4】

【發明說明書】

【中文發明名稱】

法律文件的大數據分析、預測、資料視覺化系統

【英文發明名稱】

BIG DATA ANALYSIS, PREDICTION, AND DATA VISUALIZATION SYSTEM
FOR LEGAL DOCUMENTS.

【技術領域】

【0001】 法律

【0002】 人工智慧

【0003】 資料視覺化

【先前技術】

【0004】 公開編號：200743969，統計資料的圖形顯示方法

【0005】 申請案號：104141828，資料視覺化方法與資料視覺化裝置

【0006】 申請案號：101103701，量刑決策支援系統

【0007】 申請案號：105142810，法學資料搜尋系統、搜尋處理方法及電
腦可讀取紀錄媒體

【0008】 申請案號：106211023，一種具有類神經模糊網路的法學知識學
習資料庫系統

【發明內容】

【0009】 法律用語的艱澀難懂足以使一般民眾望而生怯，並阻礙法治教育的推廣普及，使法律無法「平民化」。而在民眾遇到法律糾紛時，法官與律師需要去搜尋法律判決書資料庫，了解過去遇到同類問題時，前人的經驗是如何處理的。但人類無法在有限的時間內，閱讀完大量的判決書，目前法律實務的作法是使用有限的關鍵詞，來搜尋判決書，關鍵詞的使用需要專業與技巧，更需要經驗的累積，這對一般民眾而言，望塵莫及。

【0010】 本發明的目的是克服上述現有技術的缺點，提供了一種能夠實現判決書檢索分析和展示的大數據分析、預測、數據視覺化系統。為了實現上述目的，本發明具有如下技術：

【0011】 自然語言處理技術（Natural Language Processing），使用者可以直接把一句話或一段文章直接複製貼上到搜尋列，該自然語言處理技術把整段字串內容自動轉成多個關鍵詞，以便處理。而幾乎見於每篇文章的虛詞（虛詞如中文裡的一，的，你，我，他，一個，可是，所以）都不具獨特性的資訊，就可以省略。接下來該技術比較該多個關鍵詞與詞典中已儲存關鍵詞後，輸出相對應的回應字串至該使用者。但是文章有很多關鍵詞，需要大量計算，所以該技術通常只搜尋較小的資料集（例如學術論文或法院判決書），而不搜尋整個網絡。其常用的領域包括問答系統（Question Answering）：使用者不需要思考該使用什麼樣的問法才能夠得到理想的答案，只需要用口語化的方式直

接提問如“請問誰是美國總統？”即可。而該技術在了解使用者問句後，會非常清楚地回答“川普是美國總統”。

【0012】協同過濾技術（Collaborative Filtering），計算過去多個使用者的歷史搜尋與本次使用者搜尋的相似度，透過過去最相似的歷史搜尋的統計，找出更進一步可能的歷史搜尋，作為本次使用者輸入下一個關鍵詞的推薦。

【0013】資訊擷取技術（Information Extraction），搜尋與統計歷年訴訟審判資料，從大量文獻數據中過濾出較為相關的文字片段，分析文句語法（例如找出動詞、主詞、或受詞），再提取出特定事件包括人（Who）、事（What）、地（where）與時間（When）等事實（Fact），並獲得加總後的統計數據。當使用者要搜尋特定人、事、地、與時間等事實條件時，本發明即自動比對歷史統計數據，並輸出使用者特定條件下的分析或預測數據。

【0014】本發明提出法律文件的大數據分析、預測、資料視覺化系統。利用人工智慧大量搜尋與學習法律、法規、行政命令等，甚至千萬多筆的判決書資料，當非法律專業背景的使用者輸入所需資料，本發明將統計分析所有類似案件的相關法規與審判數據，以簡單的圖像表格顯示出來。

【0015】本發明目的不是要取代法界人士的工作，而是讓一般民眾透過查找所有相關法令、類似案件及訴訟數據，來更了解台灣的法

律與法院實務判決，預測可能的訴訟方向與結果，讓民眾決定是否興訟，節省大量法律資源。

【0016】本發明可供律師事務所、法院、甚至所有民眾都可使用，把法律資料視覺化，加速全民法律的擴散。並配合現代智慧型手機的普及與APP應用的蓬勃發展，發展對應的應用程式平台。

【圖式簡單說明】

【0017】第一圖係本發明之系統功能圖

【0018】第二圖係本發明實施例之第一範例圖表

【0019】第三圖係本發明實施例之第二範例圖表

【0020】第四圖係本發明實施例之第三範例圖表

【0021】第五圖係本發明實施例之第四範例圖表

【實施方式】

【0022】這裡將說明一些典型實施例，其中的說明在本質上係當作說明之用，而非架構於限制本發明。

【0023】本發明提出使用者操作介面，不識字或有需要的使用者可以選用麥克風唸出、或用文章方式輸入欲查詢的資料。例如唸出「殺人罪在台北地方法院會判多久的有期徒刑?」

【0024】當使用者以文章或語音方式輸入請求資訊後，自然語言處理模組將分析出關鍵詞，並輸出給該大數據分析與預測模組。例如分析出「台北地方法院」的關鍵詞、「判多久」的意圖資料。

- 【0025】該大數據分析與預測模組利用資訊擷取技術搜尋歷年訴訟審判資料後，統計所有類似案件的審判數據，並進行分類。例如以「台北地方法院」的關鍵詞搜尋後，統計並分類出「殺人罪10%」、「毒品罪21%」、「偷竊罪37%」。即可輸出到圖表顯示模組：分析數據圖表(如圖二)、與類別關鍵詞「殺人」、「毒品」、「偷竊」。
- 【0026】該使用者操作介面顯示上述的分析數據圖表(如圖二)，使用者可以直接點擊任意的類別關鍵詞，此類別關鍵詞將輸出給該大數據分析與預測模組，再次搜尋歷年訴訟審判資料，統計所有類似案件的審判數據，並進行分類。例如直接點擊「殺人罪」的關鍵詞後，該大數據分析與預測模組統計並分類出「無罪3%」、「0-2年 7%」、「2-7年 21%」…。即可輸出到圖表顯示模組：分析數據圖表(如圖三)、與類別關鍵詞「無罪」、「0-2年 7%」、「2-7年 21%」…。
- 【0027】該使用者操作介面顯示上述的分析數據圖表(如圖三)，使用者可以看到在該二個類別關鍵詞下，7-10年有期徒刑的機率最大。或可以回到上一個類別關鍵詞重新點擊，並再次統計所有類似案件的審判數據。
- 【0028】該使用者操作介面也可顯示任意的分析數據圖表(如圖四)，使用者可直接點擊機率最大的「7-10年有期徒刑」類別關鍵詞，進一步顯示出該分類的案件內容供使用者查詢。
- 【0029】該使用者操作介面也可以讓不識字或有需要的使用者聆聽使用者操作介面的資訊。

【0030】該自然語言處理模組，係配合現代智慧型手機的普及與APP應用的蓬勃發展，發展對應的應用程式平台（QQ、微信、Line等，如圖5所示），當使用者透過語音或文章輸入模組輸入一自然語言字串，例如輸入“我朋友在台北不小心撞到人了，該賠多少錢？”，該自然語言處理模組將比較該自然語言字串與詞典中已儲存字串，如“台北”、“撞”、“賠多少錢”。該自然語言處理模組將輸出相對應的字串，如“你朋友有喝酒嗎？”。當使用者再次輸入一自然語言字串，如“當然沒有喝酒”，該自然語言處理模組將比較該自然語言字串與詞典中已儲存字串，如“沒”、“酒”。該自然語言處理模組將輸出相對應的字串，如“你朋友有闖紅燈嗎？”。透過一直提問得到足夠的資訊量後，該自然語言處理模組即輸出相對應的字串至該使用者操作介面，如“好，我推薦三位台北的律師給您：林某...王某...陳某...”；或輸出該大數據分析與預測模組的分析與預測數據，如“賠償金額的圖像表格圖”。

【符號說明】

- 101 一使用者操作介面，包含至少讓使用者輸入的介面、顯示的介面、與點擊的介面。
- 102 一語音或文章輸入模組，讓使用者以文章或語音方式輸入欲搜尋的字串。
- 103 一自然語言處理模組，利用自然語言處理技術處理該搜尋的字串與該分析統計後的數據，並輸出相對應的字串或預測數據。

- 104 一大數據分析與預測模組，與該自然語言處理模組相連接，本大數據分析與預測模組將搜尋與統計歷年訴訟審判資料後，輸出相對應的分析或預測數據。
- 105 一資料視覺化模組，把該大數據分析與預測模組的預測數據，以圖像表格方式顯示到該使用者操作介面。
- 106 一自動推薦類別關鍵詞模組，讓使用者可以直接點擊自動推薦的類別關鍵詞，輸出給該大數據分析與預測模組。
- 107 一語音輸出模組，讓不識字或有需要的使用者可以聆聽使用者操作介面的資訊。



【發明摘要】

【中文發明名稱】

法律文件的大數據分析、預測、資料視覺化系統

【英文發明名稱】

BIG DATA ANALYSIS, PREDICTION, AND DATA VISUALIZATION SYSTEM
FOR LEGAL DOCUMENTS.

【中文】

本發明提出一種法律文件的大數據分析、預測、數據視覺化系統，包括協同過濾技術、自然語言處理技術、與資訊擷取技術，利用人工智慧大量搜尋與學習法律、法規、行政命令等。本發明讓一般民眾透過大數據分析所有相關法令、類似案件及訴訟數據，來更了解台灣的法律與法院實務判決，預測可能的訴訟方向與結果。本發明把法律數據視覺化，使法律不再是民眾難以觸及的領域，提供給律師事務所、法院、甚至所有民眾使用，並配合現代智慧型手機的普及與APP應用的蓬勃發展，發展對應的應用程式平台。

【指定代表圖】

第一圖

【代表圖之符號簡單說明】

101~107：係為本發明系統之功能圖。

【發明申請專利範圍】

【第1項】

一種法律文件的大數據分析、預測、資料視覺化系統，包括：

一自動推薦類別關鍵詞模組，利用協同過濾技術，計算出歷史搜尋數據與本次搜尋數據的相似度，並將與歷史搜尋數據中與本次搜尋數據相似度最高的關鍵詞，作為自動推薦的類別關鍵詞；

一使用者操作介面，以圖像表格方式顯示分析或預測數據，以及讓使用者可以直接點擊該自動推薦的類別關鍵詞；

一大數據分析與預測模組，利用資訊擷取技術來搜尋與統計歷年訴訟審判資料，提取出至少包括人、事、地、與時間的特定事件，並獲得加總後的分析數據；當使用者點擊該自動推薦類別關鍵詞時，該大數據分析與預測模組自動比對歷史統計數據，並輸出相對應的分析或預測數據；

一資料視覺化模組，把該大數據分析與預測模組的分析或預測數據，以圖像表格方式顯示到該使用者操作介面。

【第2項】

如申請專利範圍第1項所述之法律文件的大數據分析、預測、資料視覺化系統，還包括：

一語音或文章輸入模組，讓使用者以文章或語音方式輸入欲搜尋的字串；

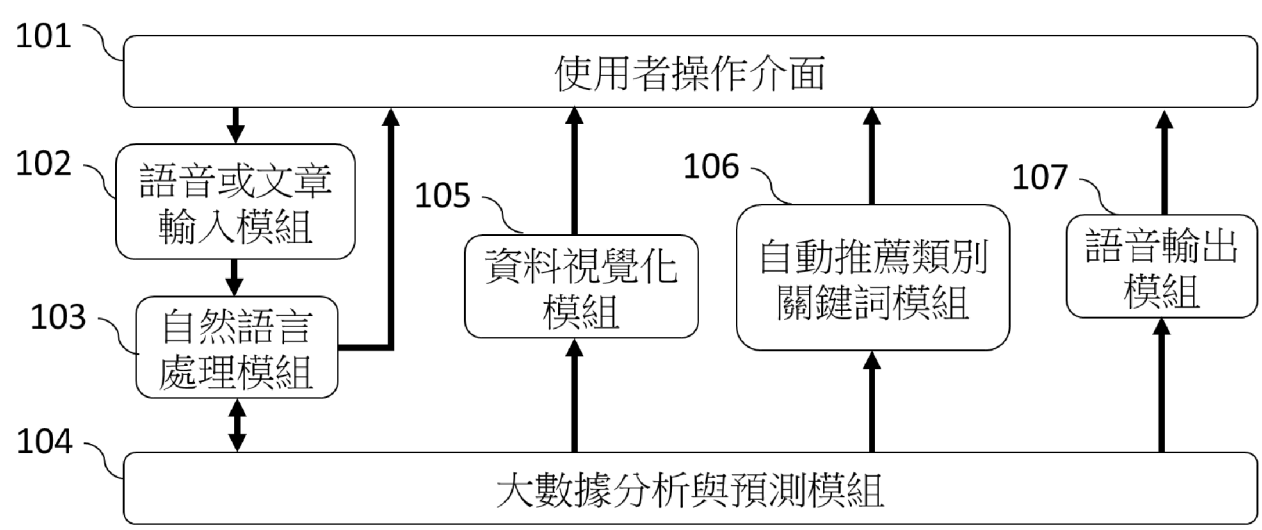
一自然語言處理模組，利用自然語言處理技術處理該搜尋的字串與該分析統計後的數據，並輸出相對應的字串或預測數據；

一語音輸出模組，讓不識字或有需要的使用者可以聆聽該使用者操作介面的資訊。

【第3項】

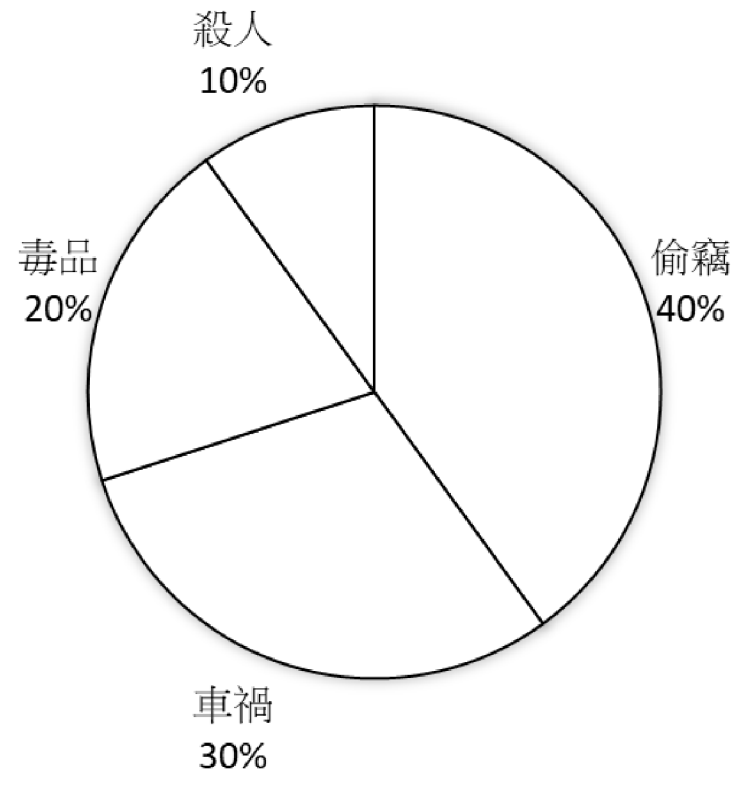
一種實現如申請專利範圍第1項所述之法律文件的大數據分析、預測、資料視覺化系統的裝置，其中該裝置係手機、行動電子裝置、平板電腦、或個人電腦。

【發明圖式】



【圖1】

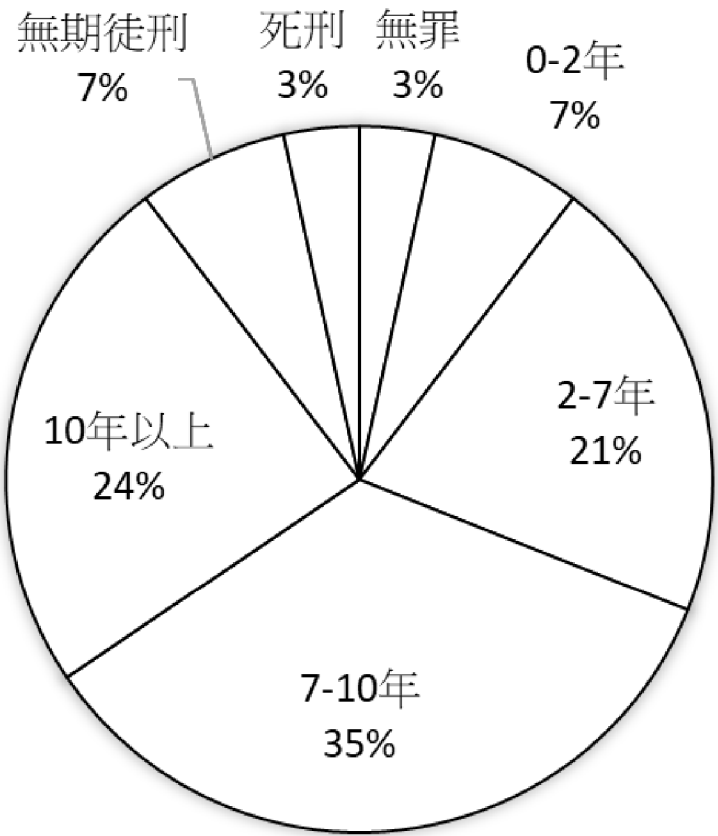
台北地方法院



【圖2】

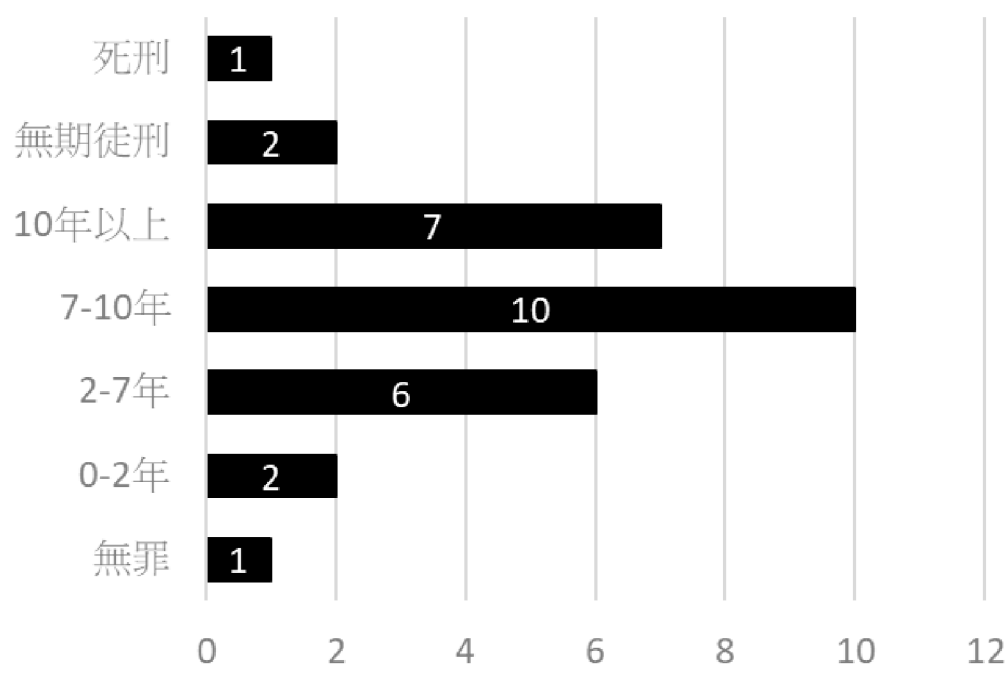
台北地方法院

殺人罪刑期



【圖3】

台北地方法院
殺人罪刑期



【圖4】



【圖5】

(19)



INTELLECTUAL PROPERTY OFFICE

(11) Publication number:

NZ 516822 A

(41) Publication date:

28.05.2004

(51) Int. Cl:

G06N5/00; G06N5/02;

(12)

Patent Application

(21) Application number: **NZ20000516822**

(71) Applicant(s):

LEXIS NEXIS

(22) Date of filing: **31.07.2000**

(72) Inventor(s):

**WILTSHIRE JAMES S
MORELOCK JOHN T
HUMPHREY TIMOTHY L
LU X ALLEN
PECK JAMES M
AHMED SALAHUDDIN**

(30) Priority: **19990147389P US 06.08.1999
2000US17872 WO 31.07.2000**

(54) **Title:**

**System and method for classifying legal concepts using
legal topic scheme**

(57) **Abstract:**

A computer implemented method of building a knowledge base for a legal topic classification system comprises imputing a plurality of training documents, parsing the plurality of training documents to extract classified legal concepts, extracting features (terms, legal phrases and case citations) generating relevance scores for each feature and storing features, topics and relevance scores in a knowledge base using an inverted index.

A computer implemented method of building a knowledge base for a legal topic classification system comprises imputing a plurality of training documents, parsing the plurality of training documents to extract classified legal concepts, extracting features (terms, legal phrases and case citations); generating relevance scores for each feature and storing features, topics and relevance scores in a knowledge base using an inverted index.

**SYSTEM AND METHOD FOR CLASSIFYING LEGAL CONCEPTS
USING LEGAL TOPIC SCHEME**

5 COPYRIGHT NOTICE. A portion of this disclosure, including Appendices, is subject to
copyright protection. Limited permission is granted to facsimile reproduction of the
patent document or patent disclosure as it appears in the U.S. Patent and Trademark
Office (PTO) patent file or records, but the copyright owner reserves all other copyright
rights whatsoever.

BACKGROUND OF THE INVENTION

1. **Field of the Invention**

15 The present invention relates to systems and methods for automated classification.
More specifically, the invention relates to automated systems and methods for classifying
concepts (such as legal concepts, including points of law from court opinions) according
to a topic scheme (such as a hierarchical legal topic classification scheme).

20 2. **Related Art**

Document classification has long been recognized as one of the most important
tasks in text processing. Classification of documents provides for quality document
retrieval, and enables browsing and linking among documents across a collection. The

benefits of such easy access are especially apparent in slowly-evolving subject domains such as law. The generally stable vocabularies and topics of the legal domain insure long-term return on any classification work.

5 There are two broad document classification approaches: unsupervised learning and supervised learning. The approaches are differentiated by whether a pre-defined classification scheme is used.

10 Unsupervised learning is a data-driven classification approach, based on the assumption that documents can be well organized by a natural structure inherent to the data. Those familiar with the data should be able to follow this natural structure to locate their information. A large body of information retrieval literature has focused on this approach, mostly related to document clustering [Borko 1963, Sparck Jones 1970, van Rijsbergen 1979, Griffiths 1984, Willett 1988, Salton 1990]. More recently some machine learning techniques have been applied to this classification task [Farkas 1993] – the term “unsupervised learning” was coined to describe this approach. The following
15 patents are associated with this approach: U.S. Patent No. 5,182,708 and U.S. Patent No. 5,832,470.

20 Opposite to the unsupervised learning approach to document classification is supervised learning. With this approach, a pre-defined “topic scheme” is given, along with the classified documents for each topic in the scheme. The topic scheme may be a simple list of discrete topics, or a complex hierarchical topic scheme. Supervised learning technology focuses on the task of feeding a computer meaningful topical descriptions so that it can learn to classify a document of unknown type.

When a topic scheme includes a simple list of discrete topics (one without a complex hierarchical relationships among the topics), the document classification becomes mere document categorization. Many machine learning techniques, including the retrieval technique of relevance feedback, have been tried for this task [Buckley 1994, Lewis 1994, and Mitchell 1997]. In addition to the effectiveness of learning methods themselves, the success of automatic categorization depends on the number of topics in the scheme, on the amount of quality training documents, and on the degree that the topics are mutually exclusive to one another. An example is disclosed in U.S. Patent No. 5,675,710.

The more difficult document classification centers on classifying documents using a hierarchical topic scheme. In this task, one has to consider *horizontal* relationships among the sister topics, which tend to be close to each other and are thus confusing to a computer. Moreover, one must also be concerned with *vertical* inheritance relationships.

Many machine learning techniques have trouble accommodating these two semantic relationships simultaneously in their learning or training, and thereafter have difficulty in classifying documents effectively. The task becomes more challenging if the topic scheme is very large, if the training documents are not topically exclusive, if the size of documents is small, or if the documents lack descriptive information.

To face these challenges, some techniques (U.S. Patent No. 5,204,812) have relied on human intervention. Others (U.S. Patent No. 5,794,236) use simple but insightful pattern matching. Still others (U.S. Patent Nos. 5,371,807 and 5,768,580) turn to linguistic knowledge to combat the ambiguity introduced in the hierarchical scheme.

However, these techniques can only handle small, domain-specific classification work. They have difficulty in scaled processing, either because of their simplicity in pattern recognition or because of the daunting demand of building expensive lexicons to support the linguistic parsing.

5 Thus, there is a need in the art to develop an economic, scalable machine learning process that can perform document classification with high accuracy using a large, hierarchical topic scheme. It is to meet this need that the present invention is directed.

Non-Patent References mentioned above:

- 10 ▪ Borko, H. and Bernick M. 1963. "Automatic document classification." *Journal of the Association for Computing Machinery*, pp.151-161.
- Sparck Jones, K. 1970. "Some thoughts on classification for retrieval." *Journal of Documentation*, pp.89-102.
- Van Rijsbergen, C. J. 1979. *Information Retrieval*, 2nd edition, Butterworths, London.
- 15 ▪ Griffiths, A and others. 1984. "Hierarchic agglomerative clustering methods for automatic document classification." *Journal of Documentation*, pp. 175-205.
- Willett, P. 1988. "Recent trends in hierarchic document clustering: A critical review." *Information Processing and Management*, pp. 577-598.
- 20 ▪ Salton, G. and Buckley C. 1990. "Flexible text matching for information retrieval." Technical Report 90-1158, Cornell University, Ithaca, New York.
- Farkas, J. 1993. "Neural networks and document classification." *Canadian Conference on Electrical and Computer Engineering*, pp. 1-4.

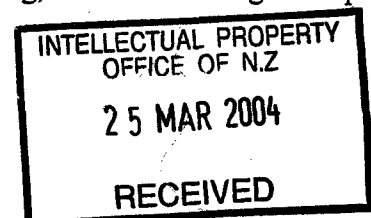
- Buckley, C and others. 1994. "Automatic routing and ad-hoc retrieval using SMART: TREC-2." *The 2nd Text Retrieval Conference*, edited by Donna Harman, NIST Special Publication 500-215, pp.45-55.
- Lewis, D. D. and Gale, W. A. 1994. "A sequential algorithm for training text classifiers." *Proceedings of the 7th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pp.3-12, London.
- Mitchell, T. 1997. *Machine Learning*, McGraw Hill, New York.

SUMMARY OF THE INVENTION

The present invention is directed to a computer-implemented method of building a knowledge base for a legal topic classification system by inputting and parsing training documents to extract classified legal concepts from such documents, extracting features from the legal concepts, generating relevance scores for the features, and storing the features, topics and relevance scores in a knowledge base using an inverted index.

The present invention is further directed to a computer-implemented method of building a knowledge base for a legal topic classification system by analyzing previously classified legal concepts to determine distinguishing features for each concept, generating relevance scores for each feature in each training concept, and storing features, topics and relevance scores in a knowledge base using an inverted index.

The present invention is yet further directed to a computer-implemented method of building a knowledge base for a legal topic classification system by identifying features within previously classified training concepts and generating relevance scores for these features. Such method includes converting the features into terms; generating, for each training concept, a term

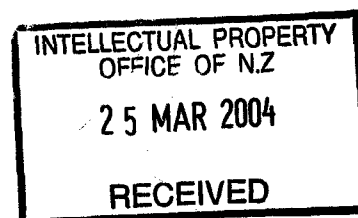


frequency, average term frequency, and document frequency for each term; determining the training set as the total number of training concepts in the knowledge base; generating an inverse document frequency for each term; and generating a relevance score for each term for each concept.

The present invention is also directed to a computer-implemented method of processing an input concept from a document text to provide, from a topic scheme, a list of one or more topics that are relevant to the input concept. Such method includes analyzing the input concept to identify candidate concept features, converting the candidate concept features to candidate terms, searching a database of concepts that were previously classified according to the topic scheme for concepts similar to the input concept based on the features, ranking the similar concepts based on relevance score, and voting on topics associated with the concepts within the database to form the list of topics relevant to the input concept.

The present invention is still further directed to a computer-implemented method of processing an input concept from a document text to provide, from a topic scheme incorporating a plurality of training concepts which have been previously classified and scored in accordance with the topic scheme, a list of one or more topics that are relevant to the input concept. Such method includes retrieving topics associated with the training concepts from a knowledge base, grouping training concepts and scores by associated topic, calculating a total relevance score for each topic as a sum of the training concept scores for each topic, and sorting topics by total topic relevance score to create a topic list relevant to the input concept.

The present invention is yet further directed to a computer-implemented method of processing an input concept from a document text to identify, within a knowledge base incorporating a plurality of training concepts, concepts similar to the input concept and to rank



these similar concepts. Such method includes identifying features of the input concept as candidate terms, retrieving from the knowledge base relevance scores for training concepts similar to the input concept, calculating a total relevance score for each retrieved training concept as a sum of candidate term relevance scores for that concept, and sorting retrieved training concepts by total relevance scores.

Other features of the present invention will be apparent to those skilled in the art upon a reading of this specification including the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

The invention is better understood by reading the following Detailed Description of the Preferred Embodiments with reference to the accompanying drawing figures, in which like reference numerals refer to like elements throughout, and in which:

FIG. 1 illustrates an exemplary hardware configuration in which the inventive classification system and method may be implemented.

FIG. 2 is a high-level flow chart schematically indicating a training process 200 and a classification process 210, alongside a knowledge base 201, a topic scheme 202 and exemplary lists 203, 204 that are used in the processes.

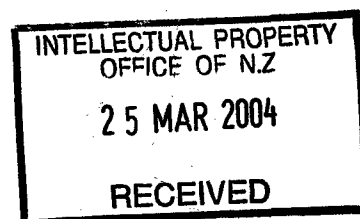
FIG. 3 is a flow chart indicating an exemplary training process 200 (FIG. 2).

FIG. 4 is a flow chart indicating details of an exemplary legal concepts extraction step (301 from FIG. 3; 701 from FIG. 7).

FIG. 5 is a flow chart indicating details of an exemplary feature extraction step (302 from FIG. 3; 702 from FIG. 7).

FIG. 6 is a flow chart representing details of an exemplary knowledge building step 303 (FIG. 3).

FIG. 7 is a flow chart indicating an exemplary embodiment of classification process 210 (FIG. 2).



Other objects, features and advantages of the present invention will be apparent to those skilled in the art upon a reading of this specification including the accompanying drawings.

5

BRIEF DESCRIPTION OF THE DRAWINGS

The invention is better understood by reading the following Detailed Description of the Preferred Embodiments with reference to the accompanying drawing figures, in which like reference numerals refer to like elements throughout, and in which:

10 FIG. 1 illustrates an exemplary hardware configuration in which the inventive classification system and method may be implemented.

FIG. 2 is a high-level flow chart schematically indicating a training process 200 and a classification process 210, alongside a knowledge base 201, a topic scheme 202 and exemplary lists 203, 204 that are used in the processes.

15 FIG. 3 is a flow chart indicating an exemplary training process 200 (FIG. 2).

FIG. 4 is a flow chart indicating details of an exemplary legal concepts extraction step (301 from FIG. 3; 701 from FIG. 7).

FIG. 5 is a flow chart indicating details of an exemplary feature extraction step (302 from FIG. 3; 702 from FIG. 7).

20 FIG. 6 is a flow chart representing details of an exemplary knowledge building step 303 (FIG. 3).

FIG. 7 is a flow chart indicating an exemplary embodiment of classification process 210 (FIG. 2).

FIG. 8 is a flow chart representing details of an exemplary concept classification step 703 (FIG. 7).

FIG. 9 is a flow chart illustrating details of an exemplary process 801 (FIG. 8) for ranking training concepts by relevance score.

5 FIG. 10 is a flow chart illustrating details of an exemplary process 802 (FIG. 8) for voting to generate a list of relevant topics.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

10 In describing preferred embodiments of the present invention illustrated in the drawings, specific terminology is employed for the sake of clarity. However, the invention is not intended to be limited to the specific terminology so selected, and it is to be understood that each specific element includes all technical equivalents that operate in a similar manner to accomplish a similar purpose.

15 **Background terminology.** As background to understanding the inventive system and method, it is understood that preferred embodiments classify "legal concepts" in a case law document (court opinion), using machine-based learning techniques. Then, the classification process is carried out according to a pre-defined scheme of "legal topics."

20 In particular, a "concept" in the legal domain, called a "legal concept," may be more widely known as a "rule of law." A topic in the legal domain, called a "legal topic," may be more widely known as an "area of law." Of course, the invention can be applied to domains other than the legal domain, and the broad terms "concept" and "topic" should not be limited to the legal domain.

Referring to the particular embodiment that is applied to the legal domain, case law documents must each have a set of distinct legal concepts. In a particular preferred embodiment, "legal concepts" may be defined as "controlling points of law, material to the disposition of the case, stated in the language of the court." Typically, a judicial opinion passage contains a legal concept if:

1. The passage is a *positive statement of a rule of law*, such as:
 - a direct expression of a legal principle (the test, the elements, the general rule, the exception to the general rule, etc.)
 - a definition of a legal term of art
 - a statement of the applicable standard of review
 - an express statement that another case is overruled or disapproved
 - an interpretation or construction—not just a quotation—of a statute or court rule
2. The rule of law stated is *significant* to the court's resolution of the case.
3. The court expressly or implicitly *adopts* the stated rule of law as its own.

A sample set of legal *concepts* is shown in Appendix A.

A sample scheme of legal *topics* is shown in Appendix B. While this sample scheme is hierarchical in nature (with general top-level topics and more specific lower-level topics), this invention is not limited to this type of scheme.

With this terminology understood, the invention is first described in general terms, followed by a more detailed explanation.

Overview. A preferred embodiment of the present invention provides a legal concept classification system and method that analyze text of a concept from a legal document and provide relevant topics for the legal concept from a given legal topic scheme. The invention uses a database of legal concepts previously classified according to the legal topic scheme, a list of legal phrases, and a list of stopwords.

The preferred embodiment involves two major processes to provide topics for legal concepts: training, and classifying. The system is first trained to distinguish topic trends in legal concepts according to the given topic scheme. Once trained, the system then classifies other legal concepts according to this same topic scheme.

Training process—overview. The process of training involves:

- gathering of training data,
- extraction of previously-classified legal concepts,
- analysis of the “features” in the training data,
- calculation of relevance scores for these features, and
- storage of this information in a knowledge base.

This training must be done to initially create information in the knowledge base, but it is envisioned that training may occur throughout the life of the application with feedback of newly classified legal concepts, to continually improve the quality of the knowledge base.

A plurality of case law documents is parsed, extracting classified legal concepts from the appropriate section of the document. This requires a considerable sample of case law documents from which legal concepts have been extracted and classified according to the same topic scheme.

Once extracted from the case, legal concepts are analyzed to determine the distinguishing “features” of these concepts. In a particular embodiment, the features for the training and classification processes have been identified:

- Terms
- Legal phrases
- Case cites

The creation of these features from the classified legal concepts is best understood with the following hypothetical example: a legal topic, followed by a legal concept as might be found in a case:

Criminal Law & Procedure—Evidence—Opinion Testimony

It is the role of the factfinder, not the appellate courts, to judge the credibility of witnesses, or lack thereof, and to decide whether to accept such testimony or to disregard it entirely; the appellate court must view the evidence in a light most favorable to the jury's verdict, as noted in *Smith v. Ohio*.

This legal concept's topic (Criminal Law & Procedure-Evidence—Opinion Testimony) is from a hierarchical (multi-tiered) legal topic scheme. More specifically:

- the top-tier topic, the most general, is “Criminal Law & Procedure”;
- the 2nd-tier topic, more specific, is “Evidence”; and
- the 3rd-tier, the most specific, “Opinion Testimony”.

From the text of this legal concept, the term features are extracted, excluding meaningless “stop-words” and depluralizing to normalize terms. For example, the terms likely to be significant features from the above sample legal concept might be “credibility”, “witness”, or “disregard”. The citation *Smith v. Ohio* would be extracted as a case cite feature. A legal phrase in this sample legal concept would be “appellate

courts". If available, opinion text identified to be relevant to each legal concept is also scanned to find relevant case citations. Of course, the exact usage of these features relative to the learning processes to be described should not limit the scope of the invention.

5 Once extracted and analyzed, relevance scores are calculated for each feature in each concept. This step of the training process uses a learning step based on a relevance-ranked text retrieval process. This step defines the relevance of features using frequency of the features, both within each legal concept and across the entire set of training legal concepts. These two frequencies -- within a legal concept and across the entire set -- are
10 combined to give a relevance score for each feature, for each legal concept. These legal concept relevance scores are then used to identify the most relevant topics for the candidate concept, during the classification process.

 The features and their scores, along with links to their related legal concepts, are then stored in the knowledge base for use during the subsequent classification process.

15 **Classification process—overview.** Once trained, the invention is used to classify previously-unclassified concepts such as legal concepts.

 According to a preferred embodiment, this classification *process* involves a classification *step* that includes:

- analyzing a "candidate" (or "target" or "input") legal concept for features,
- searching for similar training legal concepts in the knowledge base,
- ranking these similar concepts based on similarity to the candidate concepts, and
- voting to identify the most relevant topics from these similar legal concepts.

The candidate legal concept is analyzed for features. This step is identical to the analysis of features in classified legal concepts done during the training process. The same set of distinguishing features (terms, legal phrases, and case citations) used during training should be used in this step to be compatible with the knowledge base.

5 Next, the knowledge base is searched for *training* legal concepts similar in features to the *target* legal concept. These matching legal concepts are collated and ranked according to strength of match.

10 From these matching training legal concepts, the most relevant topics for the candidate legal concept are identified. Each legal concept in the knowledge base has at least one topic associated with it, so a list of topics is generated from the matching legal concepts. This topic list is sorted by relevance, and the most relevant topics chosen for the candidate legal concept.

15 The training process and the classification process (including the relevance-ranked classification step) having been described briefly above, the embodiments of the invention are now described in greater detail.

20 The inventive method is more easily understood with references to examples. The examples used herein primarily deal with classifying individual legal concepts of a case law document according to a particular hierarchical topic scheme, but this does not limit the scope of the present invention. In fact, any size text unit such as a sentence, passage, paragraph or an entire section of a document can be classified according to an arbitrary topic scheme.

Exemplary hardware embodiment. Embodiments of the inventive training and classification system may be implemented as a software system including a series of modules on a conventional computer.

As shown in FIG. 1, an exemplary hardware platform includes a central processing unit 100. The central processing unit 100 interacts with a human user through a user interface 101. The user interface is used for inputting information into the system and for interaction between the system and the human user. The user interface includes, for example, a video display, keyboard and mouse. Memory 102 provides storage for data (such as the knowledge base, stop word list and legal phrase list) and software programs (such as the training and classification processes) that are executed by the central processing unit. Auxiliary memory 103, such as a hard disk drive or a tape drive, provides additional storage capacity and a means for retrieving large batches of information.

All components shown in FIG. 1 may be of a type well known in the art. For example, the system may include a SUN workstation including the execution platform SPARCsystem 10 and SUN OS Version 5.5.1, available from SUN MICROSYSTEMS of Sunnyvale, California. Of course, the system of the present invention may be implemented on any number of computer systems. Moreover, although the preferred embodiment uses the PERL language (for text parsing tasks) and C++ (for number crunching and database access tasks), any appropriate programming language may be used to implement the present invention.

The training and classification processes. Referring to FIG. 2, the preferred embodiment includes a two step process as shown in which the system is first trained (block 200) before it classifies legal concepts (in block 210).

A knowledge base 201 is utilized to store training results during the training process. The training results stored in the knowledge base during the training process are used in the subsequent classification process.

Both the training process and the classification process make use of a predetermined topic scheme 202 (see example of a legal topic scheme in Appendix B). In a particular preferred embodiment, the training process and the classification process also make use of a Stop Word List 203 (see example in Appendix C) and a Phrase List 204 (see example of a Legal Phrase List in Appendix D).

The training process. The machine learning system is first trained, as shown in FIG. 3 and related figures, before it classifies legal concepts. This training requires case law documents with legal concepts identified, and each concept classified according to the target topic scheme.

The training process links the set of each legal concept's extracted features with that legal concept's associated topic(s). The following discussion describes an embodiment of the training process, based on a relevance-ranked approach.

The embodiment of the concept ranking step used in the training process uses frequency of occurrence of terms and features, both within each individual legal concept and across the entire set of legal concepts. Generally, a term or feature that occurs frequently within a legal concept is a strong indicator of the topic, *unless* that term or feature also frequently occurs in *many other* legal concepts. Intuitively, a highly

common legal term, like "court" or "trial", does not contribute to assigning a specific legal topic to a legal concept.

5 First, a plurality of training documents are input, as shown at block 300. Then, that plurality of case law documents is parsed to extract the legal concepts, as shown at block 301. "Features" (terms, legal phrases and embedded citations) of each legal concept are then extracted (block 302) and attached to the text in a manner suitable to the learning method being used. Relevance scores are then generated for these features (block 303). Finally, the results are stored (block 304) in knowledge base 201.

10 Referring more specifically to FIG. 3, these training steps are now discussed in more detail. Many of the steps in the training process are also used in the classification process that is discussed with reference to FIGS. 8 and 9.

The step of extracting legal concepts is used as step 301 during the training process shown in FIG. 3 and as step 701 during the classification process shown in FIG. 7. The details of steps 301 and 701 are illustrated in FIG. 4.

15 Referring to FIG. 4, block 400 illustrates the partitioning of text accomplished when a case law document is parsed and partitioned by section, to identify the section containing legal concepts.

20 Then, as shown in block 401, the legal concept section is parsed and partitioned into individual legal concepts. Each legal concept is stored along with the topic(s) associated with that legal concept. If a legal concept has no topics from the target topic scheme, then that legal concept is discarded.

The feature extraction step, used as step 302 during training and as step 702 during classification, is detailed in FIG. 5. This step involves extracting the features

necessary for accurately classifying legal concepts, and is one of the primary objects of this invention. The format of the feature when associated with the legal concept depends on the format required by the learning method. Examples are given for the relevance-ranked classification step, described in more detail with reference to FIG. 8.

5 Referring more specifically to FIG. 5, each term in the legal concept is extracted from the text, as shown in step 500. However, all instances of stop-words from the stop-word list (see an example in Appendix C) are removed from the set of terms.

10 In step 501, legal phrases such as "criminal history", "custody dispute", "eminent domain" etc. are extracted as a feature. The legal concept is searched for legal phrases (see sample in Appendix D).

Finally, step 502 involves extracting cites to other case law documents, such as "*People v. Medina* (1995) 39 Cal. App. 4th 643, 650".

After features have been extracted, then relevance scores for each feature are generated. The generation of relevance scores includes:

- 15
- the conversion of features into terms,
 - the generation of term frequencies within legal concepts and across the set,
 - the generation of (so-called) "document" frequencies (more aptly, *concept* frequencies),
 - the generation of inverse "document" frequencies (more aptly, inverse *concept* frequencies), and
 - 20 • the generation of relevance scores for each term.

Details of these steps are described as follows, with reference to FIG. 6.

A straightforward approach to using abstract features in a relevance-ranked approach is to simply convert each feature in a legal concept into a "term" (shown as step 600). In a particular preferred embodiment, a "term" is a mnemonic that uniquely distinguishes that feature from all other features.

5 For example, the legal phrase "administrative authority" can be easily converted to a "term" like "administrative_authority" (with an underline character between words). Or a case citation like "39 Cal. App. 4th 643" can be converted into a term like "39_CalApp4_643". In this way, each feature is converted into a term that is well-defined and unique across the set.

10 Then, as shown in block 601, for each "term" (including all features as well as words), the "term frequency" (TF) is calculated for each legal concept in which that term appears, using the number of occurrences of that term within the legal concept. The "average term frequency" (AVE_TF) of all terms in the legal concept is also calculated.

15 For each term, the total number of legal concepts in which a term occurs in the training set is determined, as shown in block 602. This number is determined in the same way that the conventional "document" frequency (DF) is calculated. In the text searching art, the term "document" has many meanings and can thus be ambiguous. In this specification "document" has already been used to refer to the *entire legal opinion* and not to a concept that is a part of the opinion. Therefore, to avoid ambiguity, the expanded
20 term "document frequency" will no longer be used in this specification. Instead, DF will continue to be used, with the understanding that in the context of this specification DF actually refers to a *concept* frequency.

Block 603 represents the calculation of how widely that term is used across the entire body of training legal concepts. This calculation is made in the same way that the conventional inverse "document" frequency (IDF) is calculated. Because "document" is used in this specification to refer to an entire legal opinion, the expanded term "inverse document frequency" will no longer be used. However, it is understood that IDF actually refers to an inverse *concept* frequency. In any event, this calculation is made using the DF and the total number of legal concepts in the training set DBSIZE.

Next, for each term in a legal concept, a relevance score is then calculated, as shown in block 604. This calculation involves using the term frequency (TF) for that term-legal concept pair, the AVE_TF, that term's IDF, the length of the legal concept, and the overall average length of legal concepts in the set. This scoring technique is one of the primary objects of this invention.

Exemplary formulas for calculating the relevance score are:

if (*doclength* > *aveDocLength*):

$$TFwt = \frac{TF + TF / AVE_TF}{TF + TF / AVE_TF + 2(\alpha + \beta \times (doclength - aveDocLength) / aveDocLength)}$$

if (*doclength* <= *aveDocLength*):

$$TFwt = \frac{TF + TF / AVE_TF}{TF + TF / AVE_TF + 2(\alpha + \beta \times (aveDocLength - doclength + 1) / aveDocLength)}$$

$$IDF = \log\left(\frac{DBSIZE - DF + 0.5}{DF + 0.5}\right)$$

$$score = TFwt \times IDF$$

where:

TFwt = Term frequency weight

5 TF = Term frequency within current legal concept

AVE_TF = Average term frequency of terms in the current legal concept

α, β = Scale factors, such that $\alpha + \beta = 1$

docLength = Length of current legal concept, in characters

aveDocLength = Average length, in characters, of all legal concepts in training set

10 IDF = Inverse "document" (i.e., *concept*) frequency for term across training set

DBSIZE = Total number of legal concepts in training set

DF = "Document" frequency (Number of legal concepts in which a term occurs)

15 score = Relevance score for term in a legal concept

Finally, the results of various calculations are stored in knowledge base 201, as shown in block 304 (FIG. 3). The concept frequency DF and the relevance score for each term in a legal concept are stored in an "inverted index" in the knowledge base. As will readily be understood by those skilled in the art, an inverted index in this context is a list of each term, each legal concept in which it occurs, and the relevance score for that term-legal concept, such that the list can be easily searched by term.

20

Significantly, use of an inverted index greatly increases the scalability of the invention. This is because an inverted index provides for very efficient searching on features and allows for handling much larger bodies of training data. A portion of an exemplary inverted index, given legal concepts 23, 38, and 127, sorted by term, may be represented as in TABLE I.

Also saved in the knowledge base are the basic relationships between each legal concept and its associated topic(s), determined earlier in the general training process. This set of relationships establishes the link between terms in a legal concept and the topics relevant to those terms.

The Classification Process. The *training* process 200 having been described in detail above, the inventive *classification* process 210 is now described. Shown in FIG. 7, the classification process 210 involves classifying legal concepts of an unknown topic according to a given topic scheme.

First, court case documents are input, as indicated at block 700. From each case, legal concepts are extracted, as shown in block 701. The extraction step 701 may be the same as step 301 used by the training process, shown in FIG. 4.

In block 702 (which may be the same as step 302 used by the training process, and shown in FIG. 5), features are extracted from each legal concept and are associated with that legal concept in a manner consistent with the training process used.

In concept classification step 703, each legal concept and its identified features are first input. Information gathered during training is input from the knowledge base. This information is used to generate a set of scores for the current legal concept, one each for the best-matching legal concepts in the training set. Then, the topics for the legal

concept are determined. Block 703 involves using the features found in the current candidate legal concept, and comparing them to the features found in the training legal concepts. The topics associated with the training legal concepts found most similar to the legal concept in question are collated and sorted to determine the most relevant topics.

5 A preferred embodiment of the concept classification step 703 (FIG. 7) is detailed in FIG. 8. The illustrated classification step uses frequency of features in the candidate legal concept to find similar legal concepts (and therefore similar topics) in the training knowledge base. The classification step, which may be termed a relevance-ranked classification step, includes:

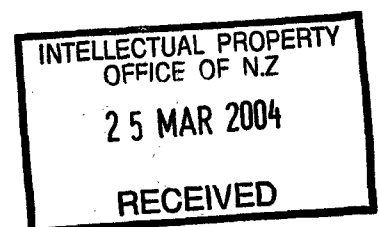
- 10
- converting features to terms
 - ranking of training concepts by relevance score for these features, and
 - voting on the topics associated with these training concepts to determine the best topics.

15 Block 800, a step of converting "features" to "terms", may be the same as that used by FIG. 6 block 600, described above.

In block 801, all training concepts are ranked by their relevance scores for these candidate terms.

20 A preferred embodiment of the ranking step 801 (FIG. 8) is detailed in FIG. 9. Referring to FIG. 9, the first two steps, illustrated as blocks 900 and 901, are optional but are useful in optimizing for later steps.

In block 900, the list of candidate terms is sorted in ascending order by DF, retrieved from the knowledge base. This orders the term list from the least common terms to the most common terms.



Next, in block 901, the candidate term list is reduced to a selection of the least common terms in the list. This in turn reduces the processing required in subsequent ranking steps; the number of terms selected depends on the amount of optimization desired. For example, refer to Table II. If optional steps 900 and 901 are not used, subsequent required ranking steps operate on the entire list of candidate terms.

In block 902 the relevance scores for training concepts, for all terms in the candidate concept, are retrieved from the knowledge base.

In block 903, for every training concept, the relevance scores for the candidate terms are summed into a total relevance score for that training concept. Table III shows an exemplary set of five candidate terms with seven training legal concepts and their relevance scores. The relevance scores for each legal concept are totaled in the bottom row. This would give a list of training legal concepts with total relevance scores, such as shown in Table IV.

In block 904, the training concepts are sorted descending by these total relevance scores, resulting in the most similar training concepts being first in this sorted list. For example, sorting Table IV by total score yields the results in Table V. This sorted list generally indicates that the most relevant training legal concept for the current candidate concept is Concept 1, the next most relevant concept is Concept 9, and so forth down the list.

This completes discussion of the details of FIG. 8 block 801 as it may be implemented in FIG. 9.

Referring again to FIG. 8, in block 802, essentially a voting step, the topics associated with the training legal concepts that are most like the candidate legal concept

are collated and sorted to generate a final list of relevant topics. In a final list of matching legal concepts, a single topic may be found more than once and therefore is more likely relevant to the candidate concept than topics found only once, for example.

Significantly, the voting process of block 802 distinguishes the more relevant topics from the less relevant topics. This voting technique is one of the primary objects of the invention.

A preferred embodiment of the voting step 802 (FIG. 8) is detailed in FIG. 10. In block 1000, the topics associated with the sorted training concepts are retrieved from the knowledge base. Then, in block 1001, these training concepts and their scores are grouped by their topics. In block 1002, the total relevance score for each topic-group is calculated to determine a topic relevance score for that topic. Finally, in block 1003, the topics are then sorted descending by these topic relevance scores. The resulting list shows the most relevant topics first in the list.

Given a flat (non-hierarchical) topic scheme, a sample list of matching concepts before voting, with their topics and relevance scores, sorted by score, might look like TABLE VI.

Thus, *before* voting, the most relevant topics from this list would be, in order of appearance:

- Admiralty Law
- Transportation Law
- Torts
- Bankruptcy Law

After voting, the scores of all the topics are accumulated and the results re-sorted, would look like TABLE VII. So the list of relevant topics *after* voting would be sorted differently than that *before* voting:

- Transportation Law (6.53)
- Admiralty Law (4.55)
- Torts (2.81)
- Bankruptcy Law (0.68)

The "Transportation Law" topic becomes the most relevant topic after voting because its accumulated relevance score is higher than the accumulated scores of the other topics.

If the legal topic scheme is a hierarchical (multi-tier) topic scheme, a second hierarchical voting may be performed on the final relevant topics. The final topic list can be grouped by 1st-tier topic, by 2nd-tier topic, and so on, and then weighted according to occurrence at each tier. These weights are then considered in the final list of topics. This technique takes into consideration similar topics and can help to improve overall quality of the topics.

For example, the list of matching concepts, before hierarchical voting, with their topics and relevance scores, sorted by score, might look like TABLE VIII.

If the relevance scores for these topics are accumulated, first by Tier 1 to give CumT1, then by Tier 2 to give CumT2, then by Tier 3 to give CumT3 (assume only a 3-tier hierarchy for this example), the topic list looks like TABLE IX.

This list is then sorted highest to lowest score, first by Tier 1, then by Tier 2, and then by Tier 3, to give TABLE X.

This would be the final list of most relevant topics, sorted by relevance. This final list of topics might be represented hierarchically as follows:

Transportation Law
 Water Transportation & Shipping
 Vessel Safety
 Carrier Liabilities & Duties
 Ratemaking
 Vehicle Transportation & Shipping
 Traffic Regulation
 Intrastate Commerce Regulation
 Foreign Commerce Regulation
 Admiralty Law
 Cargo Care & Custody
 Liability Exemptions
 Liability
 Non-Cargo Liability
 Death Actions
 Jones Act
 Torts
 Products Liability
 Negligence
 Strict Liability
 Abnormally Dangerous Activities
 Vicarious Liability
 Negligent Hiring & Supervision
 Bankruptcy Law
 Property Use, Sale, or Lease

Still other (optional) techniques can be used to further eliminate irrelevant topics. For example, if a topic's relevance score is below a predefined threshold, or if the number of times the topic occurs among the most relevant legal concepts is below a threshold, then that topic could be eliminated.

The inventive methods having been described above, the invention also encompasses apparatus (especially programmable computers) for carrying out classification of legal concepts. Further, the invention encompasses articles of manufacture, specifically, computer readable memory on which computer-readable code

embodying the methods may be stored, so that, when the code is used in conjunction with a computer, the computer can carry out the training and classification processes.

5 A non-limiting, illustrative example of an apparatus that invention envisions for executing the foregoing methods, is described above and illustrated in FIG. 1: a computer or other programmable apparatus whose actions are directed by a computer program or other software.

Non-limiting, illustrative articles of manufacture (storage media with executable code) may include the memory 103 (FIG. 1), other magnetic disks, optical disks, "floppy" diskettes, ZIP disks, or other magnetic diskettes, magnetic tapes, and the like.
10 Each constitutes a computer readable memory that can be used to direct a computer to function in a particular manner when used by the computer.

Those skilled in the art, given the preceding description of the inventive method, are readily capable of using knowledge of hardware, of operating systems and software platforms, of programming languages, and of storage media, to make and use apparatus
15 for concept classification, as well as computer readable memory articles of manufacture which, when used in conjunction with a computer can carry out concept classification. Thus, the invention's scope includes not only the method itself, but apparatus and articles of manufacture.

Modifications and variations of the above-described embodiments of the present
20 invention are possible, as appreciated by those skilled in the art in light of the above teachings. For example, the particular hardware on which the system and method are implemented, the programming languages and data formats involved, the inclusion or exclusion of optional steps, the nature of the concepts to be classified, the particular topic

scheme used, and other details of implementation, may be varied while remaining within the scope of the present invention. It is therefore to be understood that, within the scope of the appended claims and their equivalents, the invention may be practiced otherwise than as specifically described.

TABLE I:

Term	Concept	Score
appellate	23	0.08
appellate	38	0.12
court	23	0.01
court	127	0.02
court	38	0.01
credibility	23	0.23
factfinder	23	1.13
factfinder	127	1.04
judge	38	0.05
role	23	0.57
role	38	0.88
role	127	0.42
witnesses	23	0.11
witnesses	38	0.08

5

TABLE II:

Term	DF
seaworthy	1584
seaworthy	1584
admiralty	8092
admiralty	8092
admiralty	8092
admiralty	8092
negligent	12744
negligent	12744
injure	28939
injure	28939
witness	42677
witness	42677
witness	42677
witness	42677
witness	42677
witness	42677

TABLE III:

Term	Training Legal Concept						
	1	2	4	9	10	13	15
seaworthv	0.9997	0	0	0.6458	0	0	0
seaworthv	0.9997	0	0	0.6458	0	0	0
admiralty	0.8023	0	0	0.9322	0.7083	0	0.4016
admiralty	0.8023	0	0	0.9322	0.7083	0	0.4016
admiralty	0.8023	0	0	0.9322	0.7083	0	0.4016
admiralty	0.8023	0	0	0.9322	0.7083	0	0.4016
negligent	0.7922	0	0	0.5738	0	0	0.7992
negligent	0.7922	0	0	0.5738	0	0	0.7992
injure	0	0	0	0	0.9955	0	0
injure	0	0	0	0	0.9955	0	0
witness	0.1047	0.0882	0	0.0023	0	0.4771	0.1025
witness	0.1047	0.0882	0	0.0023	0	0.4771	0.1025
witness	0.1047	0.0882	0	0.0023	0	0.4771	0.1025
witness	0.1047	0.0882	0	0.0023	0	0.4771	0.1025
witness	0.1047	0.0882	0	0.0023	0	0.4771	0.1025
witness	0.1047	0.0882	0	0.0023	0	0.4771	0.1025
TOTALS	7.4212	0.5292	0	6.1818	4.8242	2.8626	3.8198

5

TABLE IV:

Training Legal Concept	Total Score
1	7.4212
2	0.5292
4	0
9	6.1818
10	4.8242
13	2.8626
15	3.8198

10

TABLE V:

Training Legal Concept	Total Score
1	7.4212
9	6.1818
10	4.8242
15	3.8198
13	2.8626
2	0.5292
4	0

TABLE VI:

Concept	Topic	Score
1	Admiralty Law	1.53
2	Admiralty Law	1.27
3	Transportation Law	0.89
4	Torts	0.88
4	Transportation Law	0.88
5	Transportation Law	0.82
6	Torts	0.79
7	Transportation Law	0.74
8	Transportation Law	0.74
9	Admiralty Law	0.69
10	Bankruptcy Law	0.68
10	Transportation Law	0.68
11	Transportation Law	0.65
12	Torts	0.59
13	Transportation Law	0.58
14	Transportation Law	0.55
14	Admiralty Law	0.55
14	Torts	0.55
15	Admiralty Law	0.51

TABLE VII:

Concept	Topic	Score
1	Admiralty Law	1.53
2	Admiralty Law	1.27
9	Admiralty Law	0.69
14	Admiralty Law	0.55
15	Admiralty Law	0.51
		4.55
4	Torts	0.88
6	Torts	0.79
12	Torts	0.59
14	Torts	0.55
		2.81
3	Transportation Law	0.89
4	Transportation Law	0.88
5	Transportation Law	0.82
7	Transportation Law	0.74
8	Transportation Law	0.74
10	Transportation Law	0.68
11	Transportation Law	0.65
13	Transportation Law	0.58
14	Transportation Law	0.55
		6.53
10	Bankruptcy Law	0.68
		0.68

TABLE VIII:

Concept	Topic	Score
1	Admiralty Law – Cargo Care & Custody – Liability Exemptions	2.03
2	Admiralty Law – Non-Cargo Liability – Death Actions	1.33
3	Transportation Law – Vehicle Transportation & Shipping – Traffic Regulation	0.89
4	Torts – Strict Liability – Abnormally Dangerous Activities	0.88
4	Transportation Law – Water Transportation & Shipping – Vessel Safety	0.88
5	Transportation Law – Carrier Liabilities & Duties – Ratemaking	0.82
6	Torts – Products Liability – Negligence	0.79
7	Transportation Law – Intrastate Commerce Regulation	0.74
8	Transportation Law – Water Transportation & Shipping – Vessel Safety	0.74
9	Admiralty Law – Cargo Care & Custody – Liability Exemptions	0.69
10	Bankruptcy Law – Property Use, Sale, or Lease	0.68
10	Transportation Law – Foreign Commerce Regulation	0.68
11	Transportation Law – Water Transportation & Shipping – Vessel Safety	0.65
12	Torts – Products Liability – Negligence	0.59
13	Transportation Law – Water Transportation & Shipping – Vessel Safety	0.58
14	Transportation Law – Carrier Liabilities & Duties – Ratemaking	0.55
14	Admiralty Law – Cargo Care & Custody – Liability Exemptions	0.55
14	Torts – Vicarious Liability – Negligent Hiring & Supervision	0.55
15	Admiralty Law – Non-Cargo Liability – Jones Act	0.51

TABLE IX:

Concept	Topic	Score	CumT1	CumT2	CumT3
9	Admiralty Law -- Cargo Care & Custody -- Liability	0.69			0.69
1	Admiralty Law -- Cargo Care & Custody -- Liability Exemptions	2.03			
14	Admiralty Law -- Cargo Care & Custody -- Liability Exemptions	0.55		3.27	2.58
2	Admiralty Law -- Non-Cargo Liability -- Death Actions	1.33			1.33
15	Admiralty Law -- Non-Cargo Liability -- Jones Act	0.51	5.11	1.84	0.51
10	Bankruptcy Law -- Property Use, Sale, or Lease	0.68	0.68	0.68	0.68
6	Torts -- Products Liability -- Negligence	0.79			
12	Torts -- Products Liability -- Negligence	0.59		1.38	1.38
4	Torts -- Strict Liability -- Abnormally Dangerous Activities	0.88		0.88	0.88
14	Torts -- Vicarious Liability -- Negligent Hiring & Supervision	0.55	2.81	0.55	0.55
5	Transportation Law -- Carrier Liabilities & Duties -- Ratemaking	0.82			
14	Transportation Law -- Carrier Liabilities & Duties -- Ratemaking	0.55		1.37	1.37
10	Transportation Law -- Foreign Commerce Regulation	0.68		0.68	0.68
7	Transportation Law -- Intrastate Commerce Regulation	0.74		0.74	0.74
3	Transportation Law -- Vehicle Transportation & Shipping -- Traffic Regulation	0.89		0.89	0.89
4	Transportation Law -- Water Transportation & Shipping -- Vessel Safety	0.88			
8	Transportation Law -- Water Transportation & Shipping -- Vessel Safety	0.74			
11	Transportation Law -- Water Transportation & Shipping -- Vessel Safety	0.65			
13	Transportation Law -- Water Transportation & Shipping -- Vessel Safety	0.58	6.53	2.85	2.85

TABLE X:

Topic	CumT1	CumT2	CumT3
Transportation Law -- Water Transportation & Shipping -- Vessel Safety	6.53	2.85	2.85
Transportation Law -- Carrier Liabilities & Duties -- Ratemaking		1.37	1.37
Transportation Law -- Vehicle Transportation & Shipping -- Traffic Regulation		0.89	0.89
Transportation Law -- Intrastate Commerce Regulation		0.74	0.74
Transportation Law -- Foreign Commerce Regulation		0.67	0.67
Admiralty Law -- Cargo Care & Custody -- Liability Exemptions	5.11	3.27	2.58
Admiralty Law -- Cargo Care & Custody -- Liability			0.69
Admiralty Law -- Non-Cargo Liability -- Death Actions		1.33	1.33
Admiralty Law -- Non-Cargo Liability -- Jones Act			0.51
Torts -- Products Liability -- Negligence	2.81	1.38	1.38
Torts -- Strict Liability -- Abnormally Dangerous Activities		0.88	0.88
Torts -- Vicarious Liability -- Negligent Hiring & Supervision		0.55	0.55
Bankruptcy Law -- Property Use, Sale, or Lease	0.68	0.68	0.68

APPENDICES

5

APPENDIX A - LEGAL CONCEPTS

APPENDIX B - HIERARCHICAL LEGAL TOPIC SCHEME

10

APPENDIX C - STOP-WORD LIST

APPENDIX D - LEGAL PHRASE LIST

15

Concerning the content of the following Appendices, see the copyright notice at the beginning of the specification.

APPENDIX A**Exemplary LEGAL CONCEPTS (Points of Law, etc.)**

5 Concepts Listed beneath respective Legal Topics from Hierarchical Legal Topic Scheme

1. **Civil Procedure--Injunctions--Permanent Injunctions**
Civil Procedure--Appeals--Standards of Review--Abuse of Discretion
10 The grant or denial of an injunction is solely within the trial court's discretion and, therefore, a reviewing court should not disturb the judgment of the trial court absent a showing of a clear abuse of discretion. An abuse of discretion involves more than an error of judgment. It is an attitude on the part of the court that is unreasonable, unconscionable, or arbitrary.
2. **Civil Procedure--Appeals--Standards of Review--Abuse of Discretion**
15 A reviewing court should presume that the trial court's findings are accurate, since the trial judge is best able to view the witnesses and observe their demeanor, gestures, and voice inflections and use these observations in weighing the credibility of the witnesses.
3. **Contract Law--Consideration--Mutual Obligation**
20 For a contract to be enforceable, it must be supported by consideration.
4. **Labor & Employment Law--Employment at Will**
Since an employer is not legally required to continue the employment of an employee at-will, continued employment is consideration for the contract not to compete.
- 25 5. **Labor & Employment Law--Trade Secrets & Unfair Competition--Noncompetition Agreements**
30 A covenant restraining an employee from competing with his former employer upon termination of employment is reasonable if it is no greater than is required for the protection of the employer, does not impose undue hardship on the employee, and is not injurious to the public. Only those covenants which are reasonable will be enforced.

APPENDIX B

Section from an exemplary

HIERARCHICAL LEGAL TOPIC SCHEME

Admiralty Law
Arbitration
Bankruptcy
Bills of Lading
Cargo Care & Custody
Due Diligence
General Average
Liability
Liability Exemptions
Limitation of Liability
Charter Parties
Contribution & Indemnity
Insurance
Jurisdiction
Law of Salvage
Liens & Mortgages
Negligence & Unseaworthiness
Non-Cargo Liability
Death Actions
Jones Act
Longshore & Harbor Workers' Compensation Act
Penalties & Forfeitures
Antitrust & Trade Regulation
Sherman Act
Clayton Act
Robinson-Patman Act
Federal Trade Commission Act
Market Definition
Restraints of Trade & Price Fixing
Exclusive or Reciprocal Dealing
Horizontal Market Allocation
Horizontal Refusals to Deal
Horizontal Restraints
Per Se Rule & Rule of Reason
Tying Arrangements
Vertical Price Restraints

...
[items purposely omitted]

APPENDIX C

Selection from an exemplary

STOP-WORD LIST

5 A
 ABLE
 ABOUT
 ABOVE
10 ACCORDING
 ACROSS
 AFTER
 AGAIN
 AGAINST
 AGO
15 ALL
 ALLOW
 ALLOWED
 ALLOWING
 ALLOWS
20 ALMOST
 ALONE
 ALONG
 ALREADY
 ALSO
25 ALTHOUGH
 ALWAYS

 [Items purposely omitted]

30 WHATEVER
 WHEN
 WHERE
 WHETHER
35 WHICH
 WHILE
 WHO
 WHOLE
 WHOSE
40 WHY
 WILL
 WITH
 WITHIN
 WITHOUT
45 WON
 WOULD
 WOULDN
 YESTERDAY
 YET
50 YOU
 YOUR
 YOURS
 YOURSELF

APPENDIX D

Selection from an exemplary

LEGAL PHRASE LIST

5 14th amendment
4th ed
5th ed
6th ed
8th ed
10 abatement act
abettor statute
abnormal sexual interest
absent class member
absent evidence
15 absolute bar
absolute discretion
absolute divorce
absolute duty
absolute equality
20 absolute immunity
absolute priority rule
absolute privilege
absolute right
absolute title
25 abstention principle
abstract book
abuse-of-discretion standard
ad damnum
ad damnum clause
30 ad valorem
ad valorem tax
additional evidence
additional fact
additional peremptory challenge
35 additional punishment
additional suspect
adequate consideration
adequate notice
adequate record
40 adequate remedy

. . .
[items purposely omitted]

45 zoning appeal
zoning authority
zoning case
zoning enabling act
zoning law
50 zoning regulation

WE CLAIM:

1. A computer-implemented method of building a knowledge base for a legal topic classification system, the method comprising:

inputting a plurality of training documents;

parsing the plurality of training documents to extract classified legal concepts;

extracting features from the legal concepts;

generating relevance scores for each feature; and

storing features, topics, and relevance scores in a knowledge base, using an inverted index.

2. The method as set forth in claim 1, the step of parsing comprising the steps of:

partitioning the text by section; and

partitioning the text by legal concept.

3. The method as set forth in claim 1, the step of extracting features comprising the steps of:

extracting terms, excluding stop words;

extracting legal phrases; and

extracting embedded case citations.

4. The method as set forth in claim 1, the step of generating relevance scores including the steps of:

converting features to terms;
generating, for each training concept, term frequency (TF) for each term, as number of occurrences of that term in that training concept;
generating, for each training concept, document frequency (DF) for each term, as total number of training concepts in which term appears;
generating inverse document frequency (IDF) for each term; and
generating a relevance score for each term for each concept.

5. A computer-implemented method of building a knowledge base for a legal topic classification system, the method comprising:

analyzing previously classified legal concepts to determine distinguishing features for each concept;
generating relevance scores for each feature in each training concept; and
storing features, topics, and relevance scores in a knowledge base, using an inverted index.

6. The method as set forth in claim 5, the step of generating relevance scores including the steps of:

converting features to terms;
generating, for each training concept, term frequency (TF) for each term, as number of occurrences of that term in that training concept;
generating, for each training concept, average term frequency of terms;

generating, for each training concept, document frequency (DF) for each term, as total number of training concepts in which term appears;

determining DBSIZE as total number of training concepts in knowledge base;

generating inverse document frequency (IDF) for each term; and

generating a relevance score for each term for each concept.

7. The method as set forth in claim 6, wherein the step of generating IDF is performed using the formula, $\log ((\text{DBSIZE} - \text{DF} + 0.5) / (\text{DF} + .05))$.

8. A computer-implemented method of processing an input concept from a document text to provide, from a topic scheme, a list of one or more topics that are relevant to the input concept, the method comprising:

analyzing the input concept to arrive at a set of distinguishing features;

converting candidate concept features to candidate terms;

searching a database of concepts, previously classified according to the topic scheme, for concepts similar to the input concept based on features;

ranking the similar concepts based on relevance score; and

voting on topics associated with the concepts within the database to form the list of topics relevant to the input concept.

9. The method as set forth in claim 8, the step of ranking including the steps of:

retrieving, for each training concept, relevance scores from a knowledge base for all candidate terms;

calculating total relevance score for each training concept, as a sum of candidate term relevance scores for that concept; and

sorting training concepts by total relevance scores.

10. The method as set forth in claim 9, the step of ranking further including, before the step of retrieving, the steps of:

sorting candidate terms by document frequency (DF) of each term, as number of knowledge base training concepts in which term occurs; and

reducing candidate term list to least common terms.

11. The method as set forth in claim 8, the step of voting including the steps of:
retrieving topics associated with each training concept from a knowledge base;
grouping training concepts and scores by associated topics;
calculating a total topic relevance score for each topic, as a sum of training concept scores for each topic; and

sorting topics by total topic relevance score to create a topic list.

12. The method as set forth in claim 11, further comprising, within a hierarchical topic scheme, the steps of:

grouping topics by tier;

weighting the topic list according to number of occurrences of each tier topic;
generating a final topic list using the weighted topic list; and
sorting the final topic list by tier.

13. The method as set forth in claim 11, the step of sorting including comparing each total topic relevance score to a threshold and eliminating from the topic list those topics having a total topic relevance score below the threshold.

14. The method as set forth in claim 11, the step of sorting including the steps of:
determining a number of times each topic occurs;
comparing the number to a threshold; and
eliminating from the topic list those topics having a number of occurrences below the threshold.

15. A computer-implemented method of processing an input concept from a document text to provide, from a topic scheme incorporating a plurality of training concepts, a list of one or more topics that are relevant to the input concept, the method comprising:
retrieving topics associated with the training concepts from a knowledge base, the training concepts having been previously classified and scored in accordance with the topic scheme;
grouping training concepts and scores by associated topics;

calculating a total topic relevance score for each topic, as a sum of training concept scores for each topic; and

sorting topics by total topic relevance score to create a topic list relevant to the input concept.

16. The method as set forth in claim 15, further comprising, within a hierarchical topic scheme, the steps of:

grouping topics by tier;

weighting the topic list according to number of occurrences of each tier topic;

generating a final topic list using the weighted topic list; and

sorting the final topic list by tier.

17. A computer-implemented method of processing an input concept from a document text to identify, within a knowledge base incorporating a plurality of training concepts, concepts similar to the input concept and to rank these similar concepts, the method comprising:

identifying features of the input concept as candidate terms;

retrieving, from the knowledge base, relevance scores for training concepts similar to the input concept;

calculating a total relevance score for each retrieved training concept, as a sum of candidate term relevance scores for that concept; and

sorting retrieved training concepts by total relevance scores.

18. A computer-implemented method of building a knowledge base for a legal topic classification system by identifying features within previously classified training concepts and generating relevance scores for these features, the method comprising the steps of:

converting the features into terms;

generating, for each training concept, term frequency (TF) for each term, as number of occurrences of that term in that training concept;

generating, for each training concept, average term frequency (AVE_TF) of terms;

generating, for each training concept, document frequency (DF) for each term, as total number of training concepts in which term appears;

determining training set DBSIZE as total number of training concepts in the knowledge base;

generating inverse document frequency (IDF) for each term; and

generating a relevance score for each term for each concept.

19. The method as set forth in claim 18, wherein when a length of a current concept, doclength, is greater than an average length of concepts in a set, aveDocLength, the relevance score is calculated using the formula $TFwt \times IDF$,
where $TFwt =$

$$\frac{TF + TF/AVE_TF}{TF + TF/AVE_TF + 2(\alpha + \beta \times (doclength - aveDocLength) / aveDocLength)}$$

and $IDF = \log ((DBSIZE - DF + 0.5) / (DF + .05))$.

20. The method as set forth in claim 18, wherein when a length of a current concept, doclength, is less than or equal to an average length of concepts in a set, aveDocLength, the relevance score is calculated using the formula $TFwt \times IDF$, where $TFwt =$

$$\frac{TF + TF/AVE_TF}{TF + TF/AVE_TF + 2(\alpha + \beta \times (aveDocLength - doclength + 1) / aveDocLength)}$$

and $IDF = \log ((DBSIZE - DF + 0.5) / (DF + .05))$.

21. A computer-implemented method of building a knowledge base for a legal topic classification system substantially as set forth in the drawings.

22. A computer-implemented method of processing an input concept from a document text substantially as set forth in the drawings.

23. A computer-implemented method of building a knowledge base for a legal topic classification system substantially as herein described with reference to any embodiment disclosed.

24. A computer-implemented method of processing an input concept from a document text substantially as herein described with reference to any embodiment disclosed.

END OF CLAIMS

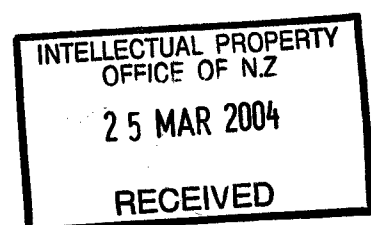


FIG. 1

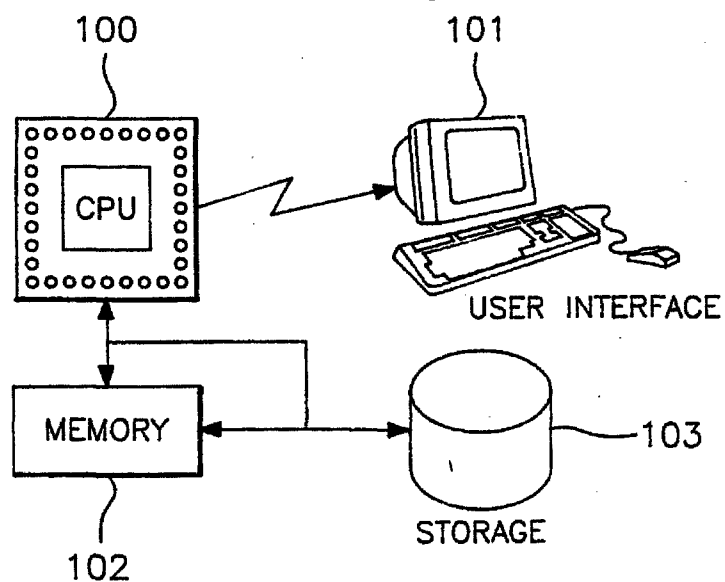


FIG. 2

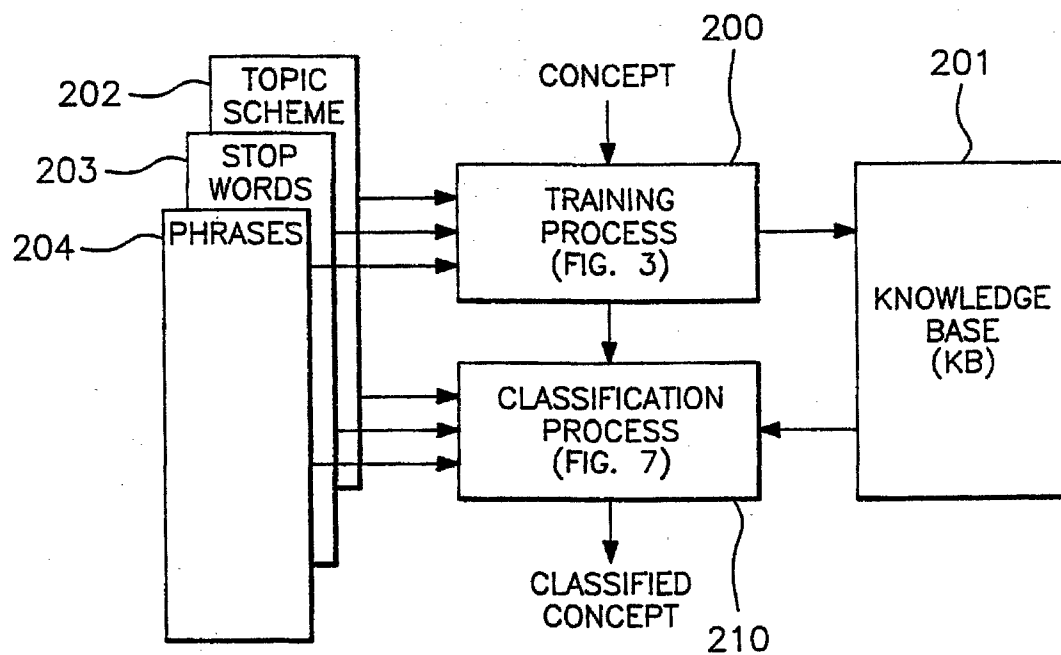


FIG. 3

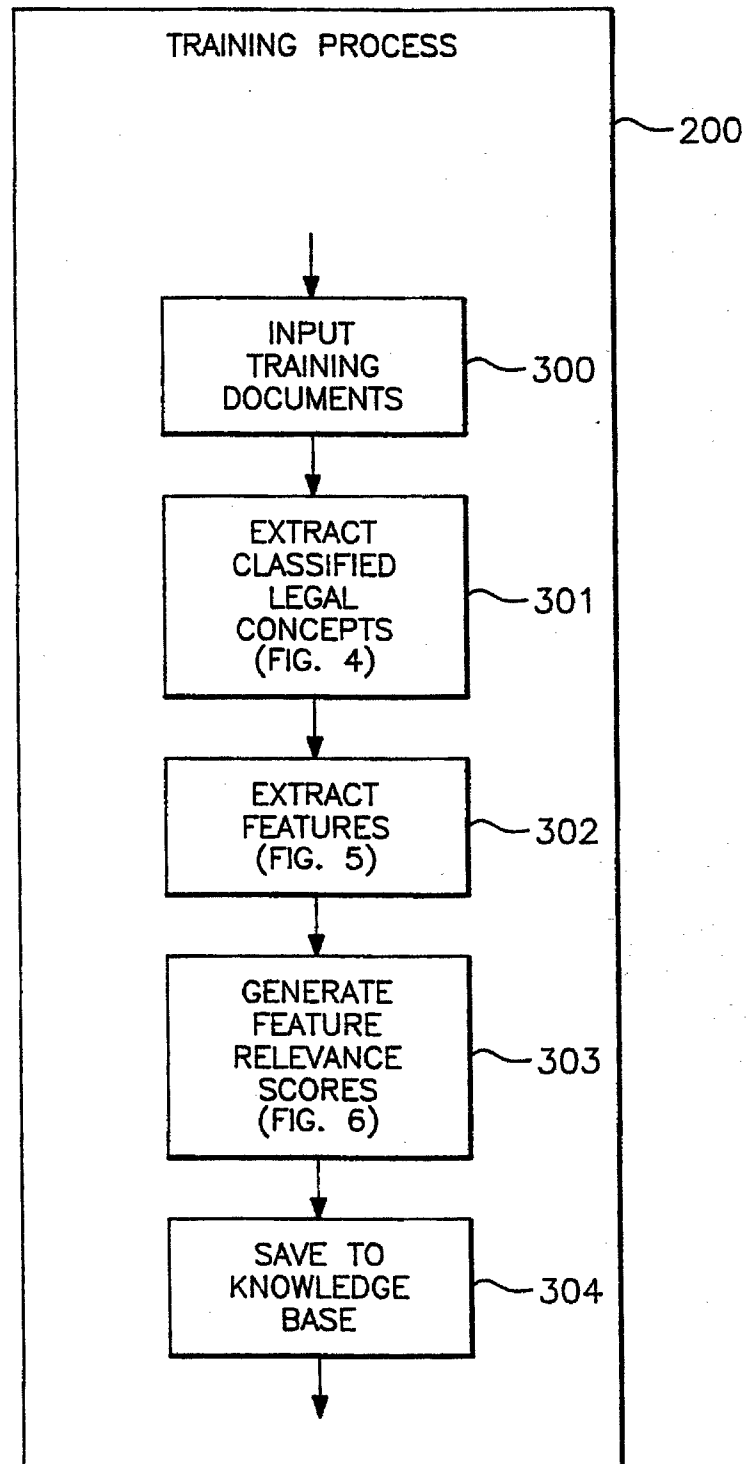


FIG. 4

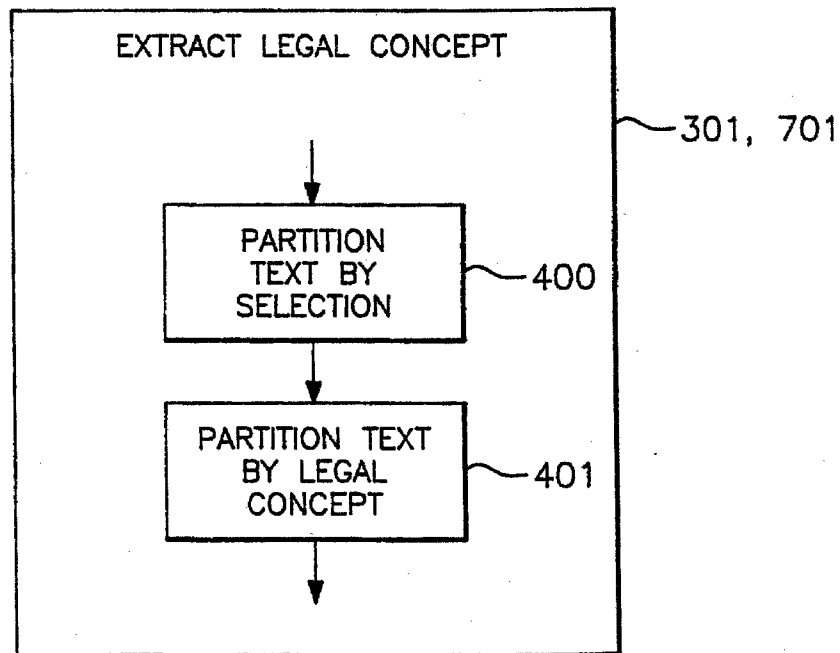


FIG. 5

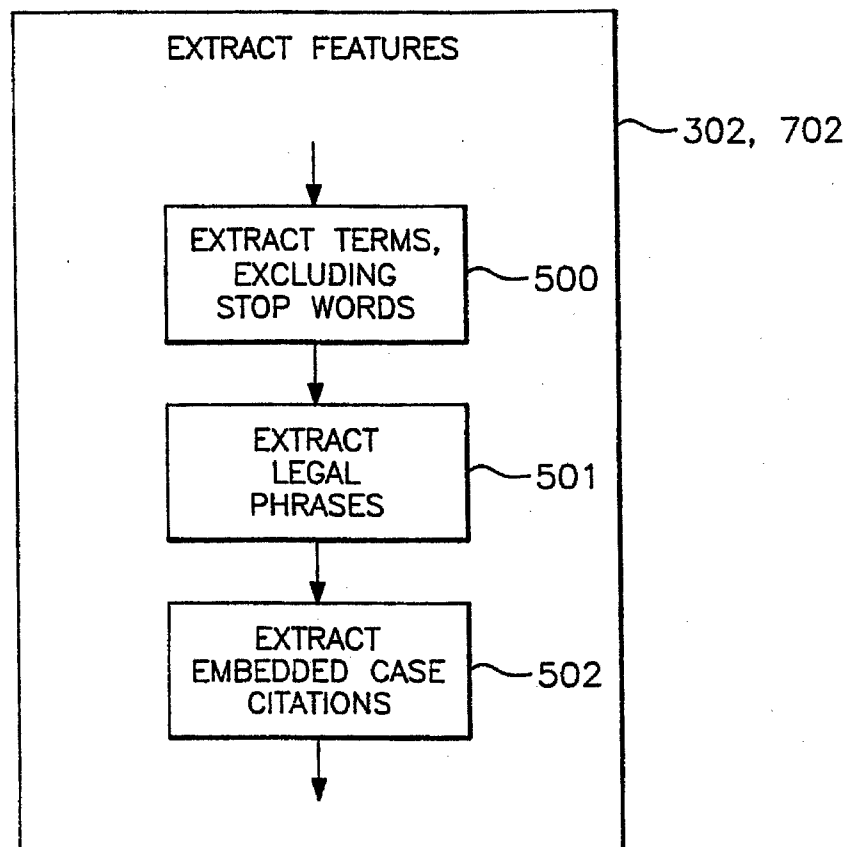


FIG. 6

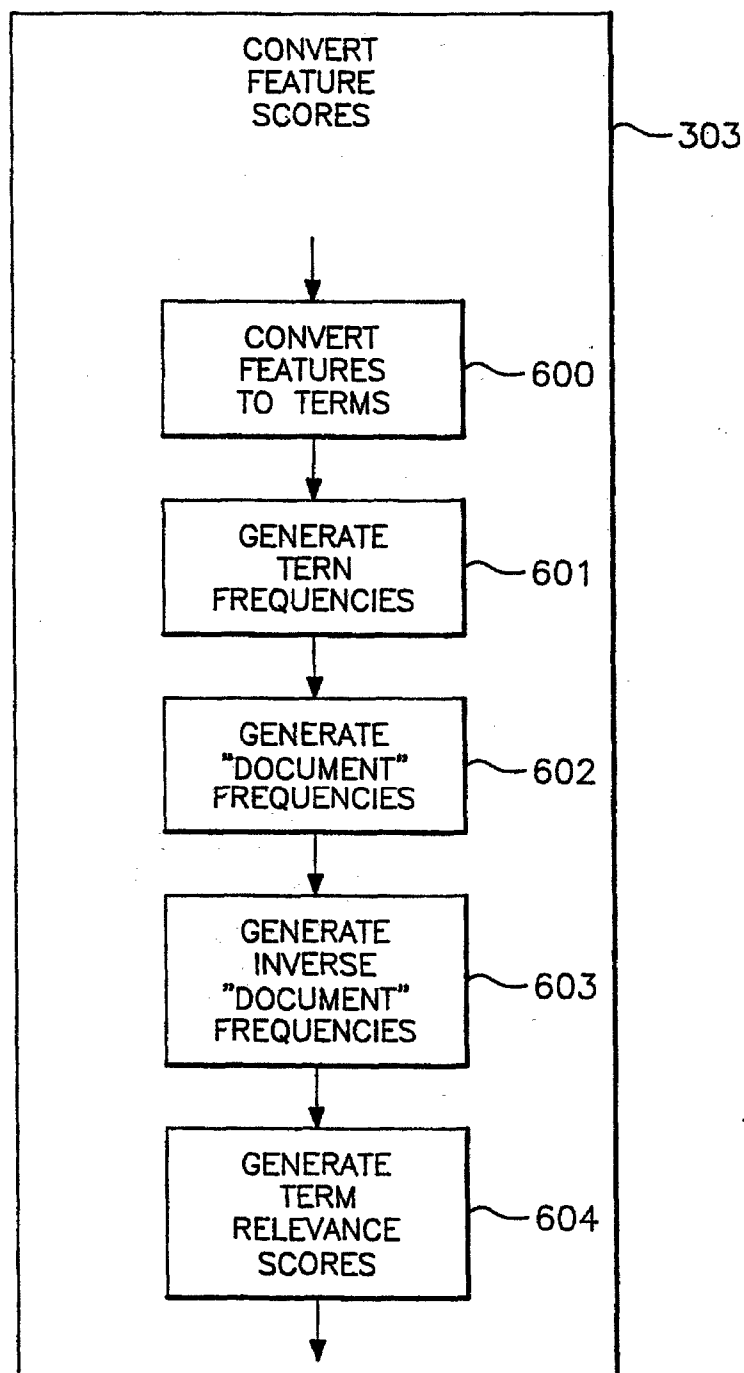


FIG. 7

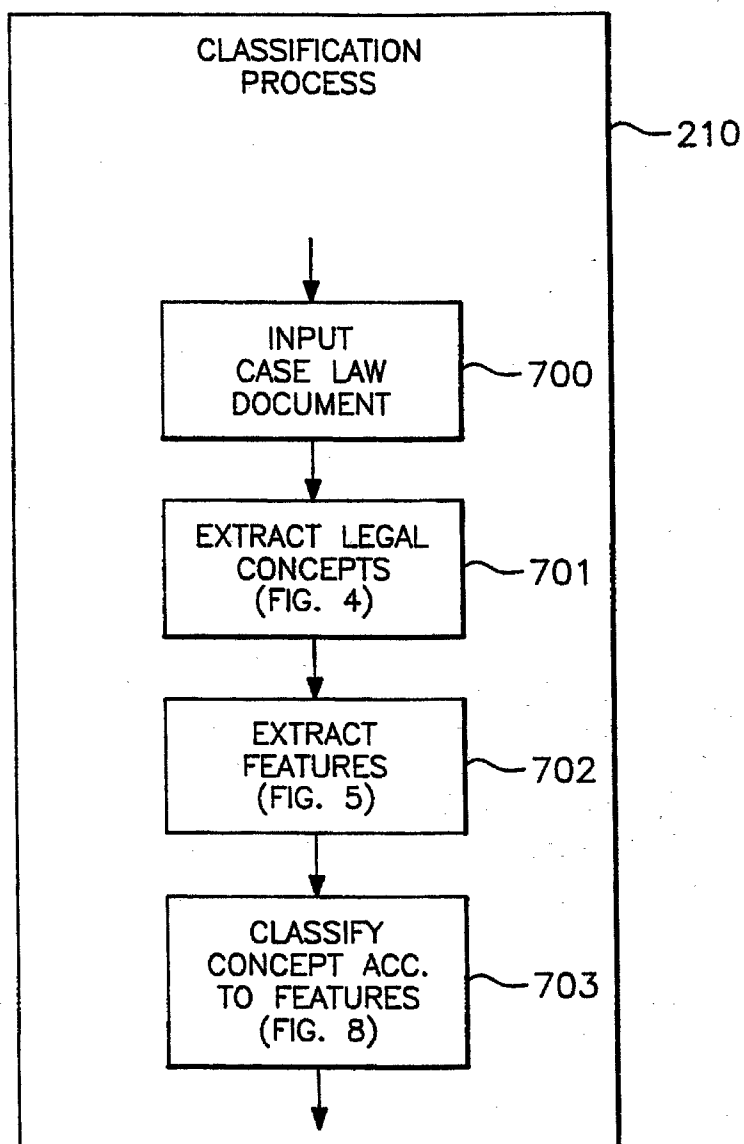


FIG. 8

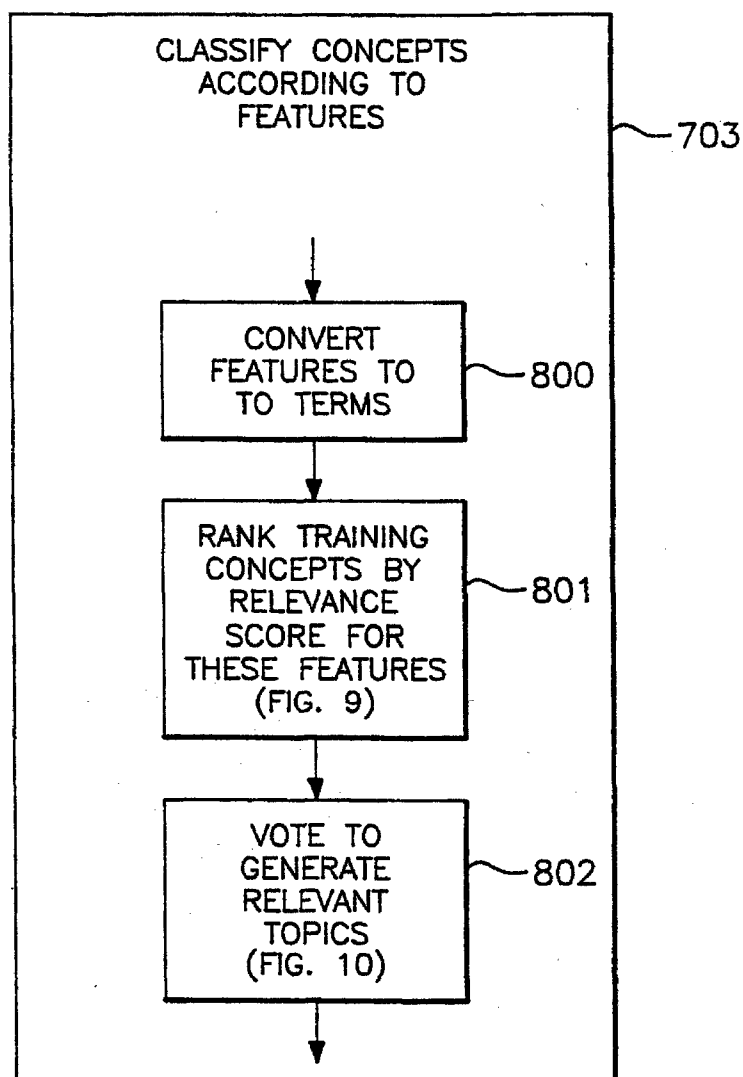


FIG. 9

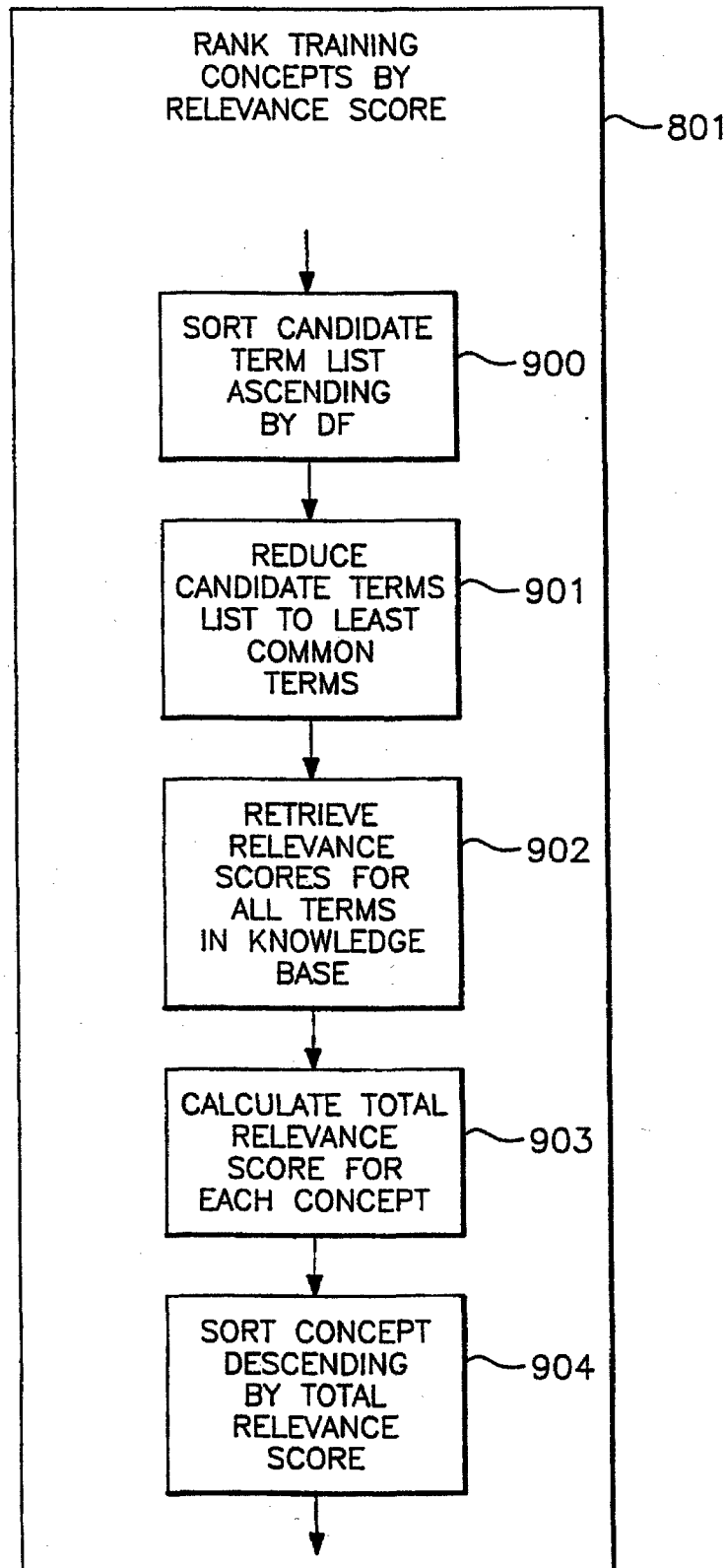
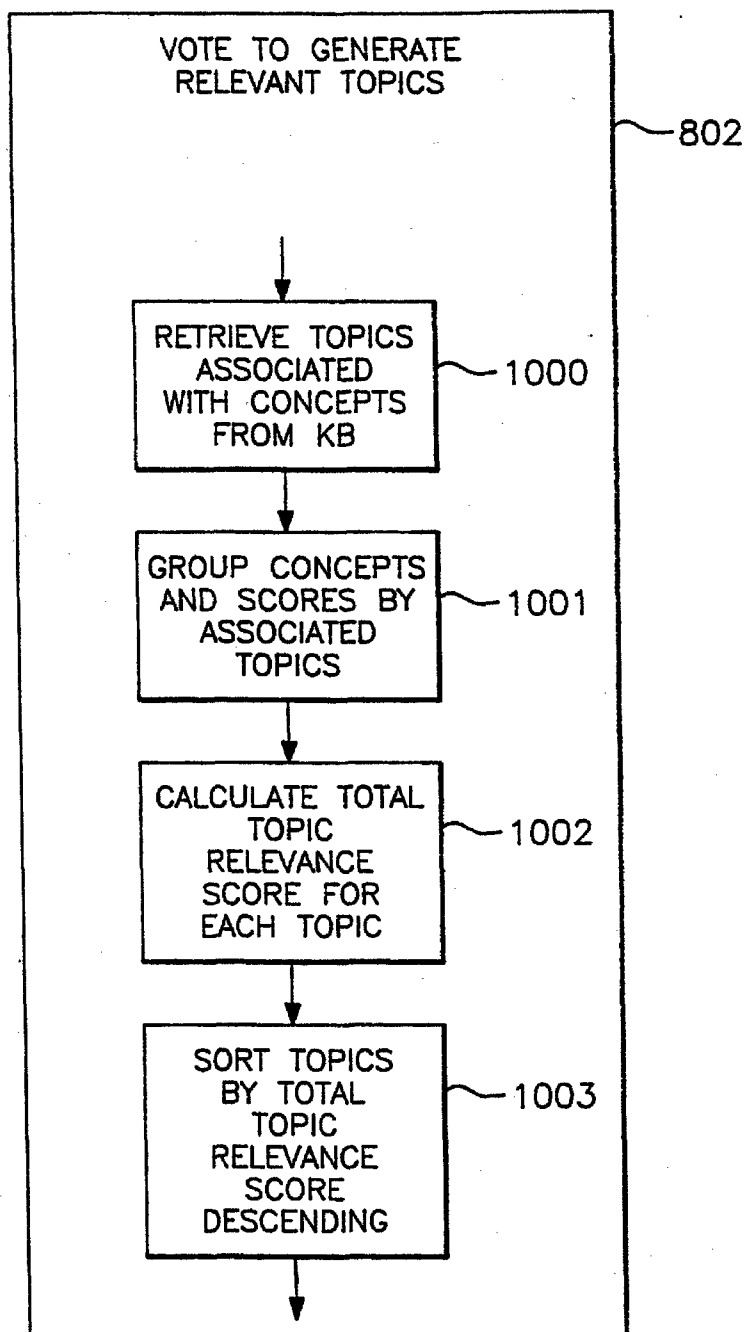


FIG. 10



END



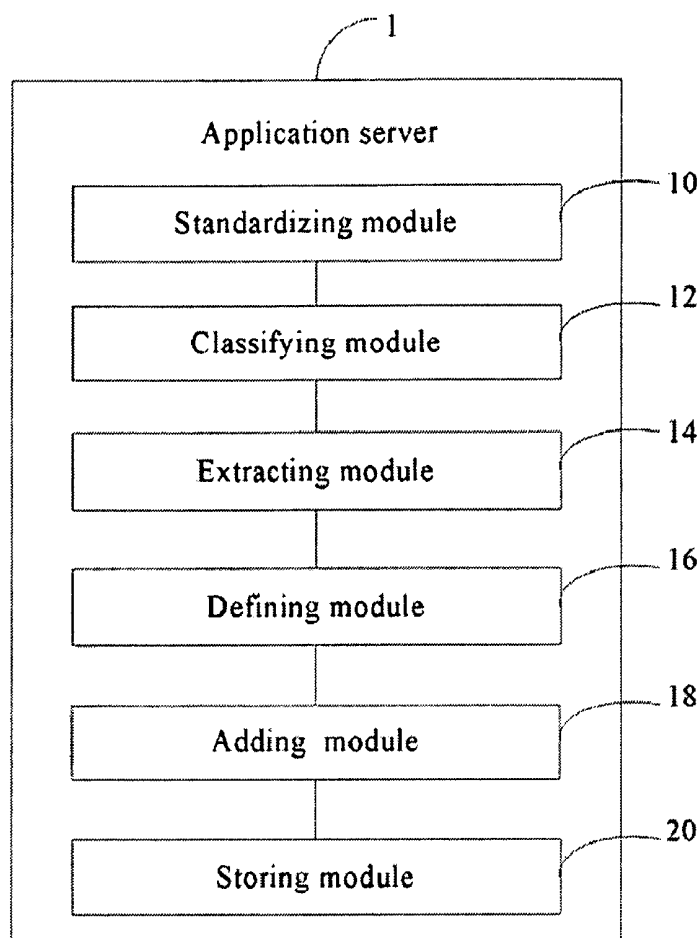
US 20080249980A1

(19) **United States**(12) **Patent Application Publication**
YU et al.(10) **Pub. No.: US 2008/0249980 A1**(43) **Pub. Date: Oct. 9, 2008**(54) **METHOD FOR CLASSIFYING TERMS OF
LEGAL DOCUMENTS**(75) Inventors: **Piao YU**, Tu-Cheng (TW);
Cheng-Yi HUANG, Tu-Cheng
(TW); **Yi-Hua TANG**, Shenzhen
(CN); **Xiao GU**, Shenzhen (CN)Correspondence Address:
PCE INDUSTRY, INC.
ATT. CHENG-JU CHIANG
458 E. LAMBERT ROAD
FULLERTON, CA 92835 (US)(73) Assignees: **HONG FU JIN PRECISION
INDUSTRY (ShenZhen) CO.,
LTD.**, Shenzhen City (CN); **HON
HAI PRECISION INDUSTRY
CO., LTD.**, Tu-Cheng (TW)(21) Appl. No.: **11/966,998**(22) Filed: **Dec. 29, 2007**(30) **Foreign Application Priority Data**

Apr. 6, 2007 (CN) 200710200406.4

Publication Classification(51) **Int. Cl.**
G06F 7/08 (2006.01)
G06F 17/30 (2006.01)
(52) **U.S. Cl.** **707/2; 707/E17.002**(57) **ABSTRACT**

A method for classifying terms and conditions of legal documents is provided. The method includes the steps of: standardizing terms and conditions of all of the legal documents stored in a database; classifying the all of the legal documents having standardized terms and conditions into different categories according to different areas of practices; extracting universal terms and conditions of each of the categories; extracting universal terms and conditions of all of the legal documents in all categories; and extracting categorical terms and conditions of each of the categories.



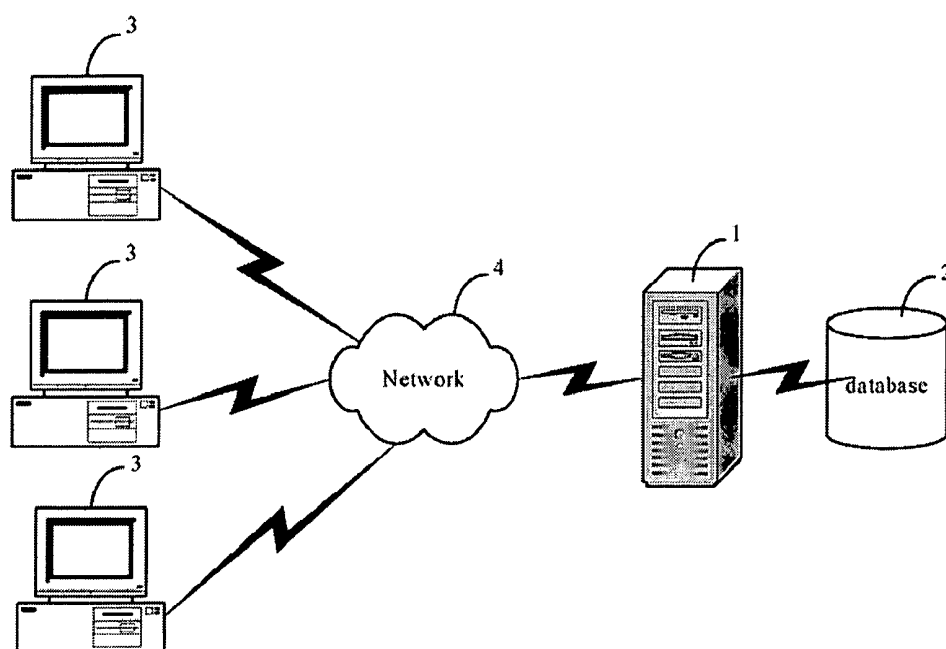


FIG. 1

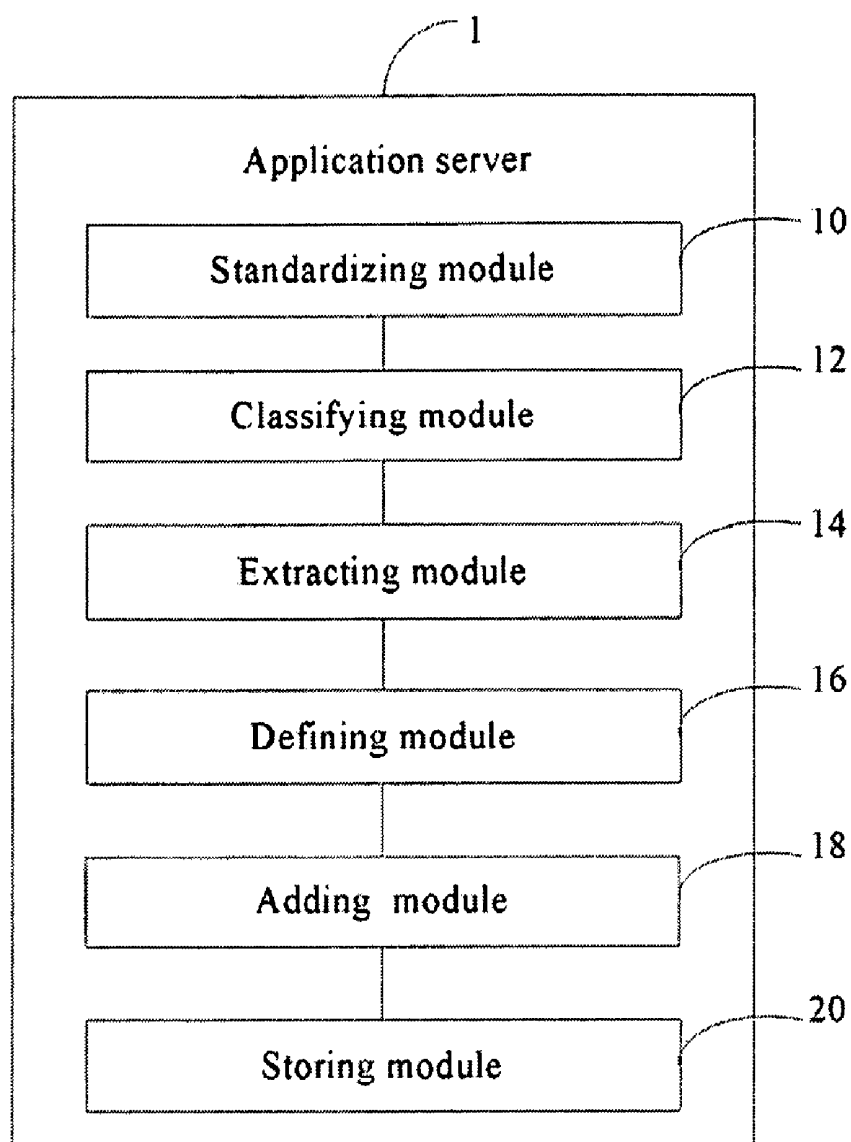


FIG. 2

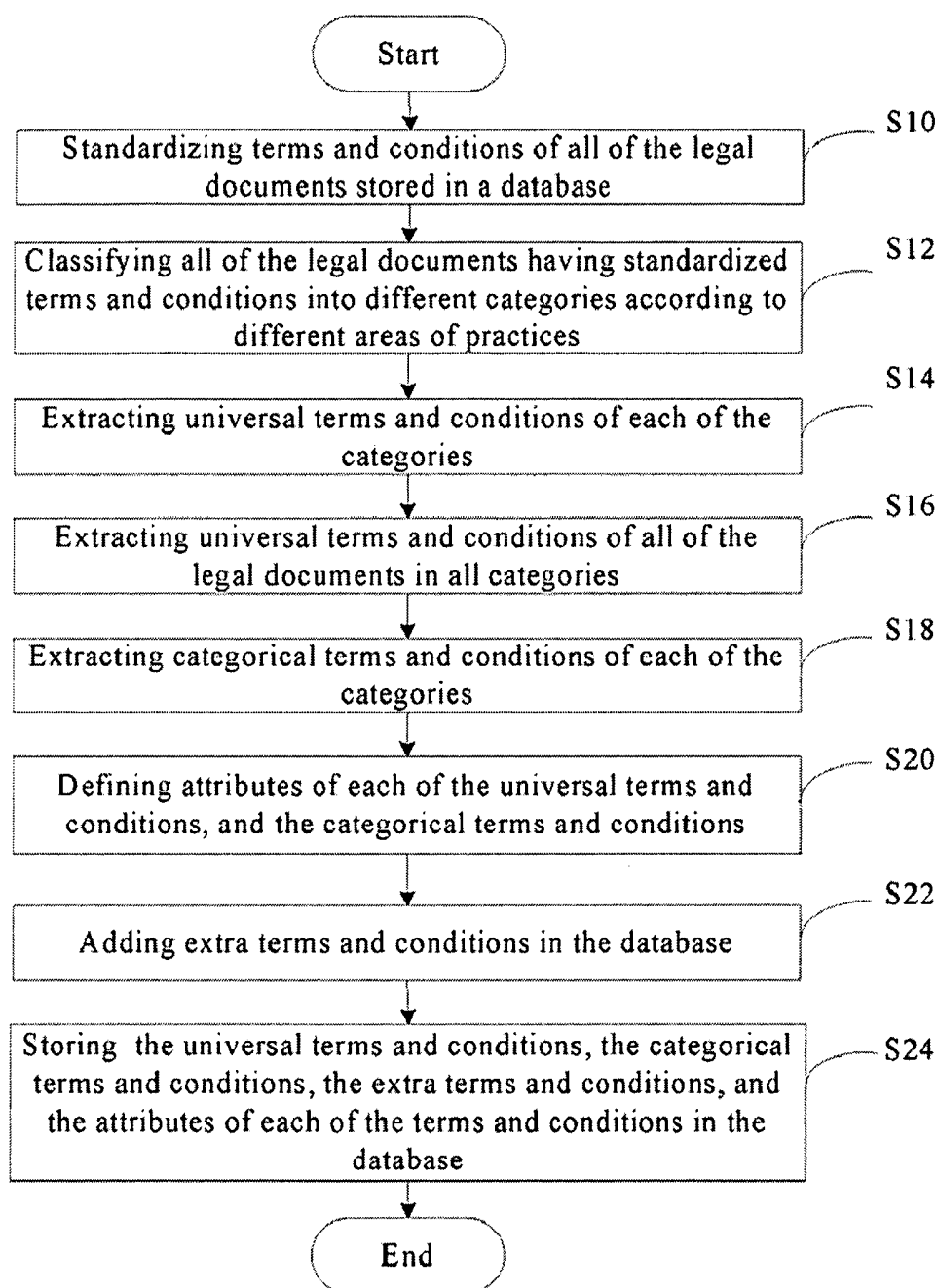


FIG. 3

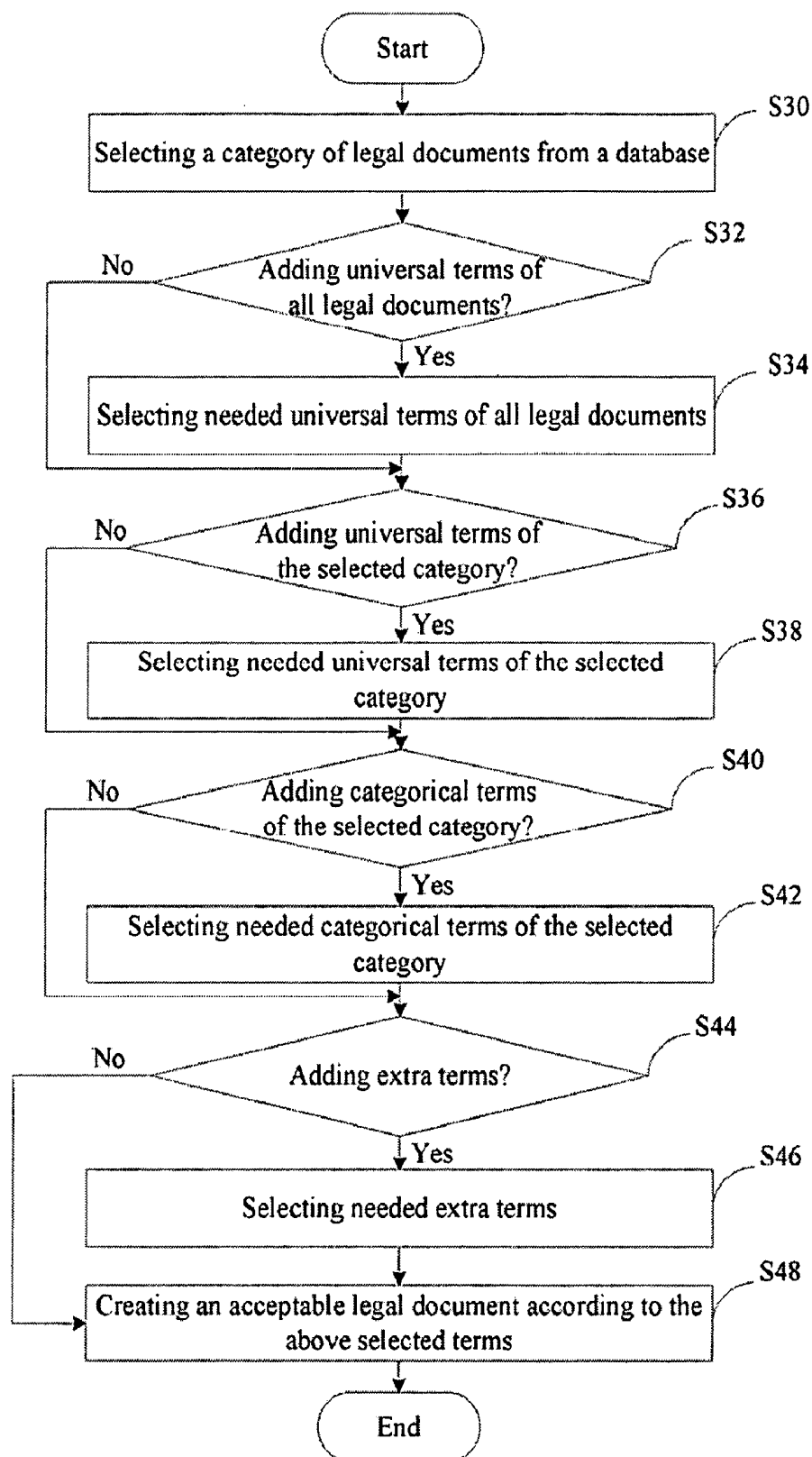


FIG.4

METHOD FOR CLASSIFYING TERMS OF LEGAL DOCUMENTS

BACKGROUND OF THE INVENTION

[0001] 1. Field of the Invention

[0002] The present invention relates to a method for classifying terms of legal documents.

[0003] 2. Description of Related Art

[0004] Legal entities use a number of tools to complete tasks regarding legal documentation. Modern commercial relationships involve multiple business partners and processes throughout the supply chain. This complexity affects terms and conditions entries corresponding to business liabilities, which relates to costs, shipments, payments, qualities, volume breaks, rebate policies, discounts, and so forth. Managing all of these requires a serious commitment of time and resources. Today, a typical enterprise has to associate with a variety of vendors, suppliers, service providers, and such by using contracts, agreements, or other types of legal documents.

[0005] Legal documents are used in almost all areas of commerce. Most conventional legal documents are reviewed, accepted, and signed using an inefficient, manually irritating process. This seriously affects the efficiency of the businesses. In recent years, software programs related to legal document management have been adopted and used by many organizations. However, interfaces provided by the software programs are often text-based and cumbersome to use. Furthermore, the software programs also lack functionalities that allow terms of legal documents to be efficiently updated and shared.

[0006] What is needed, therefore, is a method that can provide functions of classifying terms of legal documents throughout the stages of maintenance, and analysis.

SUMMARY OF THE INVENTION

[0007] A method for classifying terms of legal documents is provided. The method includes the steps of: standardizing terms and conditions of all of the legal documents stored in a database; classifying all of the legal documents having standardized terms and conditions into different categories according to different areas of practices; extracting universal terms and conditions of each of the categories; extracting universal terms and conditions of all of the legal documents in all categories; and extracting categorical terms and conditions of each of the categories.

[0008] Other advantages and novel features of the present invention will become more apparent from the following detailed description of preferred embodiments when taken in conjunction with the accompanying drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] FIG. 1 is a schematic diagram of a hardware configuration of a system for classifying terms of legal documents in accordance with a preferred embodiment;

[0010] FIG. 2 is a schematic diagram of main software function modules of an application server of FIG. 1;

[0011] FIG. 3 is a flowchart of a method for classifying terms of legal documents in accordance with a preferred embodiment; and

[0012] FIG. 4 is a flowchart of a method for a user to apply classified terms of FIG. 3 to generate an acceptable legal document.

DETAILED DESCRIPTION OF THE INVENTION

[0013] FIG. 1 is a schematic diagram of a hardware configuration of a system for classifying terms of legal documents (hereinafter, "the system") in accordance with a preferred embodiment. The system includes an application server 1, a database 2, and a plurality of clients 3 (only three shown). The application server 1 communicates with the clients 3 via a network 4. The network 4 may be an intranet, the Internet, or any other types of communications links. Each of the plurality of clients 3 is configured (i.e., structured and arranged) for providing an interactive interface that enables users to perform relevant operations so as to classify terms of the legal documents. The users may belong to different departments, such as a sales department, a purchasing department, a financial department, a legal department, or any other affiliates of an enterprise and/or a contracting party of the enterprise.

[0014] The application server 1 further connects to the database 2. The database 2 is configured for storing related data, such as legal documents and different terms of the legal documents. In the preferred embodiment, the legal documents are contracts. In other embodiments, the legal documents may be agreements, and indictments, etc.

[0015] FIG. 2 is a schematic diagram of main software function modules of the application server 1. The application server 1 may include a standardizing module 10, a classifying module 12, an extracting module 14, a defining module 16, an adding module 18, and a storing module 20.

[0016] The standardizing module 10 is configured for standardizing terms and conditions of all of legal documents stored in the database 2. The standardized terms and conditions in a particular category of legal documents may have the same expressions.

[0017] The classifying module 12 is configured for classifying all of the legal documents having standardized terms and conditions into different categories according to different areas of practices. For example, the legal documents (i.e., contracts) may be classified into the following categories: lease agreements, project contracts, service contracts, employment contracts, and business contracts, etc.

[0018] The extracting module 14 is configured for extracting universal terms and conditions of each of the categories. The universal terms and conditions of the category are applicable to all of the legal documents in the same category.

[0019] The extracting module 14 is also configured for extracting universal terms and conditions of all the categories according to the universal terms of each category. The universal terms and conditions of all the categories are selected from the universal terms and conditions of each of the categories and are applicable to each of the legal documents regardless of the category of the legal document.

[0020] The extracting module 14 is further configured for extracting categorical terms and conditions of each of the categories. The categorical terms and conditions are applicable to a particular legal document in a category.

[0021] The defining module 16 is configured for defining attributes of each of the universal terms, and conditions and the categorical terms and conditions. The attributes of each of the universal terms and conditions, and the categorical terms

and conditions may be defined as a fixed term and condition or a modifiable term and condition.

[0022] The adding module 18 is configured for adding extra terms and conditions in the database 2. The extra terms and conditions can be inserted into the universal terms and conditions, and the categorical terms and conditions so as to expand the terms and conditions.

[0023] The storing module 20 is configured for storing the universal terms and conditions, the categorical terms and conditions, extra terms and conditions, and the attributes of each of terms and conditions in the database 2.

[0024] FIG. 3 is a flowchart of a method for classifying terms of legal documents in accordance with a preferred embodiment. In step S10, the standardizing module 10 standardizes terms and conditions of all of the legal documents stored in the database 2.

[0025] In step S12, the classifying module 12 classifies all of the legal documents having standardized terms and conditions into different categories according to different areas of practices.

[0026] In step S14, the extracting module 14 extracts universal terms and conditions of each of the categories. The universal terms and conditions of a category are applicable to all of the legal documents in the same category.

[0027] In step S16, the extracting module 14 extracts universal terms and conditions of all categories according to the universal terms and conditions of each of the categories. The universal terms and conditions of all of the legal documents are selected from the universal terms and conditions of each of the categories and are applicable to each legal document regardless of the category of the legal document.

[0028] In step S18, the extracting module 14 extracts categorical terms and conditions of each of the categories. The categorical terms and conditions are applicable to a particular legal document in a category.

[0029] In step S20, the defining module 16 defines attributes of each of the universal terms and conditions, and the categorical terms and conditions.

[0030] In step S22, the adding module 18 adds extra terms and conditions in the database 2.

[0031] In step S24, the storing module 20 stores the universal terms and conditions, the categorical terms and conditions, the extra terms and conditions, and the attributes of each of the terms and conditions in the database 2.

[0032] FIG. 4 is a flowchart of a method for a user to apply the classified terms of FIG. 3 to generate an acceptable legal document. In step S30, the user selects a category of legal documents to which the legal document to be generated belongs from the database 2.

[0033] In step S32, the user determines whether there is need to add universal terms and conditions of all of the legal documents to the legal document.

[0034] In step S34, the user selects needed universal terms and conditions of all of the legal documents if there is need to add universal terms and conditions of all of the legal documents.

[0035] In step S36, the user determines whether there is need to add universal terms and conditions of the selected category to the legal document.

[0036] In step S38, the user selects needed universal terms and conditions of the selected category if there is need to add universal terms and conditions of the selected category.

[0037] In step S40, the user determines whether there is need to add categorical terms and conditions of the selected category to the legal document.

[0038] In step S42, the user selects needed categorical terms and conditions of the selected category if there is need to add categorical terms and conditions of the selected category.

[0039] In step S44, the user determines whether there is need to add extra terms and conditions to the legal document.

[0040] In step S46, the user selects needed extra terms and conditions if there is need to add extra terms and conditions.

[0041] In step S48, the system creates an acceptable legal document according to the above selected terms.

[0042] It should be emphasized that the above-described embodiments, particularly, any “preferred” embodiments, are merely possible examples of implementations, merely set forth for a clear understanding of the principles of the invention. Many variations and modifications may be made to the above-described preferred embodiment(s) without departing substantially from the spirit and principles of the invention. All such modifications and variations are intended to be included herein within the scope of this disclosure and the above-described preferred embodiment(s) and protected by the following claims.

What is claimed is:

1. A method for classifying terms of legal documents, the method comprising:

standardizing terms and conditions of all of the legal documents stored in a database;
classifying all of the legal documents having standardized terms and conditions into different categories according to different areas of practices;
extracting universal terms and conditions of each of the categories;
extracting universal terms and conditions of all of the legal documents in all categories; and
extracting categorical terms and conditions of each of the categories.

2. The method according to claim 1, further comprising:
defining attributes of each of the universal terms and conditions, and the categorical terms and conditions;
adding extra terms and conditions in the database; and
storing the universal terms and conditions, the categorical terms and conditions, the extra terms and conditions, and the attributes of each of the terms and conditions in the database.

3. The method according to claim 1, wherein the universal terms and conditions of a category are applicable to all of the legal documents in the same category.

4. The method according to claim 1, wherein the universal terms and conditions of all legal documents are selected from the universal terms and conditions of each category and are applicable to each legal document regardless of the category of the legal document.

5. The method according to claim 1, wherein the categorical terms and conditions of each category are applicable to a particular legal document in a category.

6. The method according to claim 1, further comprising a step of applying the classified terms and conditions to generate an acceptable legal document.

* * * * *



US 20200151392A1

(19) **United States**(12) **Patent Application Publication**
Crabtree et al.(10) **Pub. No.: US 2020/0151392 A1**(43) **Pub. Date: May 14, 2020**(54) **SYSTEM AND METHOD AUTOMATED
ANALYSIS OF LEGAL DOCUMENTS
WITHIN AND ACROSS SPECIFIC FIELDS**(71) Applicant: **QOMPLX, Inc.**, Reston, VA (US)(72) Inventors: **Jason Crabtree**, Vienna, VA (US);
Andrew Sellers, Monument, CO (US)(21) Appl. No.: **16/654,309**(22) Filed: **Oct. 16, 2019**

which is a continuation-in-part of application No. 15/141,752, filed on Apr. 28, 2016, which is a continuation-in-part of application No. 15/091,563, filed on Apr. 5, 2016, now Pat. No. 10,204,147, which is a continuation-in-part of application No. 14/986,536, filed on Dec. 31, 2015, now Pat. No. 10,210,255, which is a continuation-in-part of application No. 14/925,974, filed on Oct. 28, 2015, Continuation-in-part of application No. 15/847,443, filed on Dec. 19,

(Continued)

Publication Classification(51) **Int. Cl.**

G06F 40/30 (2006.01)
G06F 16/25 (2006.01)
G06F 16/93 (2006.01)
G06Q 50/18 (2006.01)
G06N 5/04 (2006.01)
G06N 5/02 (2006.01)
G06F 40/284 (2006.01)
G06F 40/295 (2006.01)

(52) **U.S. Cl.**

CPC **G06F 40/30** (2020.01); **G06F 16/254**
(2019.01); **G06F 16/93** (2019.01); **G06F**
40/295 (2020.01); **G06N 5/04** (2013.01);
G06N 5/02 (2013.01); **G06F 40/284**
(2020.01); **G06Q 50/18** (2013.01)

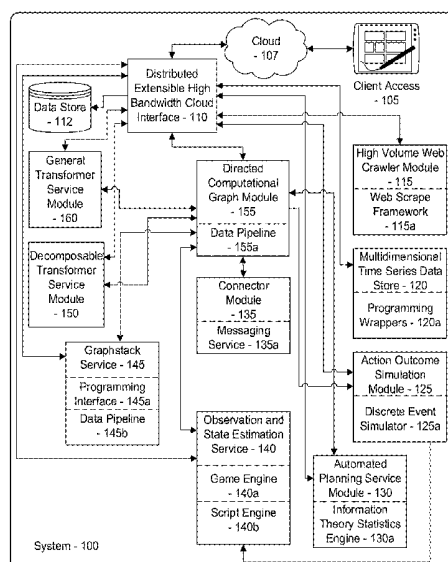
Related U.S. Application Data

(63) Continuation-in-part of application No. 15/847,443, filed on Dec. 19, 2017, which is a continuation-in-part of application No. 15/790,457, filed on Oct. 23, 2017, which is a continuation-in-part of application No. 15/790,327, filed on Oct. 23, 2017, which is a continuation-in-part of application No. 15/141,752, filed on Apr. 28, 2016, which is a continuation-in-part of application No. 15/091,563, filed on Apr. 5, 2016, now Pat. No. 10,204,147, which is a continuation-in-part of application No. 14/986,536, filed on Dec. 31, 2015, now Pat. No. 10,210,255, which is a continuation-in-part of application No. 14/925,974, filed on Oct. 28, 2015, Continuation-in-part of application No. 15/847,443, filed on Dec. 19, 2017, which is a continuation-in-part of application No. 15/489,716, filed on Apr. 17, 2017, which is a continuation-in-part of application No. 15/409,510, filed on Jan. 18, 2017, which is a continuation-in-part of application No. 15/379,899, filed on Dec. 15, 2016, which is a continuation-in-part of application No. 15/376,657, filed on Dec. 13, 2016, now Pat. No. 10,402,906, which is a continuation-in-part of application No. 15/237,625, filed on Aug. 15, 2016, now Pat. No. 10,248,910, which is a continuation-in-part of application No. 15/206,195, filed on Jul. 8, 2016, which is a continuation-in-part of application No. 15/186,453, filed on Jun. 18, 2016, which is a continuation-in-part of application No. 15/166,158, filed on May 26, 2016,

(57)

ABSTRACT

A system for automated analysis of legal documents within and across different fields is constructed using a computer system comprising at least one memory, at least one processor, and at least a first plurality of programming instructions stored in the at least one memory and operating on the at least one processor configured to allow the operation on the computer system of additional programming instructions, an extraction processor to identify, extract knowledge from data contained in the legal document and transform it into a common data form. The analysis processor develop a local and global knowledge graphs containing the key entities, relationships and concepts encoded in the text.



Related U.S. Application Data

2017, which is a continuation-in-part of application No. 15/790,457, filed on Oct. 23, 2017, which is a continuation-in-part of application No. 15/790,327, filed on Oct. 23, 2017, which is a continuation-in-part of application No. 15/616,427, filed on Jun. 7, 2017, which is a continuation-in-part of application No. 14/925,974, filed on Oct. 28, 2015, Continuation-in-part of application No. 15/847,443, filed on Dec. 19,

2017, which is a continuation-in-part of application No. 15/616,427, filed on Jun. 7, 2017, which is a continuation-in-part of application No. 14/925,974, filed on Oct. 28, 2015.

- (60) Provisional application No. 62/568,298, filed on Oct. 4, 2017, provisional application No. 62/568,291, filed on Oct. 4, 2017, provisional application No. 62/568,298, filed on Oct. 4, 2017, provisional application No. 62/568,291, filed on Oct. 4, 2017.

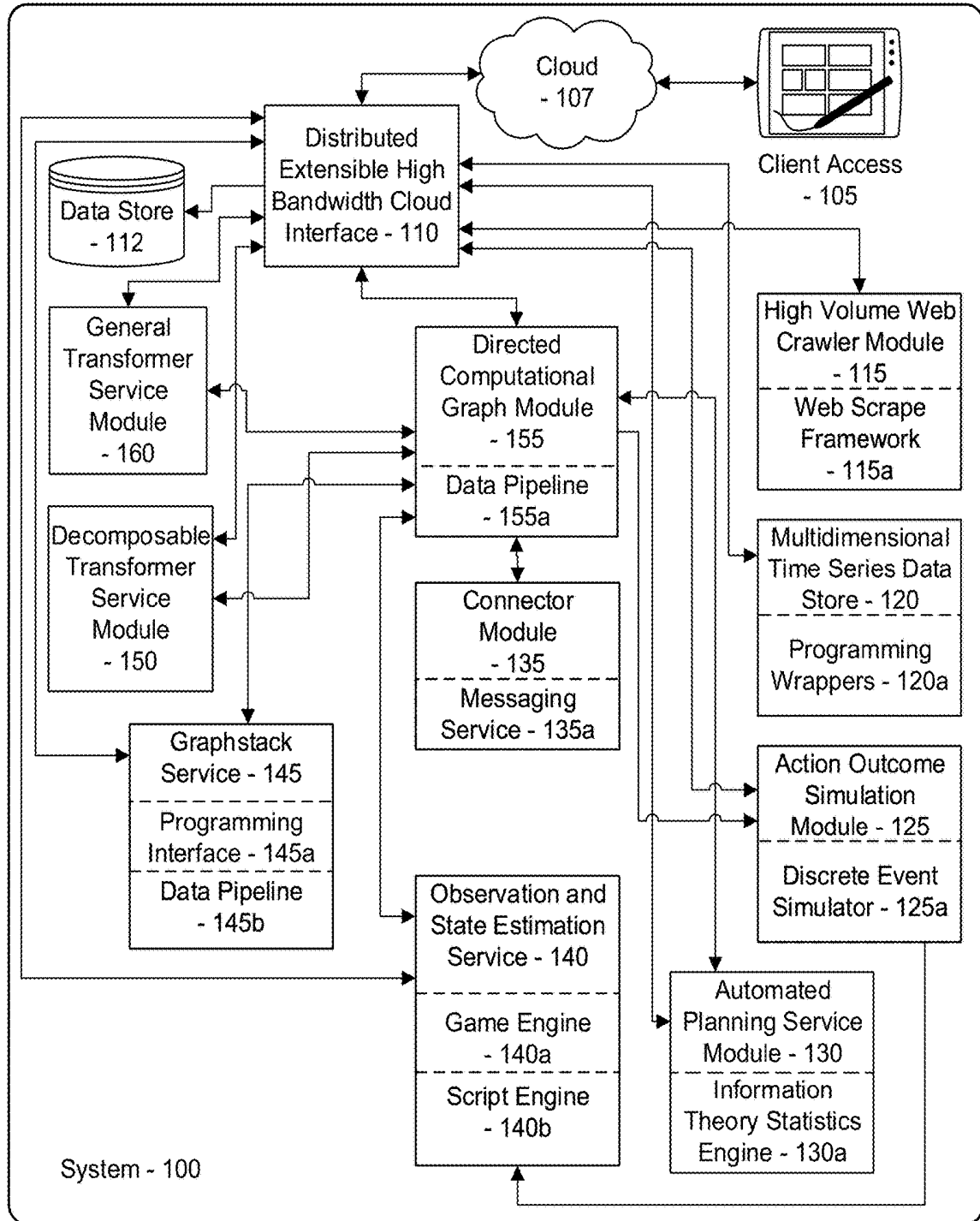
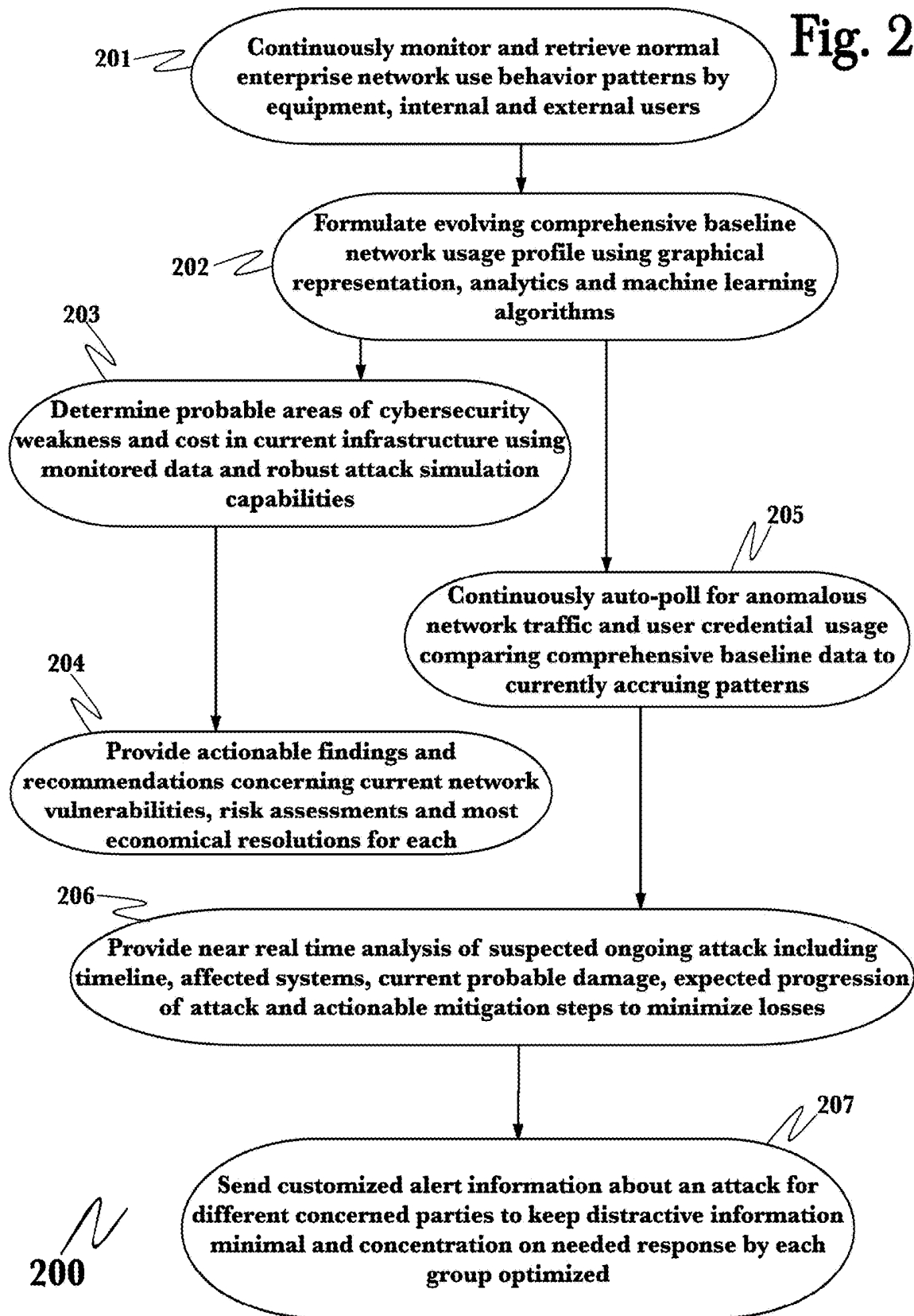


Fig. 1



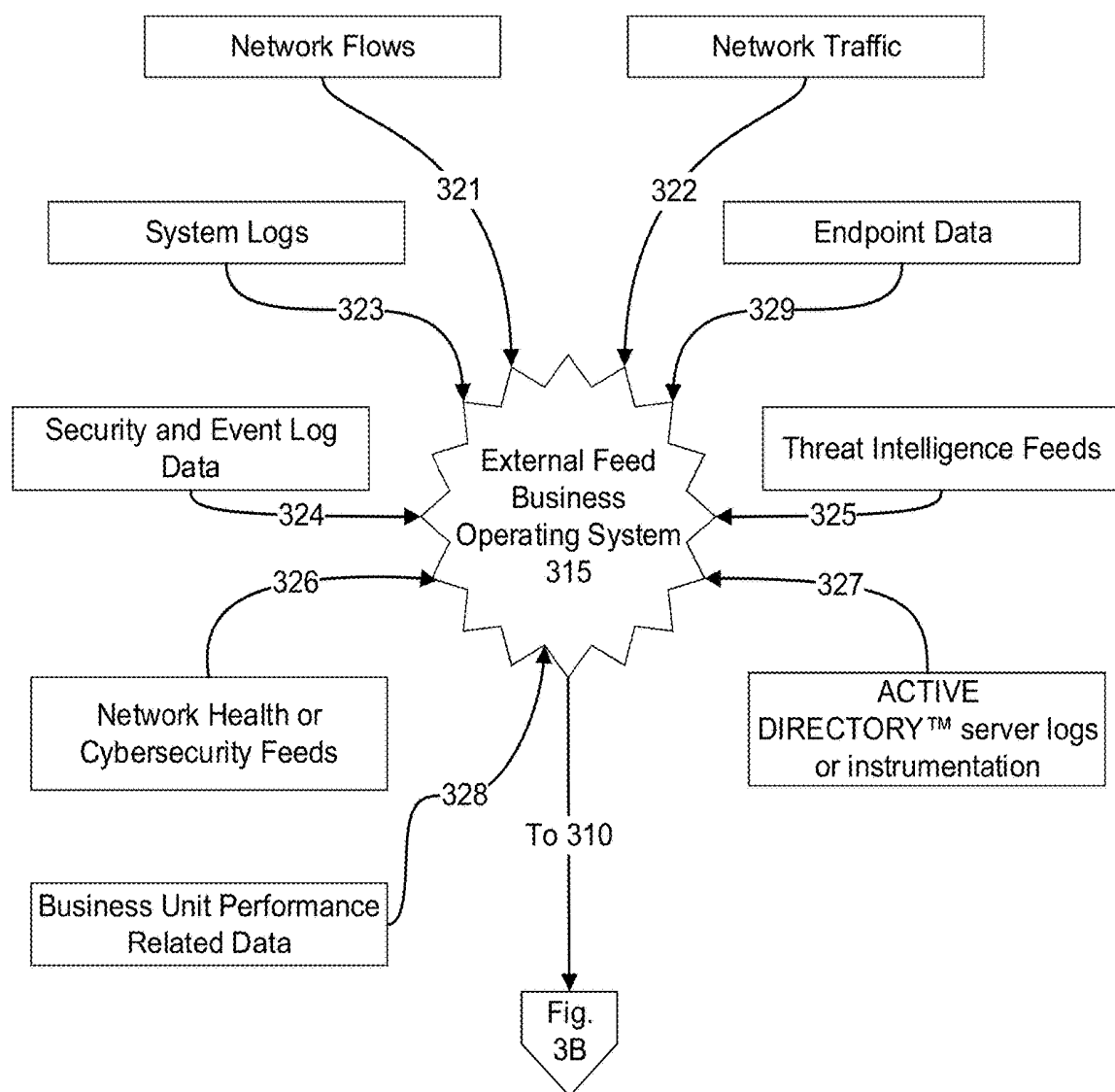
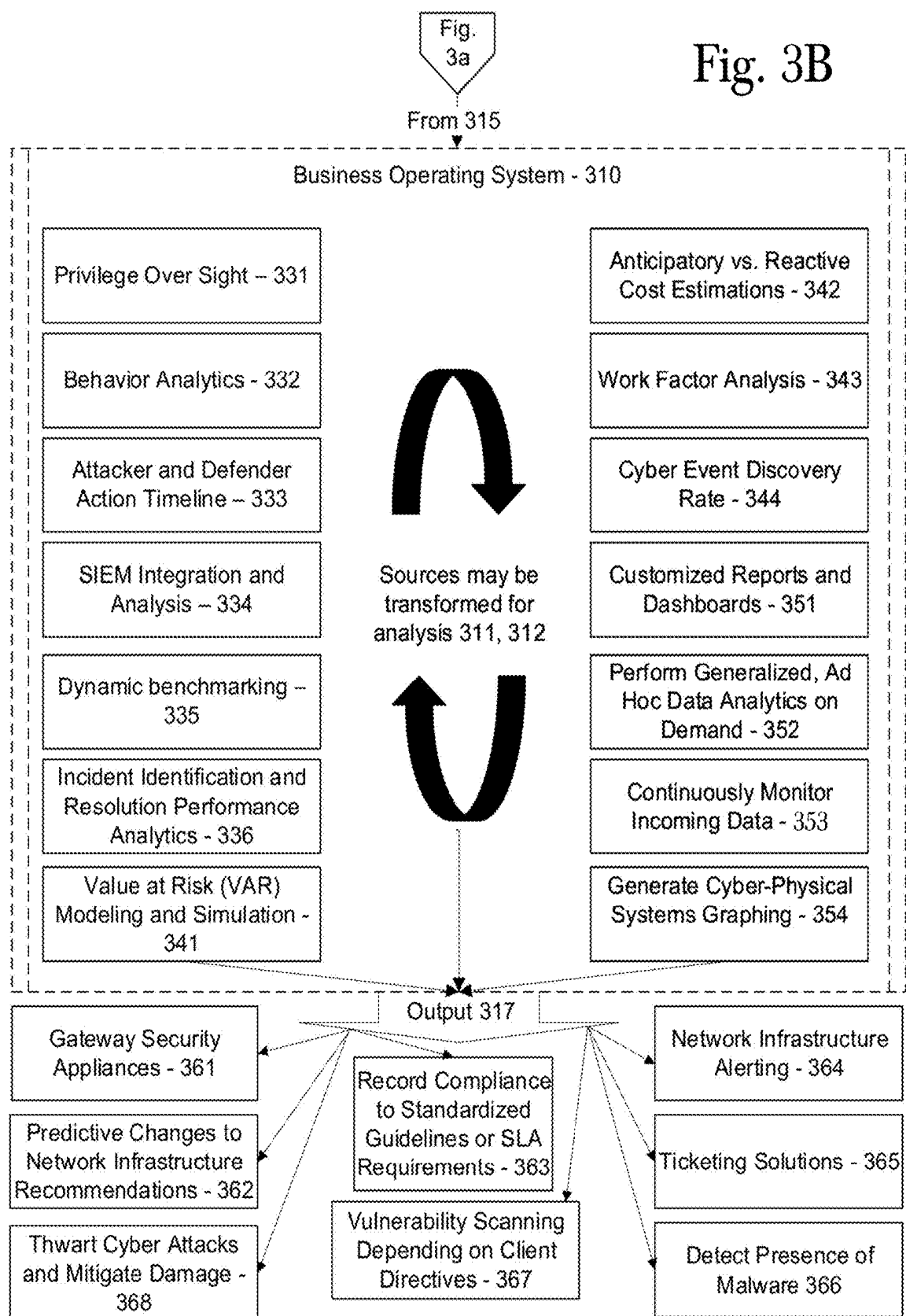
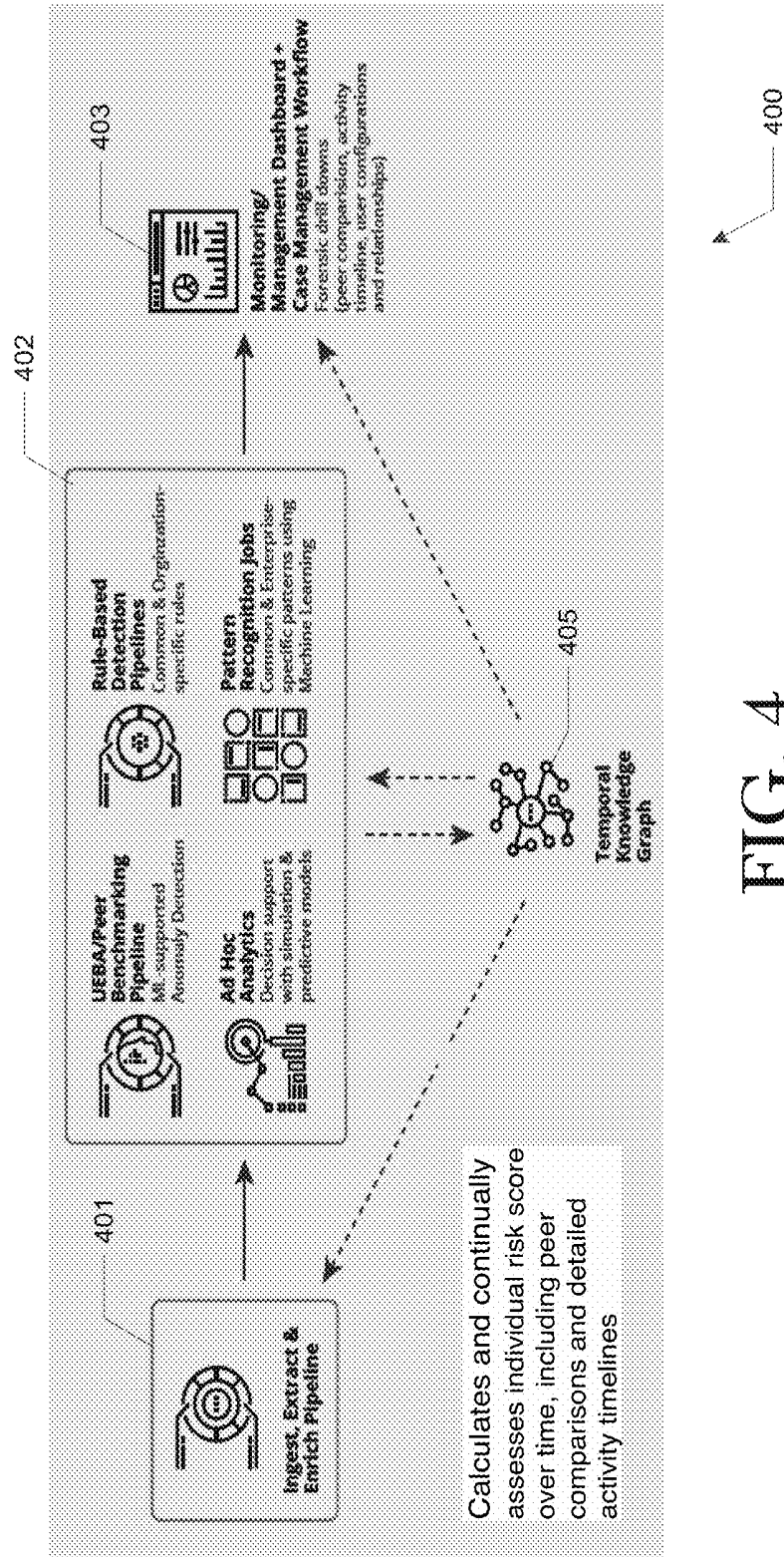


Fig. 3A





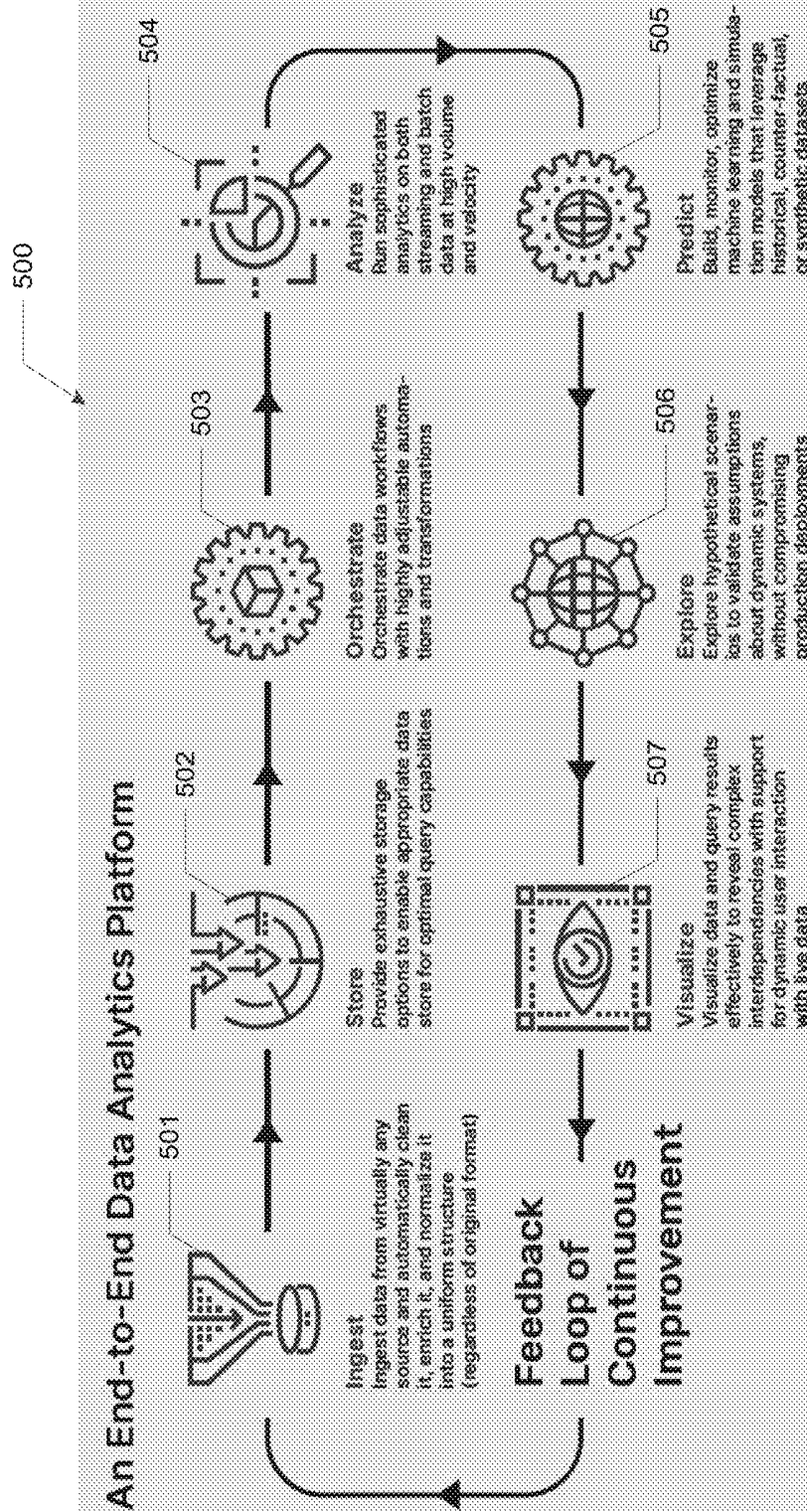


FIG. 5

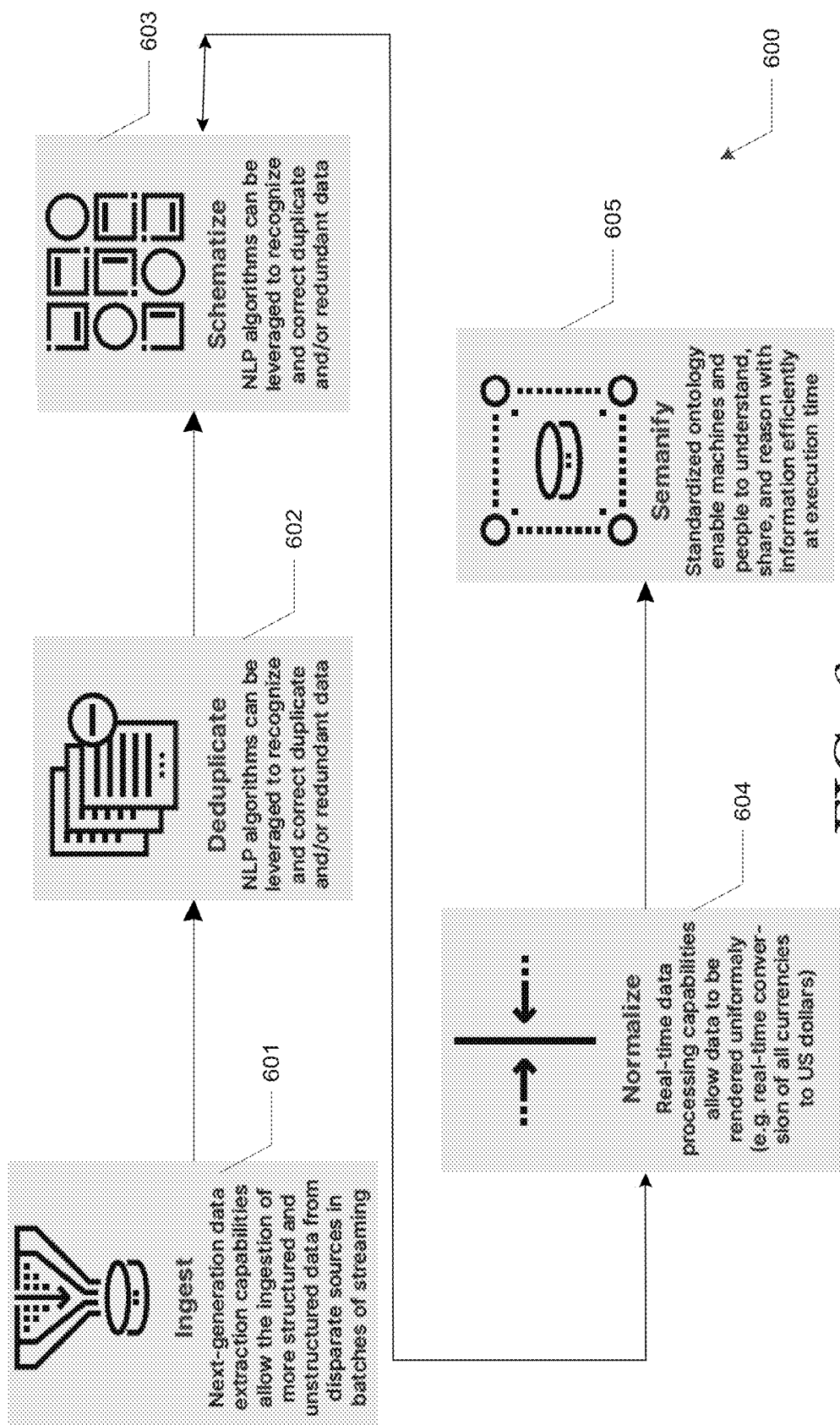


FIG. 6

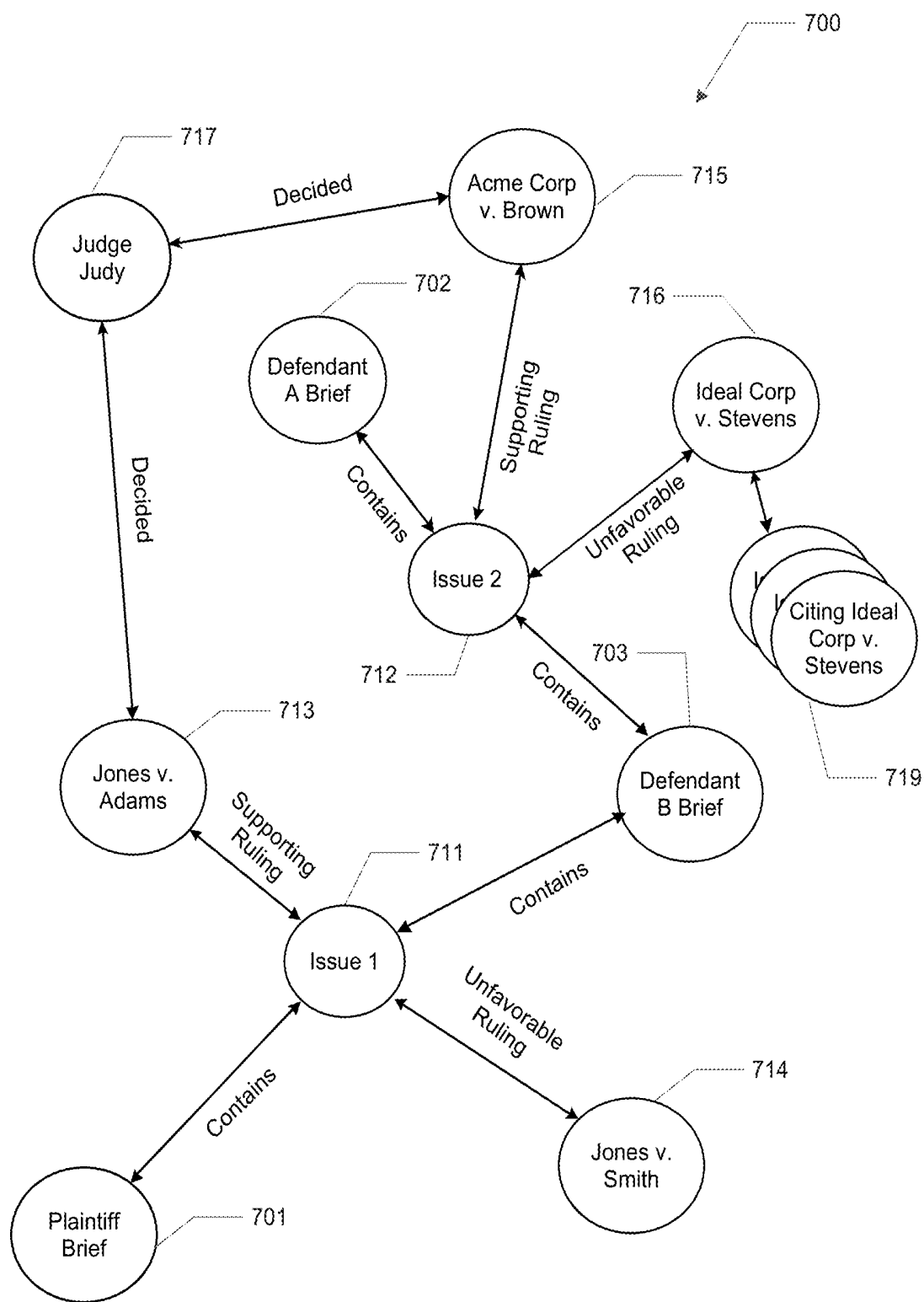


FIG. 7A

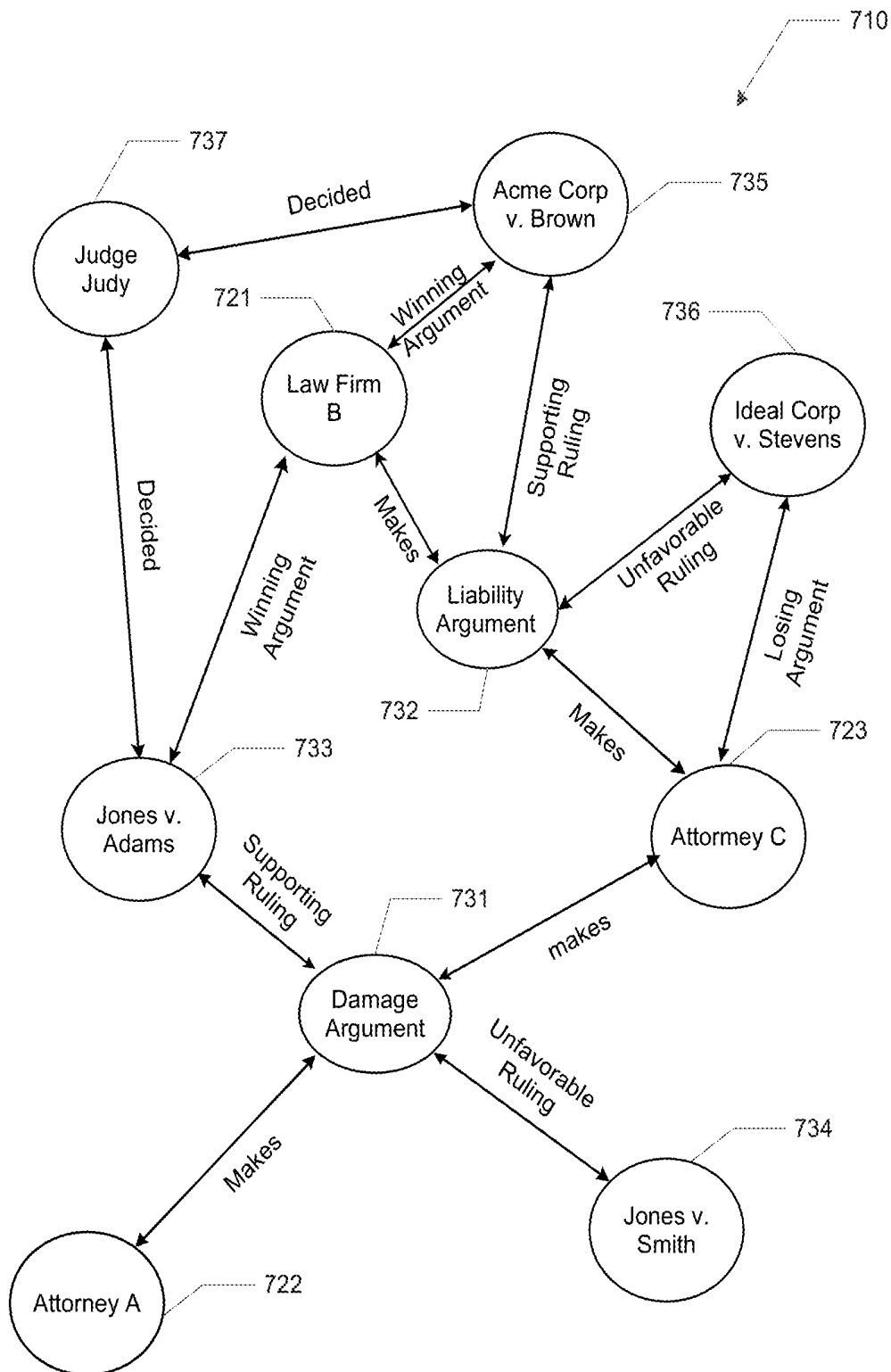


FIG. 7B

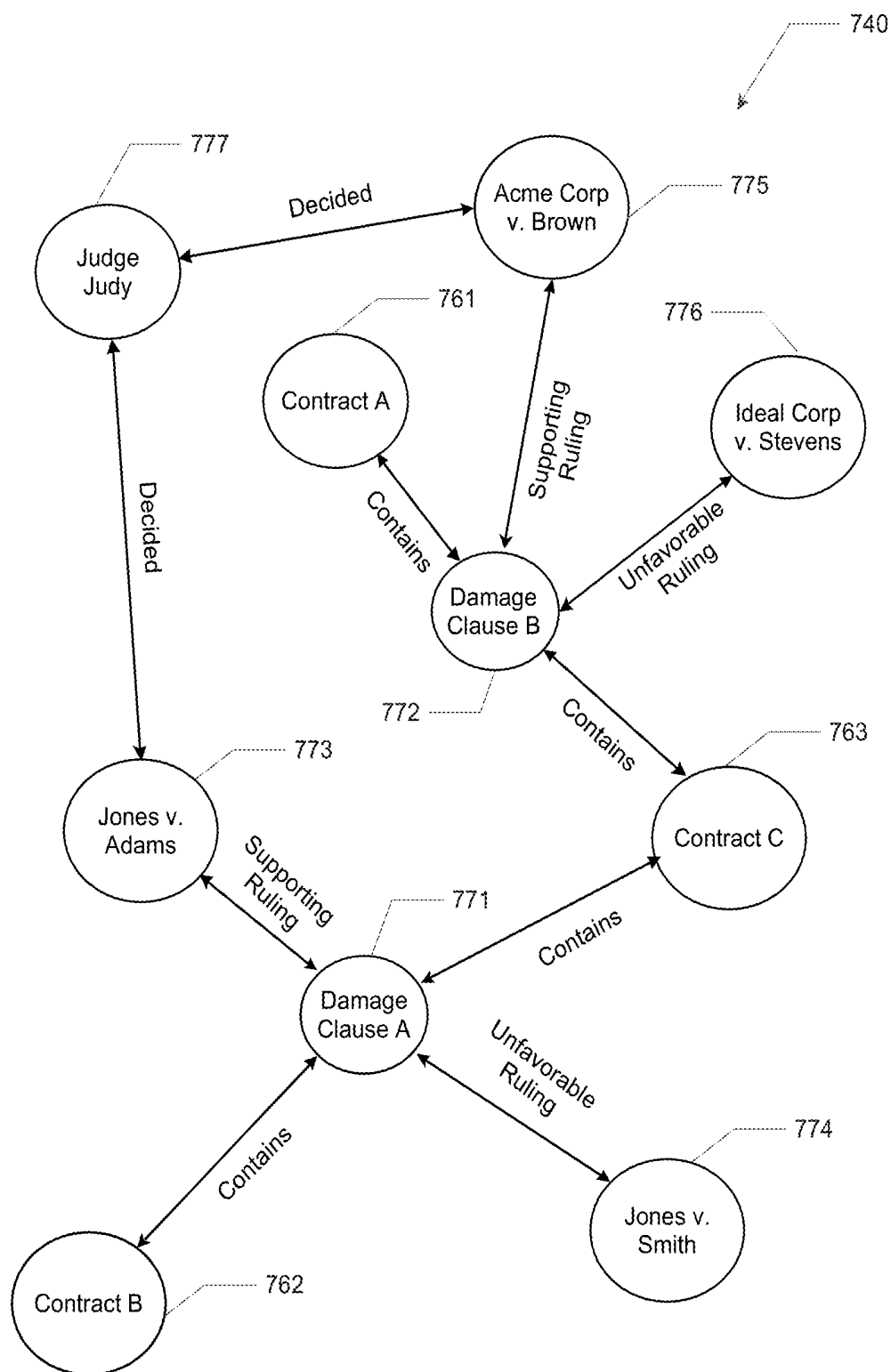


FIG. 7C

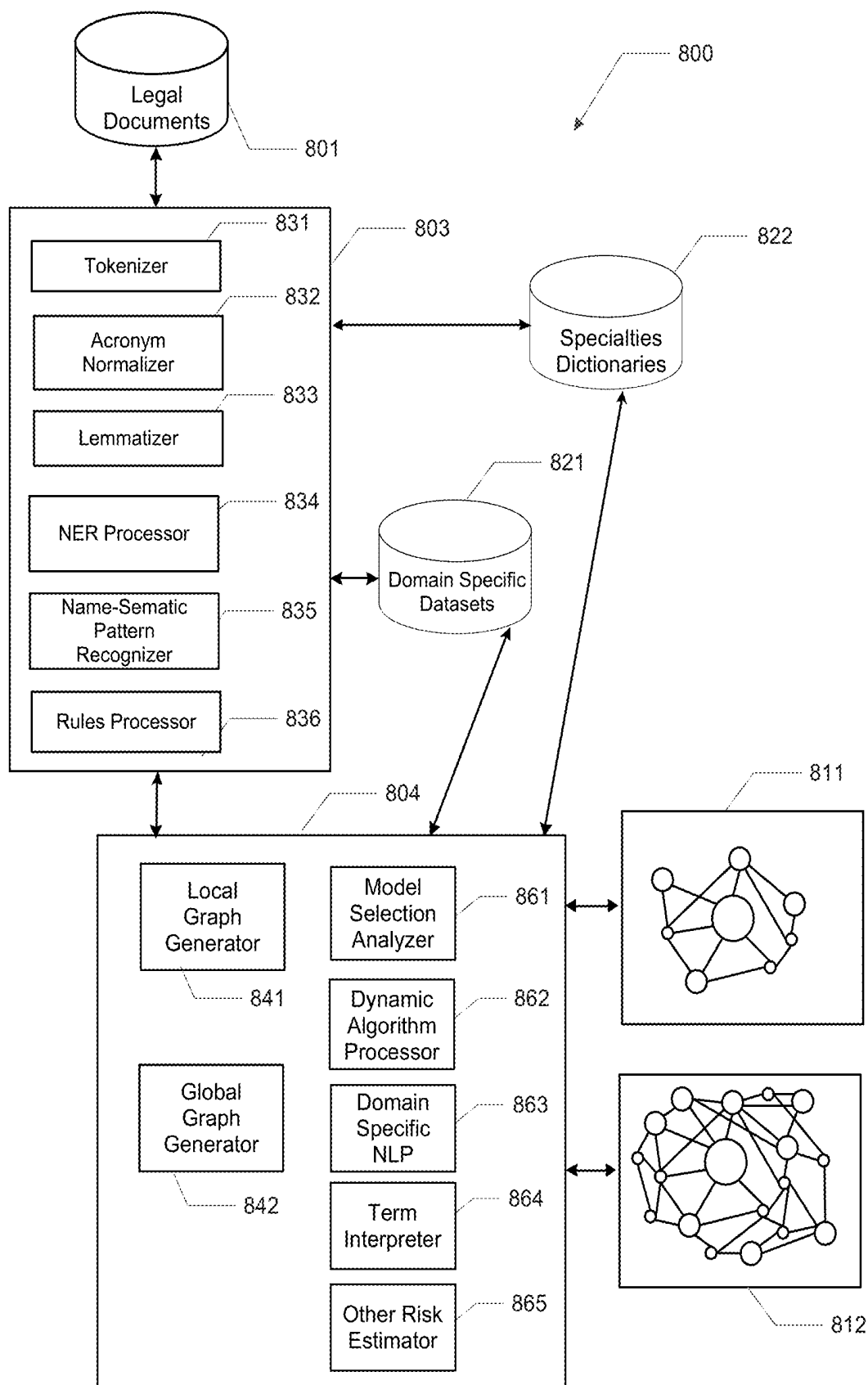


FIG. 8

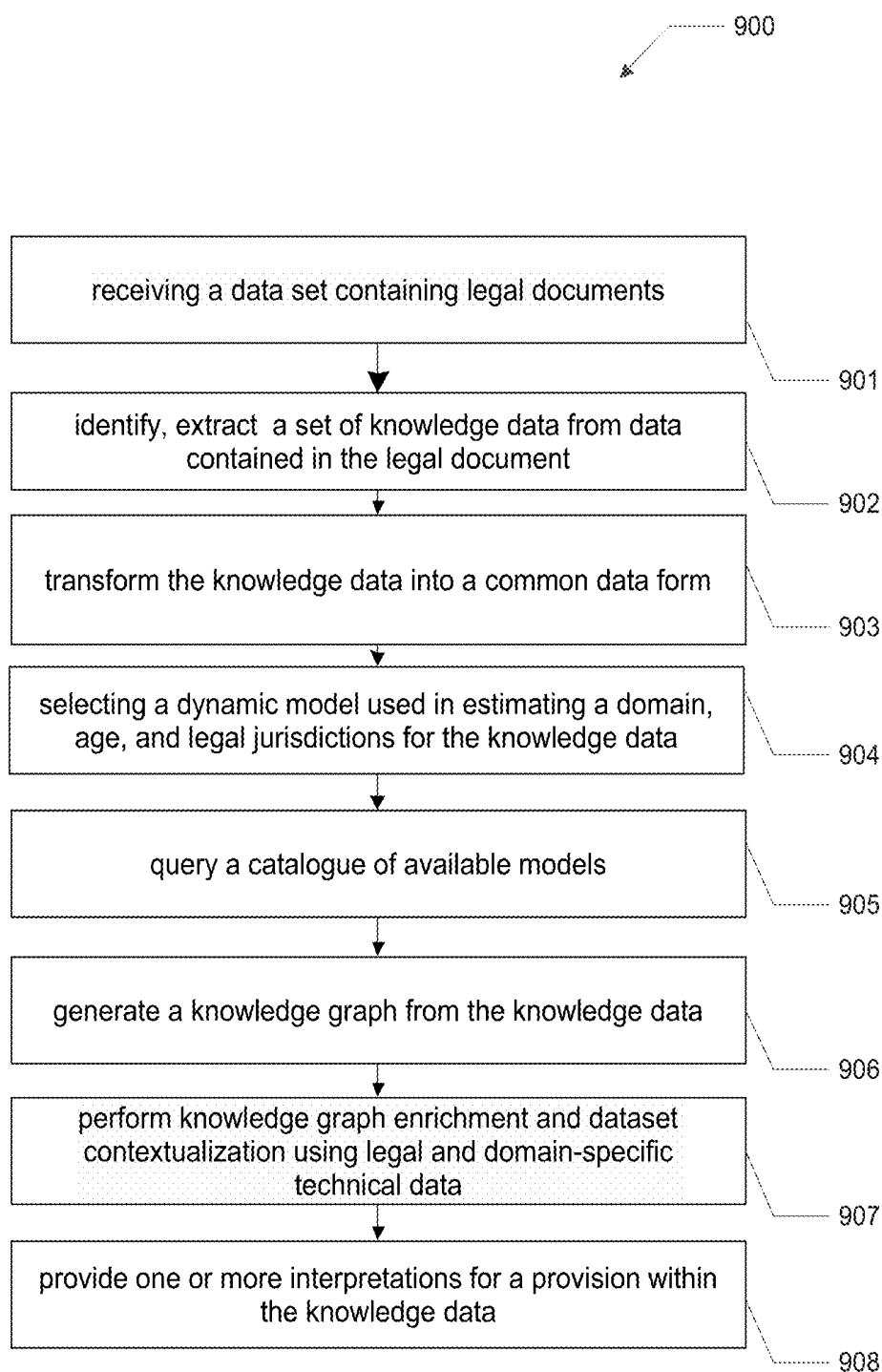


FIG. 9

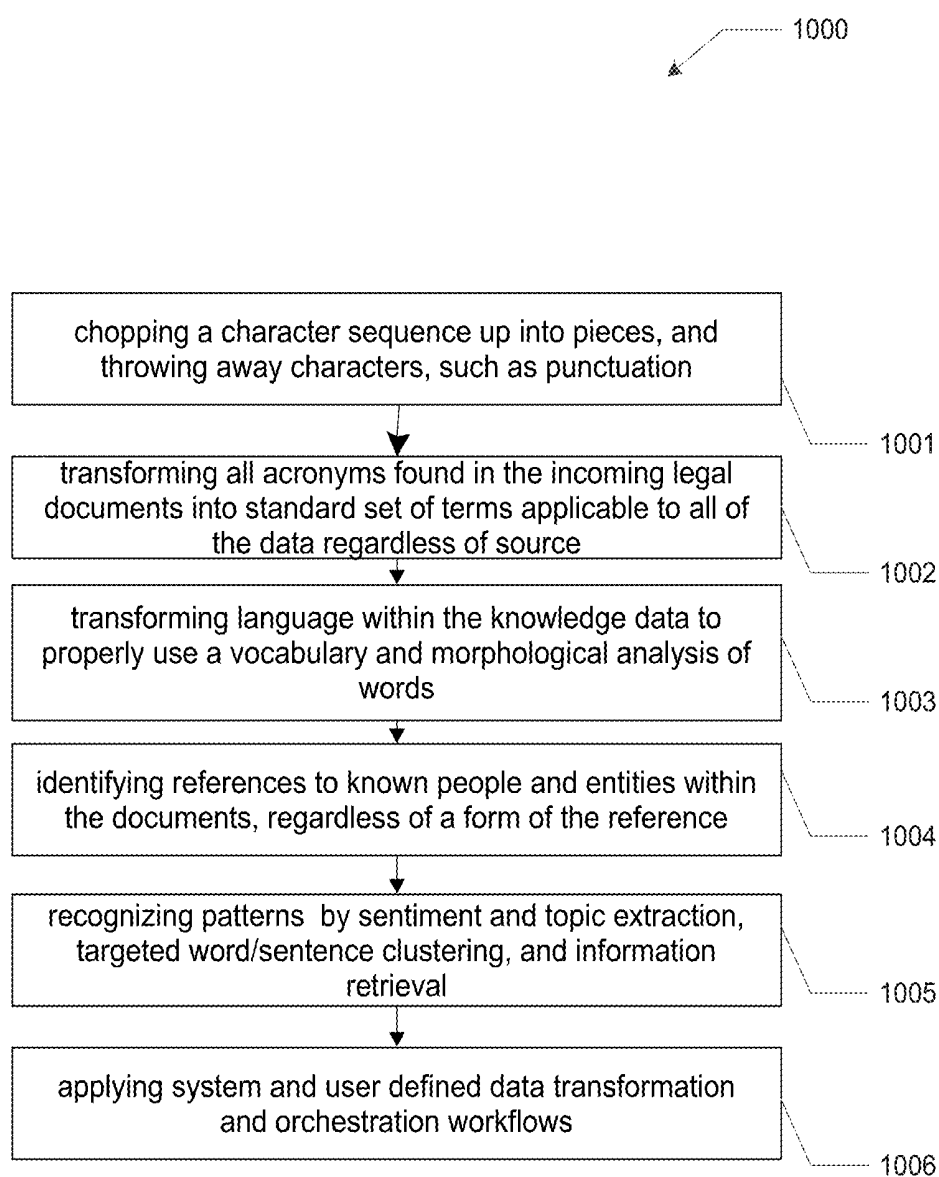


FIG. 10

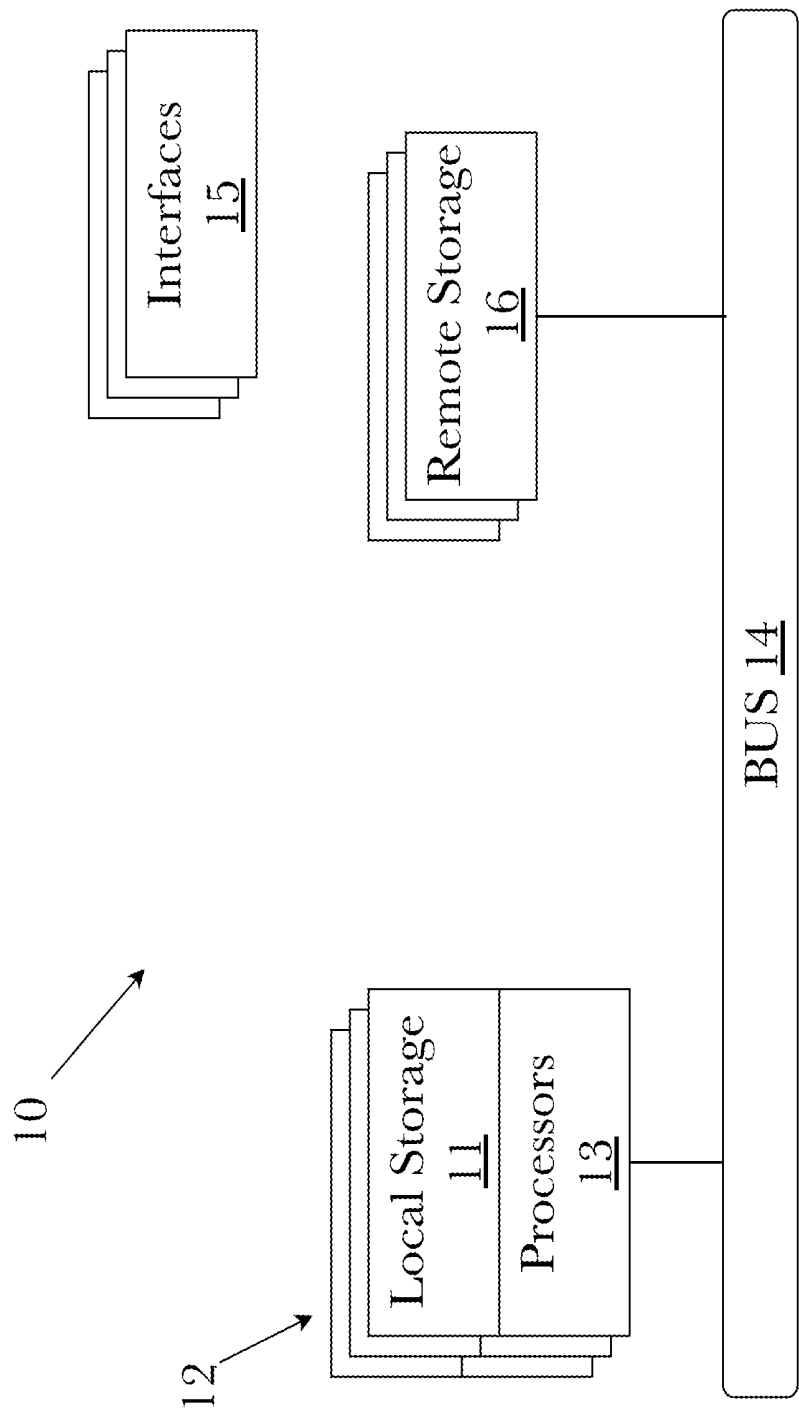


Fig. 11

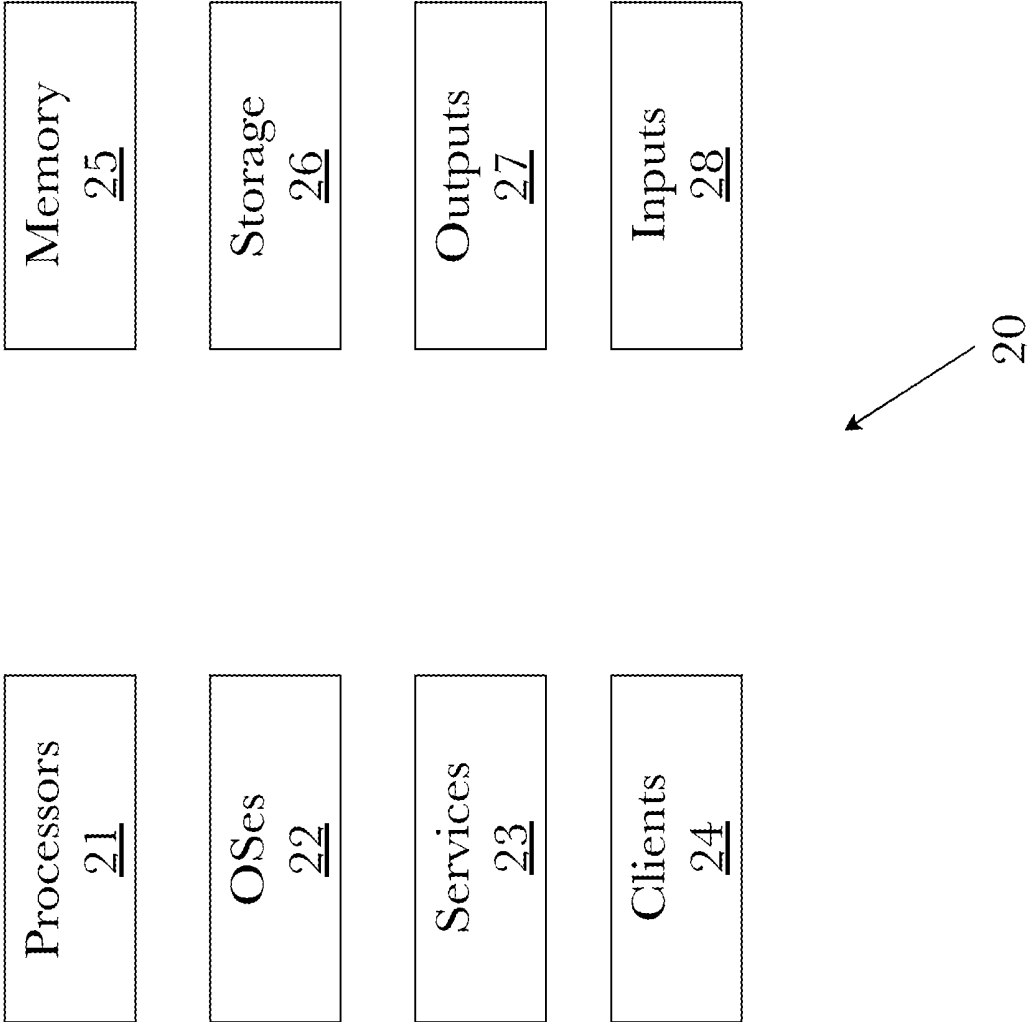


Fig. 12

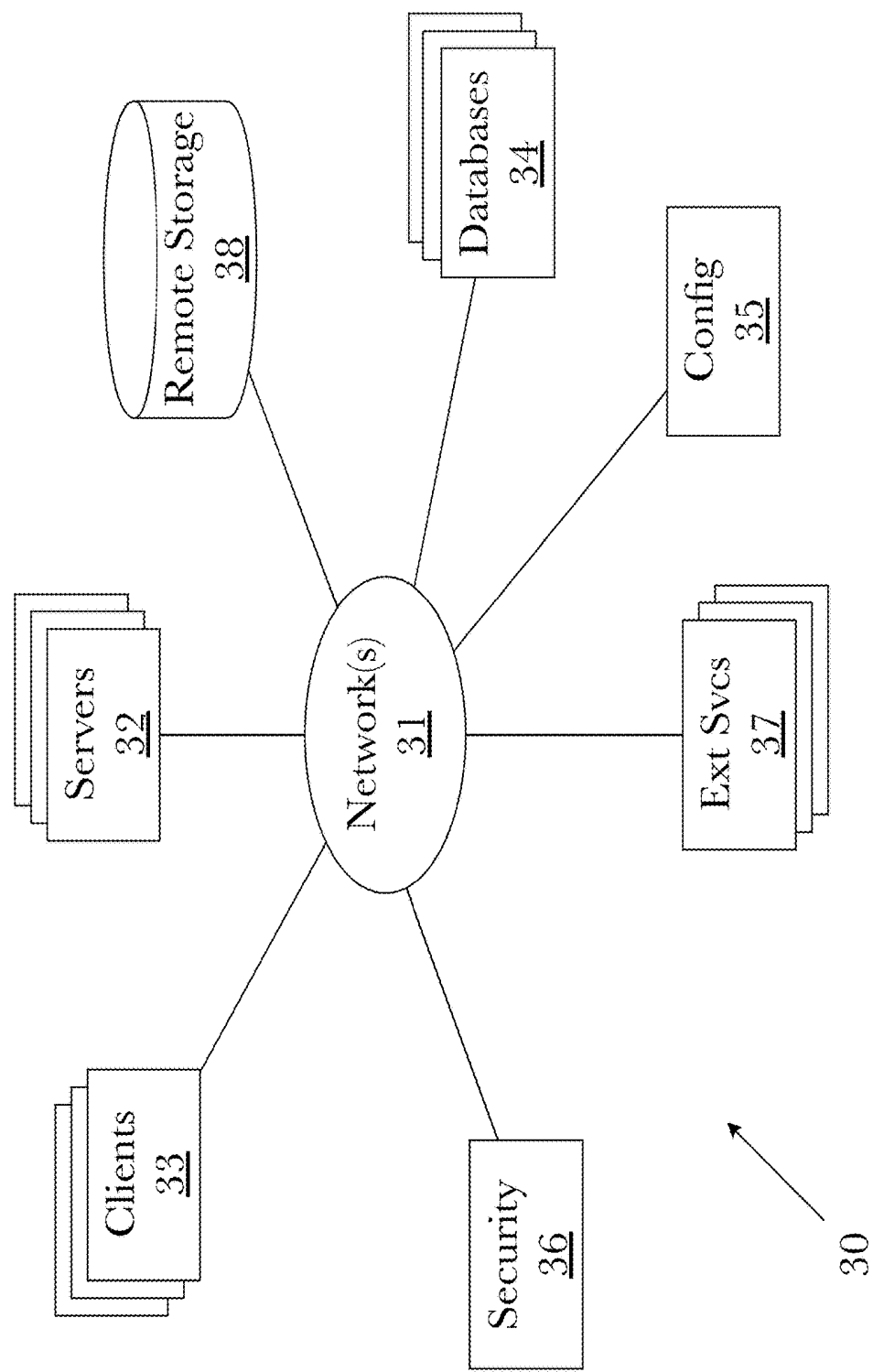


Fig. 13

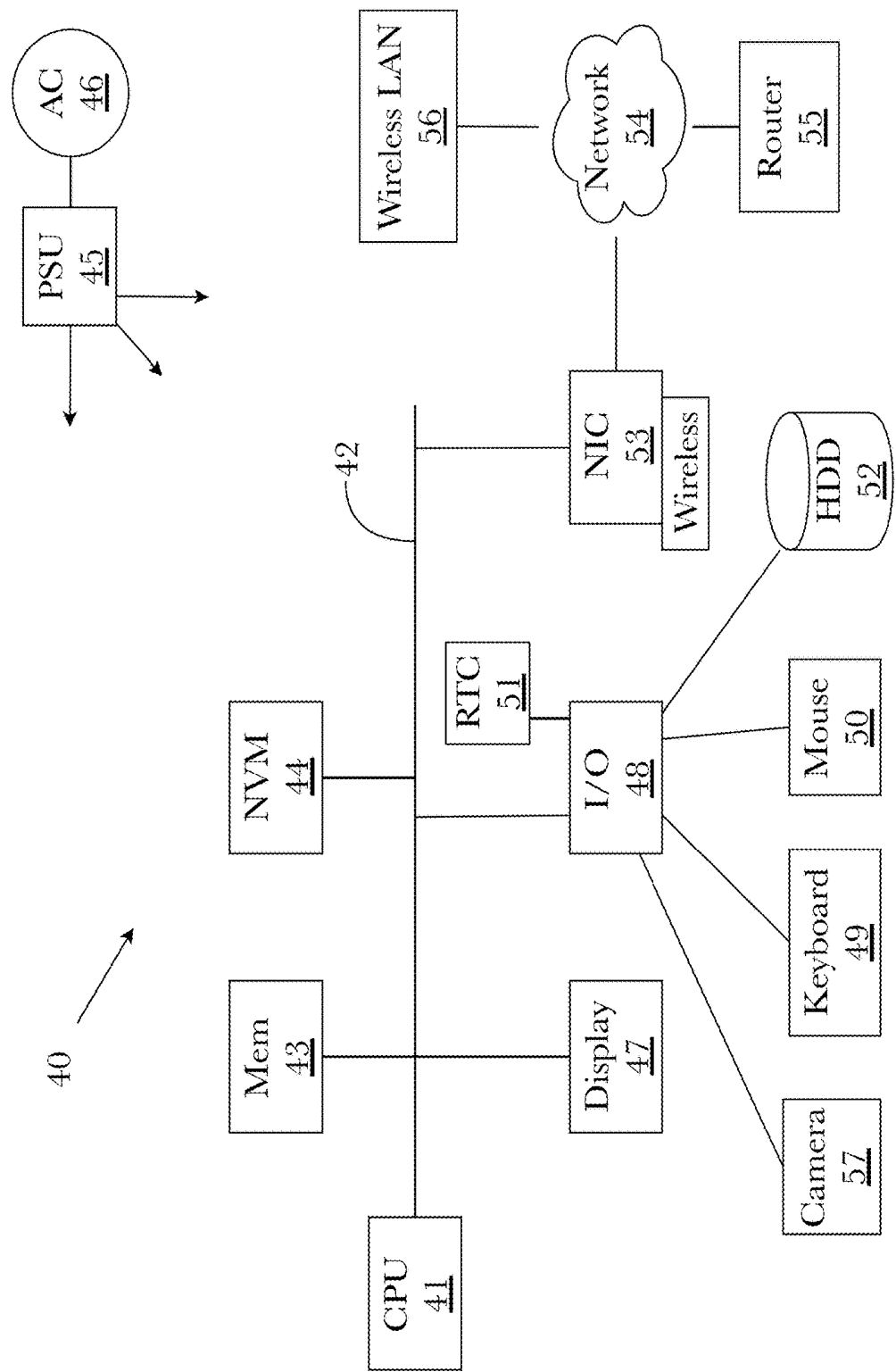


Fig. 14

**SYSTEM AND METHOD AUTOMATED
ANALYSIS OF LEGAL DOCUMENTS
WITHIN AND ACROSS SPECIFIC FIELDS**

CROSS-REFERENCE TO RELATED
APPLICATIONS

[0001]

Application No.	Date Filed	Title
Current application	Herewith	SYSTEM AND METHOD AUTOMATED ANALYSIS OF LEGAL DOCUMENTS WITHIN AND ACROSS SPECIFIC FIELDS
15/847,443	Dec. 19, 2017	Is a continuation-in-part of: SYSTEM AND METHOD FOR AUTOMATIC CREATION OF ONTOLOGICAL DATABASES AND SEMANTIC SEARCHING
15/790,457	Oct. 23, 2017	which is a continuation-in-part of: DISTRIBUTABLE MODEL WITH BIASES CONTAINED WITHIN DISTRIBUTED DATA
62/568,298	Oct. 4, 2017	which claims benefit of, and priority to: DISTRIBUTABLE MODEL WITH BIASES CONTAINED IN DISTRIBUTED DATA
15/790,327	Oct. 23, 2017	and is also a continuation-in-part of: DISTRIBUTABLE MODEL WITH DISTRIBUTED DATA
62/568,291	Oct. 4, 2017	which claims benefit of, and priority to: DISTRIBUTABLE MODEL WITH DISTRIBUTED DATA
15/141,752	Apr. 28, 2016	and is also a continuation-in-part of: SYSTEM FOR FULLY INTEGRATED CAPTURE, AND ANALYSIS OF BUSINESS INFORMATION RESULTING IN PREDICTIVE DECISION MAKING AND SIMULATION
15/091,563	Apr. 5, 2016	which is a continuation-in-part of: SYSTEM FOR CAPTURE, ANALYSIS AND STORAGE OF TIME SERIES DATA FROM SENSORS WITH HETEROGENEOUS REPORT INTERVAL PROFILES
Patent 10,204,147	Issue Date Feb. 12, 2019	and is also a continuation-in-part of: DISTRIBUTED SYSTEM FOR LARGE VOLUME DEEP WEB DATA EXTRACTION
14/986,536	Dec. 31, 2015	and is also a continuation-in-part of: RAPID PREDICTIVE ANALYSIS OF VERY LARGE DATA SETS USING THE DISTRIBUTED COMPUTATIONAL GRAPH
Patent 10,210,255	Issue Date Feb. 19, 2019	SYSTEM AND METHOD AUTOMATED ANALYSIS OF LEGAL DOCUMENTS WITHIN AND ACROSS SPECIFIC FIELDS
15/847,443	Dec. 19, 2017	Is a continuation-in-part of: SYSTEM AND METHOD FOR AUTOMATIC CREATION OF ONTOLOGICAL DATABASES AND SEMANTIC SEARCHING
15/489,716	Apr. 17, 2017	which is a continuation-in-part of: REGULATION BASED SWITCHING SYSTEM FOR ELECTRONIC MESSAGE ROUTING
15/409,510	Jan. 18, 2017	which is a continuation-in-part of: MULTI-CORPORATION VENTURE PLAN VALIDATION EMPLOYING AN ADVANCED DECISION PLATFORM
15/379,899	Dec. 15, 2016	which is a continuation-in-part of: INCLUSION OF TIME SERIES GEOSPATIAL MARKERS IN ANALYSES EMPLOYING AN ADVANCED CYBER-DECISION PLATFORM
15/376,657	Dec. 13, 2016	which is a continuation-in-part of: QUANTIFICATION FOR INVESTMENT VEHICLE MANAGEMENT EMPLOYING AN ADVANCED DECISION PLATFORM
Patent 10,402,906	Issue Date Sep. 3, 2019	

-continued

Application No.	Date Filed	Title
15/237,625 Patent	Aug. 15, 2016 Issue Date	which is a continuation-in-part of: DETECTION MITIGATION AND REMEDICATION OF CYBERATTACKS
10,248,910	Apr. 2, 2019	EMPLOYING AN ADVANCED CYBER- DECISION PLATFORM
15/206,195	Jul. 8, 2016	which is a continuation-in-part of: ACCURATE AND DETAILED MODELING OF SYSTEMS WITH LARGE COMPLEX DATASETS USING A DISTRIBUTED SIMULATION ENGINE
15/186,453	Jun. 18, 2016	which is a continuation-in-part of: SYSTEM FOR AUTOMATED CAPTURE AND ANALYSIS OF BUSINESS INFORMATION FOR RELIABLE BUSINESS VENTURE OUTCOME PREDICTION
15/166,158	May 26, 2016	which is a continuation-in-part of: SYSTEM FOR AUTOMATED CAPTURE AND ANALYSIS OF BUSINESS INFORMATION FOR SECURITY AND CLIENT-FACING INFRASTRUCTURE RELIABILITY
15/141,752	Apr. 28, 2016	which is a continuation-in-part of: SYSTEM FOR FULLY INTEGRATED CAPTURE, AND ANALYSIS OF BUSINESS INFORMATION RESULTING IN PREDICTIVE DECISION MAKING AND SIMULATION
15/091,563 Patent	Apr. 5, 2016 Issue Date	which is a continuation-in-part of: SYSTEM FOR CAPTURE, ANALYSIS AND STORAGE OF TIME SERIES DATA FROM SENSORS WITH HETEROGENEOUS REPORT INTERVAL PROFILES
10,204,147	Feb. 12, 2019	and is also a continuation-in-part of: DISTRIBUTED SYSTEM FOR LARGE VOLUME DEEP WEB DATA EXTRACTION
14/986,536 Patent	Dec. 31, 2015 Issue Date	and is also a continuation-in-part of: RAPID PREDICTIVE ANALYSIS OF VERY LARGE DATA SETS USING THE DISTRIBUTED COMPUTATIONAL GRAPH
10,210,255	Feb. 19, 2019	and is also a continuation-in-part of: RAPID PREDICTIVE ANALYSIS OF VERY LARGE DATA SETS USING THE DISTRIBUTED COMPUTATIONAL GRAPH
14/925,974	Oct. 28, 2015	and is also a continuation-in-part of: RAPID PREDICTIVE ANALYSIS OF VERY LARGE DATA SETS USING THE DISTRIBUTED COMPUTATIONAL GRAPH
Current application	Herewith	SYSTEM AND METHOD AUTOMATED ANALYSIS OF LEGAL DOCUMENTS WITHIN AND ACROSS SPECIFIC FIELDS
15/847,443	Dec. 19, 2017	Is a continuation-in-part of: SYSTEM AND METHOD FOR AUTOMATIC CREATION OF ONTOLOGICAL DATABASES AND SEMANTIC SEARCHING
15/790,457	Oct. 23, 2017	which is a continuation-in-part of: DISTRIBUTABLE MODEL WITH BIASES CONTAINED WITHIN DISTRIBUTED DATA
62/568,298	Oct. 4, 2017	which claims benefit of, and priority to: DISTRIBUTABLE MODEL WITH BIASES CONTAINED IN DISTRIBUTED DATA
15/790,327	Oct. 23, 2017	and is also a continuation-in-part of: DISTRIBUTABLE MODEL WITH DISTRIBUTED DATA
62/568,291	Oct. 4, 2017	which claims benefit of, and priority to: DISTRIBUTABLE MODEL WITH DISTRIBUTED DATA
15/616,427	Jun. 7, 2017	RAPID PREDICTIVE ANALYSIS OF VERY LARGE DATA SETS USING AN ACTOR- DRIVEN DISTRIBUTED COMPUTATIONAL GRAPH
14/925,974	Oct. 28, 2015	which is a continuation-in-part of: RAPID PREDICTIVE ANALYSIS OF VERY LARGE DATA SETS USING THE DISTRIBUTED COMPUTATIONAL GRAPH

-continued

Application No.	Date Filed	Title
Current application	Herewith	SYSTEM AND METHOD AUTOMATED ANALYSIS OF LEGAL DOCUMENTS WITHIN AND ACROSS SPECIFIC FIELDS
15/847,443	Dec. 19, 2017	Is a continuation-in-part of: SYSTEM AND METHOD FOR AUTOMATIC CREATION OF ONTOLOGICAL DATABASES AND SEMANTIC SEARCHING
15/616,427	Jun. 7, 2017	which is a continuation-in-part of: RAPID PREDICTIVE ANALYSIS OF VERY LARGE DATA SETS USING AN ACTOR-DRIVEN DISTRIBUTED COMPUTATIONAL GRAPH
14/925,974	Oct. 28, 2015	which is a continuation-in-part of: RAPID PREDICTIVE ANALYSIS OF VERY LARGE DATA SETS USING THE DISTRIBUTED COMPUTATIONAL GRAPH

the entire specification of each of which is incorporated herein by reference.

BACKGROUND

Field of the Art

[0002] The disclosure relates to the field of computer technology, more specifically to the field of computer architectures for enterprise data collection, analysis, and transmission to cloud-based services.

Discussion of the State of the Art

[0003] Legal document analysis within and across specific fields presents a tremendous challenge for applied utilization of artificial intelligence (AI) and machine learning (ML) tools based on the hyper-specialized nomenclature and concepts which may also have substantial superficial similarity but in-fact represent unique concepts across different applications of the law. Current techniques which are commonly employed to legal text analysis often over-generalize or overfit models without adequate consideration of specialized areas of the law. Distinct subspecialties employ their own parochial lexicons with highly contextualized semantics that may not be well understood even by other legal practitioners (e.g. a corpus of legal notes generated by contract lawyers may not be suitable training data for automating processing of notes generated by criminal litigators). Many of these concepts are also apposite for other specialized fields such as engineering or medicine where nomenclature and meaning can be quite distinct within subspecialties; medicine in particular is known for the specificity required in building Natural Language Processing (NLP) models with limited transferability to other (even related) specializations.

SUMMARY

[0004] Accordingly, the inventor has conceived and reduced to practice, a system and method for automated analysis of legal documents within and across different fields. A system for automated analysis of legal documents within and across different fields is disclosed. The system includes a computer system comprising at least one memory, at least one processor, and at least a first plurality of programming instructions stored in the at least one memory and operating on the at least one processor configured to

allow the operation on the computer system of additional programming instructions, an extraction processor, comprising at least a second plurality of programming instructions stored in the at least one memory of, and operating on at least one processor of, the computer system, and an analysis processor, comprising at least a second plurality of programming instructions stored in the at least one memory of, and operating on at least one processor of, the computer system. The second plurality of programming instructions, when operating on the at least one processor, cause the computer system to receive a data set containing legal documents, perform a set of systematic NLP-based data extraction generic micro-functions to identify, extract a set of knowledge data from data contained in the legal document, and transform the knowledge data into a common data form. The second plurality of programming instructions, when operating on the at least one processor, cause the computer system to perform dynamic model selection based on a series classification algorithms estimating a domain, age, and legal jurisdictions for the knowledge data, effectively query a catalogue of available models generate a knowledge graph from the knowledge data; perform knowledge graph enrichment and dataset contextualization using legal and domain-specific technical data, and using knowledge data to provide one or more interpretations for a provision within the knowledge data.

[0005] According to another aspect, a method for automated analysis of legal documents within and across different fields is disclosed. The method for automated analysis of legal documents within and across different fields receives a data set of legal documents, identifies and extracts a set of knowledge data from data contained in the legal document, transforms the knowledge data into a common data form, selects a dynamic model used in estimating a domain, age, and legal jurisdictions for the knowledge data, queries a catalogue of available models, generates a knowledge graph from the knowledge data, performs knowledge graph enrichment and dataset contextualization using legal and domain-specific technical data, and provides one or more interpretations for a provision within the knowledge data.

BRIEF DESCRIPTION OF THE DRAWING FIGURES

[0006] The accompanying drawings illustrate several aspects and, together with the description, serve to explain

the principles of the invention according to the aspects. It will be appreciated by one skilled in the art that the particular arrangements illustrated in the drawings are merely exemplary, and are not to be considered as limiting of the scope of the invention or the claims herein in any way.

[0007] FIG. 1 is a diagram of an exemplary architecture of an advanced cyber decision platform according to one aspect.

[0008] FIG. 2 is a flow diagram of an exemplary function of the business operating system in the detection and mitigation of predetermining factors leading to and steps to mitigate ongoing cyberattacks.

[0009] FIGS. 3A and 3B are a process diagram showing a general flow 300 of business operating system functions in use to mitigate cyberattacks.

[0010] FIG. 4 is an architecture diagram illustrating an exemplary system 400 for automated analysis of legal documents within and across different fields, according to aspect.

[0011] FIG. 5 is an architecture diagram illustrating an exemplary system 500 for automated analysis of legal documents within and across different fields, according to aspect.

[0012] FIG. 6 is an architecture diagram illustrating an exemplary system 600 for automated analysis of legal documents within and across different fields, according to aspect.

[0013] FIG. 7A-7C illustrate exemplary knowledge graphs for legal document analysis according to one aspect of the present invention.

[0014] FIG. 8 is an architecture diagram illustrating an exemplary system 700 for automated analysis of legal documents within and across different fields, according to aspect.

[0015] FIG. 9 is a process diagram illustrating an exemplary method 800 for automated analysis of legal documents within and across different fields, according to aspect.

[0016] FIG. 10 is a process diagram illustrating an exemplary method 900 for automated analysis of legal documents within and across different fields, according to aspect.

[0017] FIG. 11 is a block diagram illustrating an exemplary hardware architecture of a computing device.

[0018] FIG. 12 is a block diagram illustrating an exemplary logical architecture for a client device.

[0019] FIG. 13 is a block diagram showing an exemplary architectural arrangement of clients, servers, and external services.

[0020] FIG. 14 is another block diagram illustrating an exemplary hardware architecture of a computing device.

DETAILED DESCRIPTION

[0021] The inventor has conceived, and reduced to practice, automated analysis of legal documents within and across different fields.

[0022] One or more different aspects may be described in the present application. Further, for one or more of the aspects described herein, numerous alternative arrangements may be described; it should be appreciated that these are presented for illustrative purposes only and are not limiting of the aspects contained herein or the claims presented herein in any way. One or more of the arrangements may be widely applicable to numerous aspects, as may be readily apparent from the disclosure. In general, arrangements are described in sufficient detail to enable those

skilled in the art to practice one or more of the aspects, and it should be appreciated that other arrangements may be utilized and that structural, logical, software, electrical and other changes may be made without departing from the scope of the particular aspects. Particular features of one or more of the aspects described herein may be described with reference to one or more particular aspects or figures that form a part of the present disclosure, and in which are shown, by way of illustration, specific arrangements of one or more of the aspects. It should be appreciated, however, that such features are not limited to usage in the one or more particular aspects or figures with reference to which they are described. The present disclosure is neither a literal description of all arrangements of one or more of the aspects nor a listing of features of one or more of the aspects that must be present in all arrangements.

[0023] Headings of sections provided in this patent application and the title of this patent application are for convenience only, and are not to be taken as limiting the disclosure in any way.

[0024] Devices that are in communication with each other need not be in continuous communication with each other, unless expressly specified otherwise. In addition, devices that are in communication with each other may communicate directly or indirectly through one or more communication means or intermediaries, logical or physical.

[0025] A description of an aspect with several components in communication with each other does not imply that all such components are required. To the contrary, a variety of optional components may be described to illustrate a wide variety of possible aspects and in order to more fully illustrate one or more aspects. Similarly, although process steps, method steps, algorithms or the like may be described in a sequential order, such processes, methods and algorithms may generally be configured to work in alternate orders, unless specifically stated to the contrary. In other words, any sequence or order of steps that may be described in this patent application does not, in and of itself, indicate a requirement that the steps be performed in that order. The steps of described processes may be performed in any order practical. Further, some steps may be performed simultaneously despite being described or implied as occurring non-simultaneously (e.g., because one step is described after the other step). Moreover, the illustration of a process by its depiction in a drawing does not imply that the illustrated process is exclusive of other variations and modifications thereto, does not imply that the illustrated process or any of its steps are necessary to one or more of the aspects, and does not imply that the illustrated process is preferred. Also, steps are generally described once per aspect, but this does not mean they must occur once, or that they may only occur once each time a process, method, or algorithm is carried out or executed. Some steps may be omitted in some aspects or some occurrences, or some steps may be executed more than once in a given aspect or occurrence.

[0026] When a single device or article is described herein, it will be readily apparent that more than one device or article may be used in place of a single device or article. Similarly, where more than one device or article is described herein, it will be readily apparent that a single device or article may be used in place of the more than one device or article.

[0027] The functionality or the features of a device may be alternatively embodied by one or more other devices that are

not explicitly described as having such functionality or features. Thus, other aspects need not include the device itself.

[0028] Techniques and mechanisms described or referenced herein will sometimes be described in singular form for clarity. However, it should be appreciated that particular aspects may include multiple iterations of a technique or multiple instantiations of a mechanism unless noted otherwise. Process descriptions or blocks in figures should be understood as representing modules, segments, or portions of code which include one or more executable instructions for implementing specific logical functions or steps in the process. Alternate implementations are included within the scope of various aspects in which, for example, functions may be executed out of order from that shown or discussed, including substantially concurrently or in reverse order, depending on the functionality involved, as would be understood by those having ordinary skill in the art.

Definitions

[0029] “Artificial intelligence” or “AI” as used herein means a computer system or component that has been programmed in such a way that it mimics some aspect or aspects of cognitive functions that humans associate with human intelligence, such as learning, problem solving, and decision-making. Examples of current AI technologies include understanding human speech, competing successfully in strategic games such as chess and Go, autonomous operation of vehicles, complex simulations, and interpretation of complex data such as images and video.

[0030] “Machine learning” as used herein is an aspect of artificial intelligence in which the computer system or component can modify its behavior or understanding without being explicitly programmed to do so. Machine learning algorithms develop models of behavior or understanding based on information fed to them as training sets, and can modify those models based on new incoming information. An example of a machine learning algorithm is AlphaGo, the first computer program to defeat a human world champion in the game of Go. AlphaGo was not explicitly programmed to play Go. It was fed millions of games of Go, and developed its own model of the game and strategies of play.

[0031] “Neural network” as used herein means a computational model, architecture, or system made up of a number of simple, highly interconnected processing elements which process information by their dynamic state response to external inputs, and is thus able to “learn” information by recognizing patterns or trends. Neural networks, also sometimes known as “artificial neural networks” are based on our understanding of the structure and functions of biological neural networks, such as the brains of mammals. A neural network is a framework for application of machine learning algorithms.

Conceptual Architecture

[0032] FIG. 1 is a diagram of an exemplary architecture of an advanced cyber decision platform (ACDP) 100 according to one aspect. Client access to the system 105 for specific data entry, system control and for interaction with system output such as automated predictive decision making and planning and alternate pathway simulations, occurs through the system’s distributed, extensible high bandwidth cloud interface 110 which uses a versatile, robust web application

driven interface for both input and display of client-facing information via network 107 and operates a data store 112 such as, but not limited to MONGODB™, COUCHDB™, CASSANDRA™ or REDIS™ according to various arrangements. Much of the business data analyzed by the system both from sources within the confines of the client business, and from cloud based sources, also enter the system through the cloud interface 110, data being passed to the connector module 135 which may possess the API routines 135a needed to accept and convert the external data and then pass the normalized information to other analysis and transformation components of the system, the directed computational graph module 155, high volume web crawler module 115, multidimensional time series database 120 and the graph stack service 145. The directed computational graph module 155 retrieves one or more streams of data from a plurality of sources, which includes, but is in no way not limited to, a plurality of physical sensors, network service providers, web based questionnaires and surveys, monitoring of electronic infrastructure, crowd sourcing campaigns, and human input device information. Within the directed computational graph module 155, data may be split into two identical streams in a specialized pre-programmed data pipeline 155a, wherein one sub-stream may be sent for batch processing and storage while the other sub-stream may be reformatted for transformation pipeline analysis. The data is then transferred to the general transformer service module 160 for linear data transformation as part of analysis or the decomposable transformer service module 150 for branching or iterative transformations that are part of analysis. The directed computational graph module 155 represents all data as directed graphs where the transformations are nodes and the result messages between transformations edges of the graph. The high volume web crawling module 115 uses multiple server hosted preprogrammed web spiders, which while autonomously configured are deployed within a web scraping framework 115a of which SCRAPY™ is an example, to identify and retrieve data of interest from web based sources that are not well tagged by conventional web crawling technology. The multiple dimension time series data store module 120 may receive streaming data from a large plurality of sensors that may be of several different types. The multiple dimension time series data store module may also store any time series data encountered by the system such as but not limited to enterprise network usage data, component and system logs, performance data, network service information captures such as, but not limited to news and financial feeds, and sales and service related customer data. The module is designed to accommodate irregular and high volume surges by dynamically allotting network bandwidth and server processing channels to process the incoming data. Inclusion of programming wrappers for languages examples of which are, but not limited to C++, PERL, PYTHON, and ERLANG™ allows sophisticated programming logic to be added to the default function of the multidimensional time series database 120 without intimate knowledge of the core programming, greatly extending breadth of function. Data retrieved by the multidimensional time series database 120 and the high volume web crawling module 115 may be further analyzed and transformed into task optimized results by the directed computational graph 155 and associated general transformer service 150 and decomposable transformer service 160 modules. Alternately, data from the multidimensional time series database and

high volume web crawling modules may be sent, often with scripted cuing information determining important vertexes **145a**, to the graph stack service module **145** which, employing standardized protocols for converting streams of information into graph representations of that data, for example, open graph internet technology although the invention is not reliant on any one standard. Through the steps, the graph stack service module **145** represents data in graphical form influenced by any pre-determined scripted modifications **145a** and stores it in a graph-based data store **145b** such as GIRAPH™ or a key value pair type data store REDIST™, or RIAK™, among others, all of which are suitable for storing graph-based information.

[0033] Results of the transformative analysis process may then be combined with further client directives, additional business rules and practices relevant to the analysis and situational information external to the already available data in the automated planning service module **130** which also runs powerful information theory **130a** based predictive statistics functions and machine learning algorithms to allow future trends and outcomes to be rapidly forecast based upon the current system derived results and choosing each a plurality of possible business decisions. The using all available data, the automated planning service module **130** may propose business decisions most likely to result is the most favorable business outcome with a usably high level of certainty. Closely related to the automated planning service module in the use of system derived results in conjunction with possible externally supplied additional information in the assistance of end user business decision making, the action outcome simulation module **125** with its discrete event simulator programming module **125a** coupled with the end user facing observation and state estimation service **140** which is highly scriptable **140b** as circumstances require and has a game engine **140a** to more realistically stage possible outcomes of business decisions under consideration, allows business decision makers to investigate the probable outcomes of choosing one pending course of action over another based upon analysis of the current available data.

[0034] For example, the Information Assurance department is notified by the system **100** that principal X is using credentials K (Kerberos Principal Key) never used by it before to access service Y. Service Y utilizes these same credentials to access secure data on data store Z. This correctly generates an alert as suspicious lateral movement through the network and will recommend isolation of X and Y and suspension of K based on continuous baseline network traffic monitoring by the multidimensional time series data store **120** programmed to process such data **120a**, rigorous analysis of the network baseline by the directed computational graph **155** with its underlying general transformer service module **160** and decomposable transformer service module **150** in conjunction with the AI and primed machine learning capabilities **130a** of the automated planning service module **130** which had also received and assimilated publicly available from a plurality of sources through the multi-source connection APIs of the connector module **135**. Ad hoc simulations of these traffic patterns are run against the baseline by the action outcome simulation module **125** and its discrete event simulator **125a**, which is used here to determine probability space for likelihood of legitimacy. The system **100**, based on this data and analysis, was able to detect and recommend mitigation of a cyberattack that represented an existential threat to all business

operations, presenting, at the time of the attack, information most needed for an actionable plan to human analysts at multiple levels in the mitigation and remediation effort through use of the observation and state estimation service **140** which had also been specifically preprogrammed to handle cybersecurity events **140b**.

[0035] According to one aspect, the advanced cyber decision platform, a specifically programmed usage of the business operating system, continuously monitors a client enterprise's normal network activity for behaviors such as but not limited to normal users on the network, resources accessed by each user, access permissions of each user, machine to machine traffic on the network, sanctioned external access to the core network and administrative access to the network's identity and access management servers in conjunction with real-time analytics informing knowledge of cyberattack methodology. The system then uses this information for two purposes: First, the advanced computational analytics and simulation capabilities of the system are used to provide immediate disclosure of probable digital access points both at the network periphery and within the enterprise's information transfer and trust structure and recommendations are given on network changes that should be made to harden it prior to or during an attack. Second, the advanced cyber decision platform continuously monitors the network in real-time both for types of traffic and through techniques such as deep packet inspection for pre-decided analytically significant deviation in user traffic for indications of known cyberattack vectors such as, but not limited to, ACTIVE DIRECTORY™/Kerberos pass-the-ticket attack, ACTIVE DIRECTORY™/Kerberos pass-the-hash attack and the related ACTIVE DIRECTORY™/Kerberos overpass-the-hash attack, ACTIVE DIRECTORY™/Kerberos Skeleton Key, ACTIVE DIRECTORY™/Kerberos golden and silver ticket attack, privilege escalation attack, compromised user credentials, and ransomware disk attacks. When suspicious activity at a level signifying an attack (for example, including but not limited to skeleton key attacks, pass-the-hash attacks, or attacks via compromised user credentials) is determined, the system issues action-focused alert information to all predesignated parties specifically tailored to their roles in attack mitigation or remediation and formatted to provide predictive attack modeling based upon historic, current, and contextual attack progression analysis such that human decision makers can rapidly formulate the most effective courses of action at their levels of responsibility in command of the most actionable information with as little distractive data as possible. The system then issues defensive measures in the most actionable form to end the attack with the least possible damage and exposure. All attack data are persistently stored for later forensic analysis.

[0036] FIG. 2 is a flow diagram of an exemplary function of the business operating system in the detection and mitigation of predetermining factors leading to and steps to mitigate ongoing cyberattacks **200**. The system continuously retrieves network traffic data **201** which may be stored and preprocessed by the multidimensional time series data store **120** and its programming wrappers **120a**. All captured data are then analyzed to predict the normal usage patterns of network nodes such as internal users, network connected systems and equipment and sanctioned users external to the enterprise boundaries for example off-site employees, contractors and vendors, just to name a few likely participants. Of course, normal other network traffic may also be known

to those skilled in the field, the list given is not meant to be exclusive and other possibilities would not fall outside the design of the invention. Analysis of network traffic may include graphical analysis of parameters such as network item to network usage using specifically developed programming in the graph stack service **145**, **145a**, analysis of usage by each network item may be accomplished by specifically pre-developed algorithms associated with the directed computational graph module **155**, general transformer service module **160** and decomposable service module **150**, depending on the complexity of the individual usage profile **201**. These usage pattern analyses, in conjunction with additional data concerning an enterprise's network topology; gateway firewall programming; internal firewall configuration; directory services protocols and configuration; and permissions profiles for both users and for access to sensitive information, just to list a few non-exclusive examples may then be analyzed further within the automated planning service module **130**, where machine learning techniques which include but are not limited to information theory statistics **130a** may be employed and the action outcome simulation module **125**, specialized for predictive simulation of outcome based on current data **125a** may be applied to formulate a current, up-to-date and continuously evolving baseline network usage profile **202**. This same data would be combined with up-to-date known cyberattack methodology reports, possibly retrieved from several divergent and exogenous sources through the use of the multi-application programming interface aware connector module **135** to present preventative recommendations to the enterprise decision makers for network infrastructure changes, physical and configuration-based to cost effectively reduce the probability of a cyberattack and to significantly and most cost effectively mitigate data exposure and loss in the event of attack **203**, **204**.

[0037] While some of these options may have been partially available as piecemeal solutions in the past, we believe the ability to intelligently integrate the large volume of data from a plurality of sources on an ongoing basis followed by predictive simulation and analysis of outcome based upon that current data such that actionable, business practice efficient recommendations can be presented is both novel and necessary in this field.

[0038] Once a comprehensive baseline profile of network usage using all available network traffic data has been formulated, the specifically tasked business operating system continuously polls the incoming traffic data for activities anomalous to that baseline as determined by pre-designated boundaries **205**. Examples of anomalous activities may include a user attempting to gain access several workstations or servers in rapid succession, or a user attempting to gain access to a domain server of server with sensitive information using random userIDs or another user's userID and password, or attempts by any user to brute force crack a privileged user's password, or replay of recently issued ACTIVE DIRECTORY™/Kerberos ticket granting tickets, or the presence on any known, ongoing exploit on the network or the introduction of known malware to the network, just to name a very small sample of the cyberattack profiles known to those skilled in the field. The invention, being predictive as well as aware of known exploits is designed to analyze any anomalous network behavior, formulate probable outcomes of the behavior, and to then issue any needed alerts regardless of whether the attack follows a

published exploit specification or exhibits novel characteristics deviant to normal network practice. Once a probable cyberattack is detected, the system then is designed to get needed information to responding parties **206** tailored, where possible, to each role in mitigating the attack and damage arising from it **207**. This may include the exact subset of information included in alerts and updates and the format in which the information is presented which may be through the enterprise's existing security information and event management system. Network administrators, then, might receive information such as but not limited to where on the network the attack is believed to have originated, what systems are believed currently affected, predictive information on where the attack may progress, what enterprise information is at risk and actionable recommendations on repelling the intrusion and mitigating the damage, whereas a chief information security officer may receive alert including but not limited to a timeline of the cyberattack, the services and information believed compromised, what action, if any has been taken to mitigate the attack, a prediction of how the attack may unfold and the recommendations given to control and repel the attack **207**, although all parties may access any network and cyberattack information for which they have granted access at any time, unless compromise is suspected. Other specifically tailored updates may be issued by the system **206**, **207**.

[0039] FIGS. 3A and 3B are a process diagram showing a general flow **300** of business operating system functions in use to mitigate cyberattacks. Input network data which may include network flow patterns **321**, the origin and destination of each piece of measurable network traffic **322**, system logs from servers and workstations on the network **323**, endpoint data **329**, any security event log data from servers or available security information and event (SIEM) systems **324**, external threat intelligence feeds **324a**, identity or assessment context **325**, external network health or cybersecurity feeds **326**, Kerberos domain controller or ACTIVE DIRECTORY™ server logs or instrumentation **327** and business unit performance related data **328**, among many other possible data types for which the invention was designed to analyze and integrate, may pass into **315** the business operating system **310** for analysis as part of its cyber security function. These multiple types of data from a plurality of sources may be transformed for analysis **311**, **312** using at least one of the specialized cybersecurity, risk assessment or common functions of the business operating system in the role of cybersecurity system, such as, but not limited to network and system user privilege oversight **331**, network and system user behavior analytics **332**, attacker and defender action timeline **333**, SIEM integration and analysis **334**, dynamic benchmarking **335**, and incident identification and resolution performance analytics **336** among other possible cybersecurity functions; value at risk (VAR) modeling and simulation **341**, anticipatory vs. reactive cost estimations of different types of data breaches to establish priorities **342**, work factor analysis **343** and cyber event discovery rate **344** as part of the system's risk analytics capabilities; and the ability to format and deliver customized reports and dashboards **351**, perform generalized, ad hoc data analytics on demand **352**, continuously monitor, process and explore incoming data for subtle changes or diffuse informational threads **353** and generate cyber-physical systems graphing **354** as part of the business operating system's common capabilities. Output **317** can be

used to configure network gateway security appliances **361**, to assist in preventing network intrusion through predictive change to infrastructure recommendations **362**, to alert an enterprise of ongoing cyberattack early in the attack cycle, possibly thwarting it but at least mitigating the damage **362**, to record compliance to standardized guidelines or SLA requirements **363**, to continuously probe existing network infrastructure and issue alerts to any changes which may make a breach more likely **364**, suggest solutions to any domain controller ticketing weaknesses detected **365**, detect presence of malware **366**, perform one time or continuous vulnerability scanning depending on client directives **367**, thwart cyber attacks and mitigate damage **368**. These examples are, of course, only a subset of the possible uses of the system, they are exemplary in nature and do not reflect any boundaries in the capabilities of the invention.

Detailed Description of Exemplary Aspects

[0040] FIG. 4 is an architecture diagram illustrating an exemplary system **400** for automated analysis of legal documents within and across different fields, according to aspect. System **400** provides a scalable and extensible framework for addressing the fragmented, but typically hierarchical, nature of these increasingly specialized uses of legal and technical language intermixed inside contract documents, patents, employment agreements and other common documents (e.g. transactional documents associated with M&A or investment activity). Legal documents are received and processed by input module **401**. This module **401** ingest all documents to be processed into a format that is used throughout the system **400**. Data may be received from a variety of different data sources containing data useful in interpreting the legal documents. The input module **401** also extracts all relevant data using natural language processing (NLP) processing pipelines. To properly prepare the data for analysis, input module **401** deduplicated the data to reduce dataset size and streamline it for future processing. This deduplication involves leveraging various mathematical formalisms and NLP algorithms for unstructured data to recognize and remove redundant information. The processed data is stored in a model **405** optimized for querying and analysis, such as a graph, time series, index, relation, or key-value store. For example, a block of text will be broken up and placed in rows in a table for faster processing but a highly relational data set may need to be in a graph to support efficient traversals.

[0041] Analysis engine **402** receives the data generated by the input module **401** to perform analysis processing on the data. Analysis engine **402** normalizes the data by converting various data types into a consistent unified data model. The data is associated with appropriate entities and that units or values are uniform, perhaps converting all temperatures to either Celsius or Fahrenheit or all time to the GMT time zone. This process further prepares the data for analysis with various predictive models, including machine learning algorithms. Depending on the use case, it can be performed either on streaming data as it is ingested or on batches of stored data.

[0042] The analysis engine **402** uses a standardized set of naming conventions or ontologies are applied to the data to ensure there is no confusion about the definitions of objects or values within the dataset. The process of indexing data points with their logical or linguistic meaning in the context of surrounding data points is referred to as semantifying

data. Use of semantified data facilitates deeper reasoning and automated analysis of entities within the data and their relationships to one another.

[0043] Sometimes simple expressions and logic are enough to glean sufficient insight from this semantified data, but other times more advanced techniques involving various machine learning or deep learning approaches may be needed. For example, model-driven analysis extracts intended meaning and sentiment from widely varying sources of structured or unstructured data, allowing infusion or enrichment of the semantified data with even clearer meaning and greater context for increasingly improved automated processing and more insightful interpretation.

[0044] Analysis engine **402** uses the semantified data to build dynamic models **405** that enable better understanding, prediction, or simulation of the world. Additional self-improvement for models may leverage orchestration tools or reinforcement learning techniques to recognize and correct for model bias to continually maximize the accuracy and effectiveness of those models over time. Contextualized information can now be presented in a way that enables Human+AI collaboration to optimize decision-making and performance, at scale, to any domain.

[0045] This analysis processing is domain-agnostic, meaning the processing can be applied to virtually any structured or unstructured data set using Domain-Specific Languages (DSLs) and common data formalisms. FIG. 6 illustrates a process for performing this data input. The data is ingested **601** into the input module **401**. The data may be de-duplicated **602** to identify and correct duplicate and redundant information that exists in the raw input data. The input data may be schematized **603** to organize the data into a common framework. This data may be further processed **604** to normalize it into a common frame of reference for use when comparing data from differing sources. The initial processing is completed by semantifying the data into a standardized domain-specific ontology **605** enabling both machines and humans to understand, share, and reason with all of the data as a single entity.

[0046] By organizing data into a unified model **405** upon ingestion, the model **405** enables databases to evolve into knowledge bases where querying and analysis by monitoring module **403** is intuitive and efficient, even when reasoning about data from heterogeneous and disparate sources. Actual insights are delivered across data sources and not just a storage bill associated with a large but often unusable data lake. This module **403** is able to process logs and instrumented interactions at massive scale to provide the immediate context needed to understand what is present within the legal. As a result, the monitoring module **403**'s process of decomposing both data and analytic work enables the delivery of visibility and detection capabilities that continually improve as more data is gradually collected and integrated.

[0047] This visibility supports understanding and predictive awareness useful when coupled with mathematical models and statistical analytics. This module **403** includes simulation modeling to explore what if scenarios and the use of other model management tools to improve predictive model performance and management. User's exploration of hypothetical strategies and outcomes with continuously enriched allows for better understand the intricacies of their entire legal position from the adoption and/or used of the legal documents. This analysis may be useful in preparing and negotiating legal agreements given complex interdepen-

dencies that would otherwise be very difficult to recognize or comprehend. This analysis may also be useful in determining whether a legal claim is either strong or weak when the legal documents are litigated. This processing may continue as new data from court opinions, litigation results, additional negotiations and similar events provides data relevant to the analysis.

[0048] FIG. 5 is an architecture diagram illustrating an exemplary process 500 for automated analysis of legal documents within and across different fields, according to aspect. An entire feedback loop of analysis steps is supported by an analytics processing system 400. Initial data is received and ingested 501 in which it is processed and formatted into a common framework for consideration within ongoing analysis. All of the data is stored 502 into any number of different datastores for use by the analysis processing.

[0049] Stored data may be orchestrated 503 into different automated workflows and varying transformations as users explore the data. This user exploration uses analysis processing 504 that generates various analytics using the common set of data from multiple sources. Once some analysis has been done, the data may be used to predict 505 outcomes from various simulations using historical, counter-factual, and synthetic data sources.

[0050] Users may use all the above results to explore 506 additional hypothetical scenarios to validate assumption about dynamic systems that would otherwise be difficult to measure. The user may visualize 507 the results from all of the analysis, predictions, and explorations into a common set of results as users identify complex interdependencies in the data from the multitude of sources.

[0051] FIG. 7A-7C illustrate exemplary knowledge graphs for legal document analysis according to one aspect of the present invention. This analysis processing 400 may be used in several example analysis modeling. FIG. 7A illustrates one example knowledge graph. Legal briefs 701, 702, 703 filed in a case by all parties, or legal briefs from all cases on the same claim 711, 712, can be analyzed against each other, and the relative strength and weaknesses of the cited caselaw can be determined. The briefs themselves 701, 702, 703 provide one set of relevant data, caselaw opinions 713-716 provide standards to be applied and competing cases, can be compared to the facts at issue to estimate any strength or weakness in the arguments presented. For example, a cited case 713 can have its opinion compared against the decisions of all subsequent cases 719, 714, 715 that cite the asserted opinion 713. The facts described in these cases can be compared against the facts outlined in the briefs. The analysis may be performed with finds of fact for issues in dispute resolved in all possible ways to provide predictive outcomes in all of these possible scenarios. Such analysis by each party permits each to develop responses to the strongest points of opposing parties, and when performed before a responsive brief is filed permit drafting of arguments that best overcomes the opponents' case. The identity of a judge 717 that rules in the current case or in prior decisions may also be tracked to determine whether a favorable decision may be expected.

[0052] FIG. 7B illustrates another example knowledge graph in which this same data may be applied to similar cases that use differing law firms or that form the basis of selecting a law firm by a client. Analysis of the arguments made by a particular law firm 721, or a particular lawyer

722-723 in similar cases may provide a mechanism to identify the attorneys 722-723 who present the strongest arguments separate from the rulings 736 made in a case, or an argument that has mixed results 732, 731. The results of an action, motion, or the like depends upon the facts and witnesses as well as the arguments made by counsel 732, 731. The stronger work product and arguments may be made by attorneys for clients who do not prevail because of a weak factual situation. Hiring the best attorney for a case, or understanding the abilities of opposing counsel, can help clients resolve cases of all types to better negotiated resolutions. A particular judge who ruled over relevant cases may also be examined 737 for the purposes of determining the sorts of arguments and cases a given judge that may be relevant to a current case has ruled on, and how those rulings ended.

[0053] Inherent in both of the about uses of this NLP analysis is the analysis of the best legal arguments that may be made for a given set of facts. FIG. 7B illustrates one such use of an exemplary knowledge graph. Thus, attorneys performing legal research in support of development of a litigation or negotiation strategy may analyze relevant case-law 733-736 to identify the strongest approach. Fact comparison of current facts and ones found in various opinions may result in identification of specific factual circumstances that permit a particular opinion to be applicable or distinguishable that may not be as readily apparent upon an initial reading of an opinion. All relationships of facts to opinions can be readily explored when the data is analyzed and represented by a knowledge graph of the present invention.

[0054] FIG. 7C illustrates one example knowledge graph in which analysis of contracts 761, 763, insurance policies, and similar agreements of all types that may be processed as part of any negotiation process. If the universe of available agreements is large enough, facts related to litigation 773, 774, 775, 776 over various provisions within the agreement 771-772, may be present and permit identification of the facts and circumstances that gave rise to a dispute. From the knowledge of the relevant facts, parties can determine how likely such facts arising may be, and whether other approaches may mitigate the risk of an adverse result be identified or adopted, such as a judge's ruling on related cases 777. This type of predictive analysis uses data beyond just the agreement language but benefits from consideration of many diverse data sources that would be otherwise difficult to analyze.

[0055] The examples within FIG. 7A-7C are simplified to emphasize the points described herein, these various graphs 700, 710, 740 may be smaller subsets of a larger and more complex graph as a large universe of facts and legal documents are entered into the system. Any other relationships between the disputes, agreements, parties, attorneys and the like may be separately explored as part of an analysis activity. In such analysis, many hard to otherwise identify relationships may be discovered that are relevant to a case at hand.

[0056] FIG. 8 is an architecture diagram illustrating an exemplary system 800 for automated analysis of legal documents within and across different fields, according to aspect. System 800 includes two main functions: a hierarchical extraction and semantification processor 803 to identify, extract knowledge from data contained in legal documents 801 and transform it into a common data form and an analyzer 804 a development of the local 811 and global

knowledge graphs **812** containing the key entities, relationships and concepts encoded in the text.

[0057] Extraction processor **803** performs a set of systematic NLP-based data extraction single-purpose generic micro-functions including Tokenizer **831**, Acronym Normalizer **832**, Lemmatizer **833**, Name Entity Recognizer (NER) **834**, pattern recognizer **836**, and a rules processor **836**. Tokenizer **831**, given a character sequence and a defined document unit, tokenizes the character sequence up into pieces, called tokens, and possibly, at the same time, throwing away certain characters, such as punctuation. Acronym Normalizer **832** transforms all acronyms found in the incoming legal documents into standard set of terms applicable to all of the data regardless of source. Lemmatizer **833** transforming language within the documents to properly use a vocabulary and morphological analysis of words, normally aiming to remove inflectional endings only, and to return the base or dictionary form of a word. Name Entity Recognizer (NER) **834** identifies references to known people and entities within the documents, regardless of the form of the name. For example, reference to IBM or Apple and IBM Corp. and Apple Inc. will identified as referring to the same respective entities. Similar variations in references to an individual's name, including use or omission of middle initials or Jr. Pattern recognizer **835** performs other structured term-extraction features to document-wide semantic NLP pattern recognition macro-functions including sentiment and topic extraction, as well as targeted word/sentence clustering and information retrieval. Rules processor **836** performs system and user defined data transformation and orchestration workflows.

[0058] The results of hierarchical extraction and semantification processor **803** allow a model selection analyzer **861**, within analysis processor **804**, to perform dynamic model selection based on a series of more efficient classification types of algorithms which look at estimating the domain, age, legal jurisdictions, etc. . . . associated with a document and applying relevant NER, gazetteers, and ontologies. This dynamic model selection enables a dynamic algorithm processor **762** to effectively query a catalogue of available models **821** and recommend an available model to best extract, parse, interpret, schematize, normalize and then semantify the data as discussed above in FIG. 6 with a specialized natural language processor **863**, term interpreter **864**, and risk estimator **865**. The recommended model may have been trained already or is dynamically trained on available source data and labels.

[0059] Domain specific NLP processor **863** may feed legal and domain-specific technical data into workflows for both knowledge graph enrichment and dataset contextualization, together with a local and global graph generator **841**, **842**. Such graph generators **841**, **842** take data and the results of processes done by other components in an analysis processor **861-865** and may produce localized knowledge graphs for specific groups of data, or global graphs for wider ranges of data and graph-edges. These processes are only possible by using NLP-based tagging and mapping capabilities to provide a bridge between raw/semi-processed datasets and context-aware graph ontologies. Ultimately, the analysis processor **804** continuously enhance these knowledge bases through feedback loops with new data from systematic events, so that the development of local **811** and global knowledge graphs **812** can be both informed by, and inform, the extraction and analysis processes.

[0060] System **800** leverages the hierarchical extraction and semantification processor **803** to map raw legal document data to our domain specific languages (DSL). Use of the DSLs allows for capturing individual different levels of granularity in the knowledge graphs **811-812** within specific investment products in legal, finance, or multi-level risk insurance policies. Within these DSLs, and at each of these levels, the analysis processor **804** tags individual clauses or terms with contextual information, and flags problematic terms according to both endogenous ambiguity where historical information or legal precedent isn't accessible or existent, as well as exogenous risk dimensions that are specific to these industries.

[0061] Domain-language ambiguity is addressed by establishing an array of more clear-cut interpretations of that vague clause, using likelihood values that estimate a valuation distribution based on the document's language. Specific dictionaries **822** for each legal specialty provide additional data and term definition for use in processing any particular legal document. System **800** captures systemic risk changes through time-varying pattern analysis where the system can map a cross-sectional snapshot of the current state of the system's events, be it natural catastrophe incidents, political & market sentiment or regulatory and macro-prudential policy changes, to the clause or term affecting the valuation/pricing of a given product/policy. These approaches explore the state space of pricing/valuation possibilities with a dimensionality beyond what individual agents can scale to, utilizing rule-based thresholds to make efficient use of human capital to review a targeted subset of valuation or loss estimation results.

[0062] FIG. 9 is a flowchart diagram illustrating an exemplary method **900** for automated analysis of legal documents within and across different fields, according to aspect. The method begins with step **901** receiving a data set of legal documents. Step **902** identifies and extracts a set of knowledge data from data contained in the legal document. The knowledge data is transformed by step **903** into a common data form. Step **904** selects a dynamic model used in estimating a domain, age, and legal jurisdictions for the knowledge data and step **905** queries a catalogue of available models.

[0063] A knowledge graph is generated from the knowledge data by step **906**, Step **907** performs knowledge graph enrichment and dataset contextualization using legal and domain-specific technical data. The method concludes with step **908** providing one or more interpretations for a provision within the knowledge data.

[0064] FIG. 10 is a flowchart diagram illustrating an exemplary method **1000** for automated analysis of legal documents within and across different fields, according to aspect. Step **1001** tokenizes a character sequence up into pieces, and throwing away characters, such as punctuation. Step **1002** transforms all acronyms found in the incoming legal documents into standard set of terms applicable to all of the data regardless of source. Language within the knowledge data is transformed by step **1003** to properly use a vocabulary and morphological analysis of words.

[0065] Step **1004** identifies references to known people and entities within the documents, regardless of a form of the reference. Step **1005** performed pattern recognition on the knowledge data including sentiment and topic extraction, targeted word/sentence clustering, and information

retrieval. Step **1006** applies system and user defined data transformation and orchestration workflows.

Hardware Architecture

[0066] Generally, the techniques disclosed herein may be implemented on hardware or a combination of software and hardware. For example, they may be implemented in an operating system kernel, in a separate user process, in a library package bound into network applications, on a specially constructed machine, on an application-specific integrated circuit (ASIC), or on a network interface card.

[0067] Software/hardware hybrid implementations of at least some of the aspects disclosed herein may be implemented on a programmable network-resident machine (which should be understood to include intermittently connected network-aware machines) selectively activated or reconfigured by a computer program stored in memory. Such network devices may have multiple network interfaces that may be configured or designed to utilize different types of network communication protocols. A general architecture for some of these machines may be described herein in order to illustrate one or more exemplary means by which a given unit of functionality may be implemented. According to specific aspects, at least some of the features or functionalities of the various aspects disclosed herein may be implemented on one or more general-purpose computers associated with one or more networks, such as for example an end-user computer system, a client computer, a network server or other server system, a mobile computing device (e.g., tablet computing device, mobile phone, smartphone, laptop, or other appropriate computing device), a consumer electronic device, a music player, or any other suitable electronic device, router, switch, or other suitable device, or any combination thereof. In at least some aspects, at least some of the features or functionalities of the various aspects disclosed herein may be implemented in one or more virtualized computing environments (e.g., network computing clouds, virtual machines hosted on one or more physical computing machines, or other appropriate virtual environments).

[0068] Referring now to FIG. 11, there is shown a block diagram depicting an exemplary computing device **10** suitable for implementing at least a portion of the features or functionalities disclosed herein. Computing device **10** may be, for example, any one of the computing machines listed in the previous paragraph, or indeed any other electronic device capable of executing software- or hardware-based instructions according to one or more programs stored in memory. Computing device **10** may be configured to communicate with a plurality of other computing devices, such as clients or servers, over communications networks such as a wide area network a metropolitan area network, a local area network, a wireless network, the Internet, or any other network, using known protocols for such communication, whether wireless or wired.

[0069] In one aspect, computing device **10** includes one or more central processing units (CPU) **12**, one or more interfaces **15**, and one or more busses **14** (such as a peripheral component interconnect (PCI) bus). When acting under the control of appropriate software or firmware, CPU **12** may be responsible for implementing specific functions associated with the functions of a specifically configured computing device or machine. For example, in at least one aspect, a computing device **10** may be configured or designed to

function as a server system utilizing CPU **12**, local memory **11** and/or remote memory **16**, and interface(s) **15**. In at least one aspect, CPU **12** may be caused to perform one or more of the different types of functions and/or operations under the control of software modules or components, which for example, may include an operating system and any appropriate applications software, drivers, and the like.

[0070] CPU **12** may include one or more processors **13** such as, for example, a processor from one of the Intel, ARM, Qualcomm, and AMD families of microprocessors. In some aspects, processors **13** may include specially designed hardware such as application-specific integrated circuits (ASICs), electrically erasable programmable read-only memories (EEPROMs), field-programmable gate arrays (FPGAs), and so forth, for controlling operations of computing device **10**. In a particular aspect, a local memory **11** (such as non-volatile random access memory (RAM) and/or read-only memory (ROM), including for example one or more levels of cached memory) may also form part of CPU **12**. However, there are many different ways in which memory may be coupled to system **10**. Memory **11** may be used for a variety of purposes such as, for example, caching and/or storing data, programming instructions, and the like. It should be further appreciated that CPU **12** may be one of a variety of system-on-a-chip (SOC) type hardware that may include additional hardware such as memory or graphics processing chips, such as a QUALCOMM SNAP-DRAGON™ or SAMSUNG EXYNOS™ CPU as are becoming increasingly common in the art, such as for use in mobile devices or integrated devices.

[0071] As used herein, the term “processor” is not limited merely to those integrated circuits referred to in the art as a processor, a mobile processor, or a microprocessor, but broadly refers to a microcontroller, a microcomputer, a programmable logic controller, an application-specific integrated circuit, and any other programmable circuit.

[0072] In one aspect, interfaces **15** are provided as network interface cards (NICs). Generally, NICs control the sending and receiving of data packets over a computer network; other types of interfaces **15** may for example support other peripherals used with computing device **10**. Among the interfaces that may be provided are Ethernet interfaces, frame relay interfaces, cable interfaces, DSL interfaces, token ring interfaces, graphics interfaces, and the like. In addition, various types of interfaces may be provided such as, for example, universal serial bus (USB), Serial, Ethernet, FIREWIRE™, THUNDERBOLT™, PCI, parallel, radio frequency (RF), BLUETOOTH™, near-field communications (e.g., using near-field magnetics), 802.11 (WiFi), frame relay, TCP/IP, ISDN, fast Ethernet interfaces, Gigabit Ethernet interfaces, Serial ATA (SATA) or external SATA (ESATA) interfaces, high-definition multimedia interface (HDMI), digital visual interface (DVI), analog or digital audio interfaces, asynchronous transfer mode (ATM) interfaces, high-speed serial interface (HSSI) interfaces, Point of Sale (POS) interfaces, fiber data distributed interfaces (FDDIs), and the like. Generally, such interfaces **15** may include physical ports appropriate for communication with appropriate media. In some cases, they may also include an independent processor (such as a dedicated audio or video processor, as is common in the art for high-fidelity A/V hardware interfaces) and, in some instances, volatile and/or non-volatile memory (e.g., RAM).

[0073] Although the system shown in FIG. 11 illustrates one specific architecture for a computing device 10 for implementing one or more of the aspects described herein, it is by no means the only device architecture on which at least a portion of the features and techniques described herein may be implemented. For example, architectures having one or any number of processors 13 may be used, and such processors 13 may be present in a single device or distributed among any number of devices. In one aspect, a single processor 13 handles communications as well as routing computations, while in other aspects a separate dedicated communications processor may be provided. In various aspects, different types of features or functionalities may be implemented in a system according to the aspect that includes a client device (such as a tablet device or smartphone running client software) and server systems (such as a server system described in more detail below).

[0074] Regardless of network device configuration, the system of an aspect may employ one or more memories or memory modules (such as, for example, remote memory block 16 and local memory 11) configured to store data, program instructions for the general-purpose network operations, or other information relating to the functionality of the aspects described herein (or any combinations of the above). Program instructions may control execution of or comprise an operating system and/or one or more applications, for example. Memory 16 or memories 11, 16 may also be configured to store data structures, configuration data, encryption data, historical system operations information, or any other specific or generic non-program information described herein.

[0075] Because such information and program instructions may be employed to implement one or more systems or methods described herein, at least some network device aspects may include nontransitory machine-readable storage media, which, for example, may be configured or designed to store program instructions, state information, and the like for performing various operations described herein. Examples of such nontransitory machine-readable storage media include, but are not limited to, magnetic media such as hard disks, floppy disks, and magnetic tape; optical media such as CD-ROM disks; magneto-optical media such as optical disks, and hardware devices that are specially configured to store and perform program instructions, such as read-only memory devices (ROM), flash memory (as is common in mobile devices and integrated systems), solid state drives (SSD) and “hybrid SSD” storage drives that may combine physical components of solid state and hard disk drives in a single hardware device (as are becoming increasingly common in the art with regard to personal computers), memristor memory, random access memory (RAM), and the like. It should be appreciated that such storage means may be integral and non-removable (such as RAM hardware modules that may be soldered onto a motherboard or otherwise integrated into an electronic device), or they may be removable such as swappable flash memory modules (such as “thumb drives” or other removable media designed for rapidly exchanging physical storage devices), “hot-swappable” hard disk drives or solid state drives, removable optical storage discs, or other such removable media, and that such integral and removable storage media may be utilized interchangeably. Examples of program instructions include both object code, such as may be produced by a compiler, machine code, such as may be produced by an

assembler or a linker, byte code, such as may be generated by for example a JAVA™ compiler and may be executed using a Java virtual machine or equivalent, or files containing higher level code that may be executed by the computer using an interpreter (for example, scripts written in Python, Perl, Ruby, Groovy, or any other scripting language).

[0076] In some aspects, systems may be implemented on a standalone computing system. Referring now to FIG. 12, there is shown a block diagram depicting a typical exemplary architecture of one or more aspects or components thereof on a standalone computing system. Computing device 20 includes processors 21 that may run software that carry out one or more functions or applications of aspects, such as for example a client application 24. Processors 21 may carry out computing instructions under control of an operating system 22 such as, for example, a version of MICROSOFT WINDOWS™ operating system, APPLE macOS™ or iOS™ operating systems, some variety of the Linux operating system, ANDROID™ operating system, or the like. In many cases, one or more shared services 23 may be operable in system 20, and may be useful for providing common services to client applications 24. Services 23 may for example be WINDOWS™ services, user-space common services in a Linux environment, or any other type of common service architecture used with operating system 21. Input devices 28 may be of any type suitable for receiving user input, including for example a keyboard, touchscreen, microphone (for example, for voice input), mouse, touchpad, trackball, or any combination thereof. Output devices 27 may be of any type suitable for providing output to one or more users, whether remote or local to system 20, and may include for example one or more screens for visual output, speakers, printers, or any combination thereof. Memory 25 may be random-access memory having any structure and architecture known in the art, for use by processors 21, for example to run software. Storage devices 26 may be any magnetic, optical, mechanical, memristor, or electrical storage device for storage of data in digital form (such as those described above, referring to FIG. 11). Examples of storage devices 26 include flash memory, magnetic hard drive, CD-ROM, and/or the like.

[0077] In some aspects, systems may be implemented on a distributed computing network, such as one having any number of clients and/or servers. Referring now to FIG. 13, there is shown a block diagram depicting an exemplary architecture 30 for implementing at least a portion of a system according to one aspect on a distributed computing network. According to the aspect, any number of clients 33 may be provided. Each client 33 may run software for implementing client-side portions of a system; clients may comprise a system 20 such as that illustrated in FIG. 12. In addition, any number of servers 32 may be provided for handling requests received from one or more clients 33. Clients 33 and servers 32 may communicate with one another via one or more electronic networks 31, which may be in various aspects any of the Internet, a wide area network, a mobile telephony network (such as CDMA or GSM cellular networks), a wireless network (such as WiFi, WiMAX, LTE, and so forth), or a local area network (or indeed any network topology known in the art; the aspect does not prefer any one network topology over any other). Networks 31 may be implemented using any known network protocols, including for example wired and/or wireless protocols.

[0078] In addition, in some aspects, servers 32 may call external services 37 when needed to obtain additional information, or to refer to additional data concerning a particular call. Communications with external services 37 may take place, for example, via one or more networks 31. In various aspects, external services 37 may comprise web-enabled services or functionality related to or installed on the hardware device itself. For example, in one aspect where client applications 24 are implemented on a smartphone or other electronic device, client applications 24 may obtain information stored in a server system 32 in the cloud or on an external service 37 deployed on one or more of a particular enterprise's or user's premises. In addition to local storage on servers 32, remote storage 38 may be accessible through the network(s) 31.

[0079] In some aspects, clients 33 or servers 32 (or both) may make use of one or more specialized services or appliances that may be deployed locally or remotely across one or more networks 31. For example, one or more databases 34 in either local or remote storage 38 may be used or referred to by one or more aspects. It should be understood by one having ordinary skill in the art that databases in storage 34 may be arranged in a wide variety of architectures and using a wide variety of data access and manipulation means. For example, in various aspects one or more databases in storage 34 may comprise a relational database system using a structured query language (SQL), while others may comprise an alternative data storage technology such as those referred to in the art as "NoSQL" (for example, HADOOP CASSANDRA™, GOOGLE BIGTABLE™, and so forth). In some aspects, variant database architectures such as column-oriented databases, in-memory databases, clustered databases, distributed databases, or even flat file data repositories may be used according to the aspect. It will be appreciated by one having ordinary skill in the art that any combination of known or future database technologies may be used as appropriate, unless a specific database technology or a specific arrangement of components is specified for a particular aspect described herein. Moreover, it should be appreciated that the term "database" as used herein may refer to a physical database machine, a cluster of machines acting as a single database system, or a logical database within an overall database management system. Unless a specific meaning is specified for a given use of the term "database", it should be construed to mean any of these senses of the word, all of which are understood as a plain meaning of the term "database" by those having ordinary skill in the art.

[0080] Similarly, some aspects may make use of one or more security systems 36 and configuration systems 35. Security and configuration management are common information technology (IT) and web functions, and some amount of each are generally associated with any IT or web systems. It should be understood by one having ordinary skill in the art that any configuration or security subsystems known in the art now or in the future may be used in conjunction with aspects without limitation, unless a specific security 36 or configuration system 35 or approach is specifically required by the description of any specific aspect.

[0081] FIG. 14 shows an exemplary overview of a computer system 40 as may be used in any of the various locations throughout the system. It is exemplary of any computer that may execute code to process data. Various

modifications and changes may be made to computer system 40 without departing from the broader scope of the system and method disclosed herein. Central processor unit (CPU) 41 is connected to bus 42, to which bus is also connected memory 43, nonvolatile memory 44, display 47, input/output (I/O) unit 48, and network interface card (NIC) 53. I/O unit 48 may, typically, be connected to peripherals such as a keyboard 49, pointing device 50, hard disk 52, real-time clock 51, a camera 57, and other peripheral devices. NIC 53 connects to network 54, which may be the Internet or a local network, which local network may or may not have connections to the Internet. The system may be connected to other computing devices through the network via a router 55, wireless local area network 56, or any other network connection. Also shown as part of system 40 is power supply unit 45 connected, in this example, to a main alternating current (AC) supply 46. Not shown are batteries that could be present, and many other devices and modifications that are well known but are not applicable to the specific novel functions of the current system and method disclosed herein. It should be appreciated that some or all components illustrated may be combined, such as in various integrated applications, for example Qualcomm or Samsung system-on-a-chip (SOC) devices, or whenever it may be appropriate to combine multiple capabilities or functions into a single hardware device (for instance, in mobile devices such as smartphones, video game consoles, in-vehicle computer systems such as navigation or multimedia systems in automobiles, or other integrated hardware devices).

[0082] In various aspects, functionality for implementing systems or methods of various aspects may be distributed among any number of client and/or server components. For example, various software modules may be implemented for performing various functions in connection with the system of any particular aspect, and such modules may be variously implemented to run on server and/or client components.

[0083] The skilled person will be aware of a range of possible modifications of the various aspects described above. Accordingly, the present invention is defined by the claims and their equivalents.

What is claimed is:

1. A system for automated analysis of legal documents within and across different fields, comprising:

a computer system comprising at least one memory, at least one processor, and at least a first plurality of programming instructions stored in the at least one memory and operating on the at least one processor configured to allow the operation on the computer system of additional programming instructions;

an extraction processor, comprising at least a second plurality of programming instructions stored in the at least one memory of, and operating on at least one processor of, the computer system, wherein the second plurality of programming instructions, when operating on the at least one processor, cause the computer system to:

receive a data set containing legal documents;

perform a set of systematic NLP-based data extraction single-purpose generic micro-functions to identify, extract a set of knowledge data from data contained in the legal document; and

transform the knowledge data into a common data form; and

an analysis processor, comprising at least a second plurality of programming instructions stored in the at least one memory of, and operating on at least one processor of, the computer system, wherein the second plurality of programming instructions, when operating on the at least one processor, cause the computer system to:

perform dynamic model selection based on classification algorithms estimating a domain, age, and legal jurisdictions for the knowledge data;
 query a catalogue of available models;
 generate a knowledge graph from the knowledge data;
 perform knowledge graph enrichment and dataset contextualization using legal and domain-specific technical data; and
 using knowledge data to provide one or more interpretations for a provision within the knowledge data.

2. The system of claim 1, wherein the set of systematic NLP-based data extraction generic micro-functions comprises:

tokenizing a character sequence up into pieces, and throwing away characters, such as punctuation;
 transforming all acronyms found in the incoming legal documents into standard set of terms applicable to all of the data regardless of source;
 transforming language within the knowledge data to properly use a vocabulary and morphological analysis of words;
 identifying references to known people and entities within the documents, regardless of a form of the reference;
 recognizing patterns by sentiment and topic extraction, targeted word/sentence clustering, and information retrieval; and
 applying system and user defined data transformation and orchestration workflows.

3. The system of claim 1, wherein the knowledge graph generator comprises a local graph generator and a global graph generator.

4. The system of claim 1, wherein the data set of legal documents includes contracts, court decisions, statutes, court pleadings, and text data.

5. The system of claim 1, wherein the extraction processor further comprises a specialty dictionary related to a specific legal subject matter.

6. A method for automated analysis of legal documents within and across different fields, comprising the steps of: receiving a data set of legal documents;
 identifying and extracting a set of knowledge data from data contained in the legal document;
 transforming the knowledge data into a common data form;
 selecting a dynamic model used in estimating a domain, age, and legal jurisdictions for the knowledge data;
 querying a catalogue of available models;
 generating a knowledge graph from the knowledge data;
 performing knowledge graph enrichment and dataset contextualization using legal and domain-specific technical data; and
 providing one or more interpretations for a provision within the knowledge data.

7. The method of claim 6, wherein the method further comprises:

tokenizing a character sequence up into pieces, and throwing away characters, such as punctuation;
 transforming all acronyms found in the incoming legal documents into standard set of terms applicable to all of the data regardless of source;
 transforming language within the knowledge data to properly use a vocabulary and morphological analysis of words;
 identifying references to known people and entities within the documents, regardless of a form of the reference;
 performing pattern recognition on the knowledge data including sentiment and topic extraction, targeted word/sentence clustering, and information retrieval; and
 applying system and user defined data transformation and orchestration workflows.

8. The method of claim 7, wherein the knowledge graph comprises a local knowledge graph and a global knowledge graph.

9. The method of claim 7, wherein the data set of legal documents includes contracts, court decisions, statutes, court pleadings, and text data.

10. The method of claim 7, wherein the extraction of the set of knowledge data utilizes a specialty dictionary related to a specific legal subject matter.

* * * * *



US 20200019601A1

(19) **United States**(12) **Patent Application Publication**
TSU et al.(10) **Pub. No.: US 2020/0019601 A1**(43) **Pub. Date: Jan. 16, 2020**(54) **METHOD & SYSTEM FOR LABELING AND ORGANIZING DATA FOR SUMMARIZING AND REFERENCING CONTENT VIA A COMMUNICATION NETWORK****Publication Classification**(51) **Int. Cl.**
G06F 17/24 (2006.01)
G06F 16/38 (2006.01)
(52) **U.S. Cl.**
CPC **G06F 17/241** (2013.01); **G06F 16/382** (2019.01)(71) Applicant: **PowerNotes LLC**, Chicago, IL (US)(72) Inventors: **Wilson TSU**, Chicago, IL (US);
Srujani PAGIDIPATI, Chicago, IL (US)(21) Appl. No.: **16/584,556**(22) Filed: **Sep. 26, 2019****Related U.S. Application Data**

(63) Continuation of application No. 14/215,843, filed on Mar. 17, 2014.

(60) Provisional application No. 61/793,717, filed on Mar. 15, 2013.

(57) **ABSTRACT**

Systems and methods of the present technology generally provide computer implemented assistance for data summary, including organizing and generating a summary of data selected from legal source documents. A system server may receive user-identified information selected from one or more legal source documents at a user device, along with at least one unique identifier selected from a unique identifier selection box via a communication link. The server may then generate a summary that includes the unique identifier and the user-identified information, and may send the summary via the communication link to be displayed at the user device.

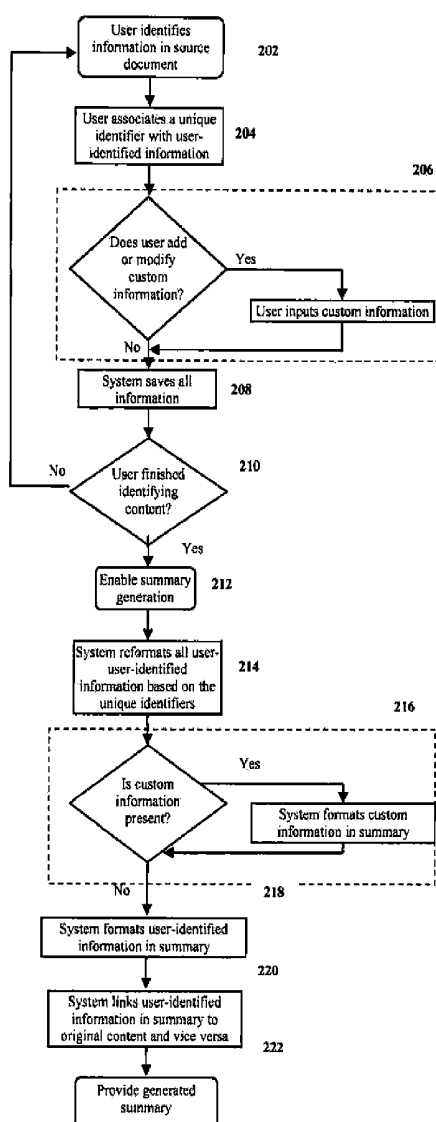


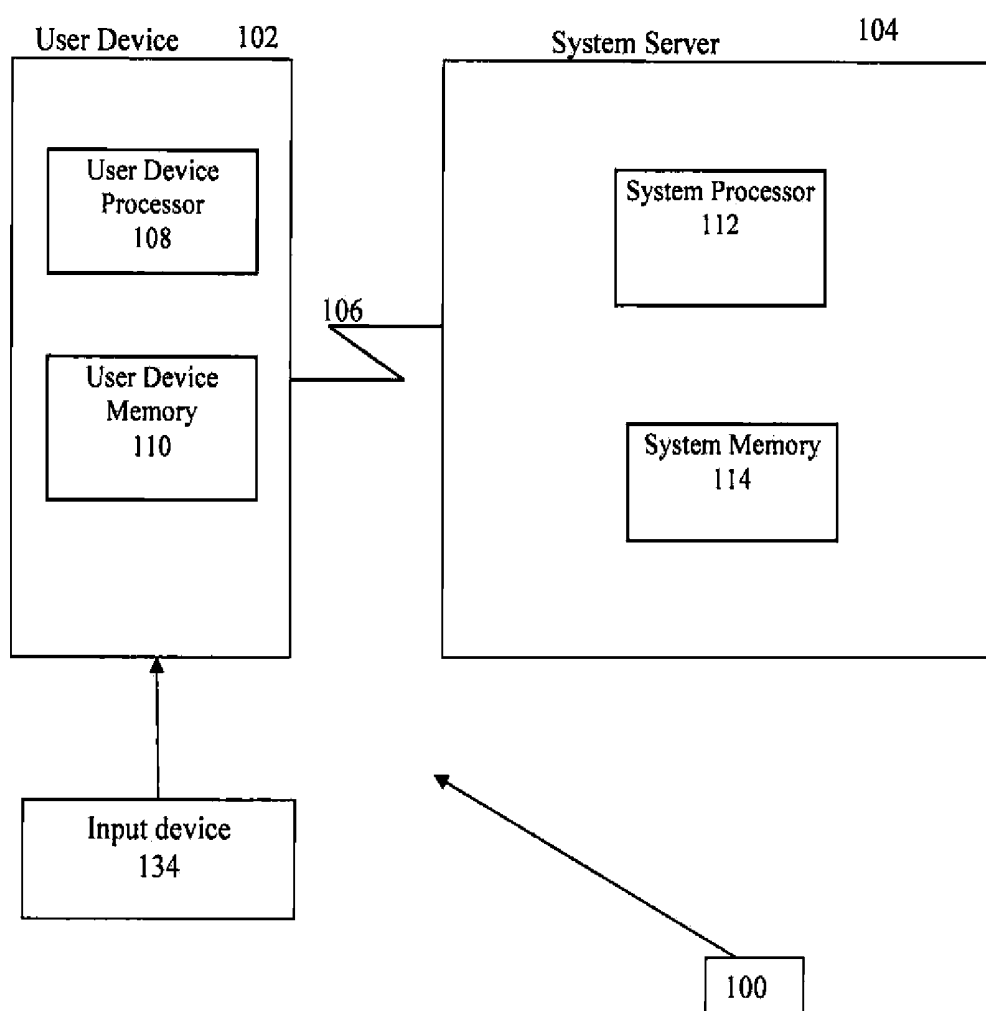
FIGURE 1

FIGURE 2

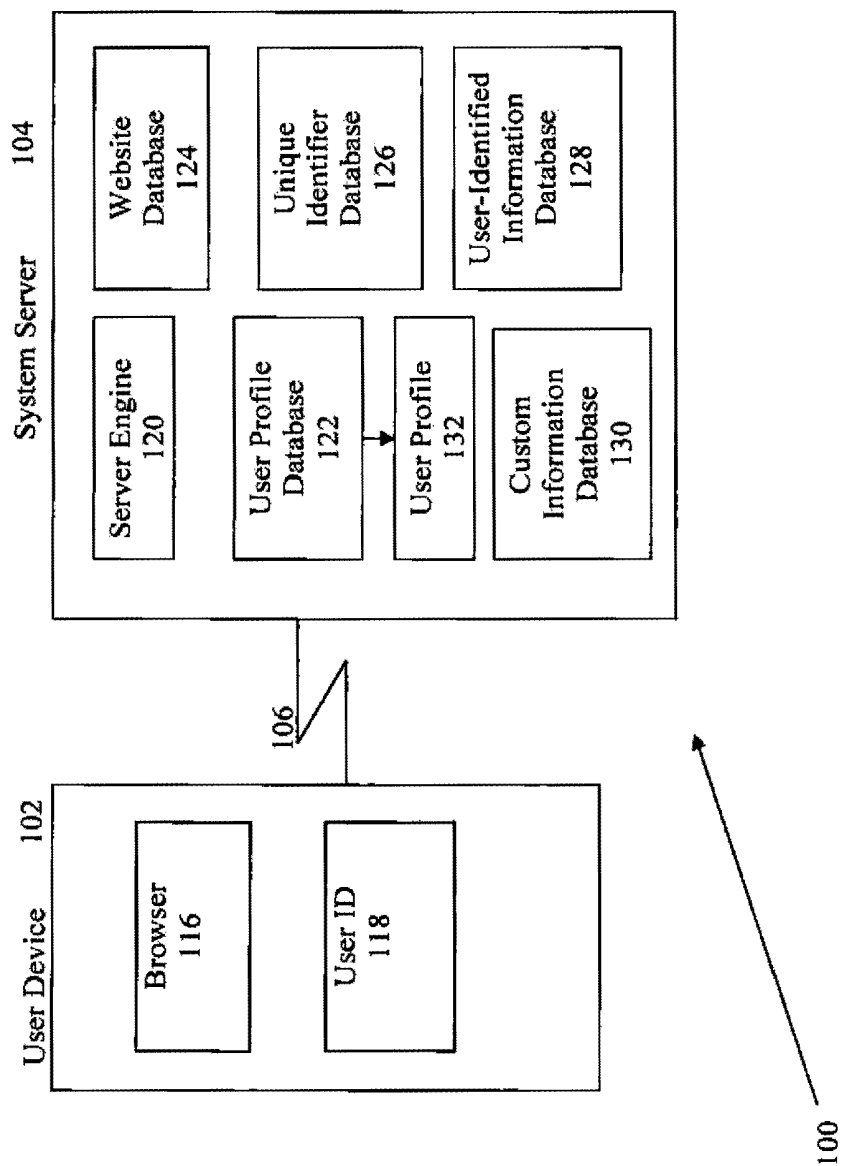
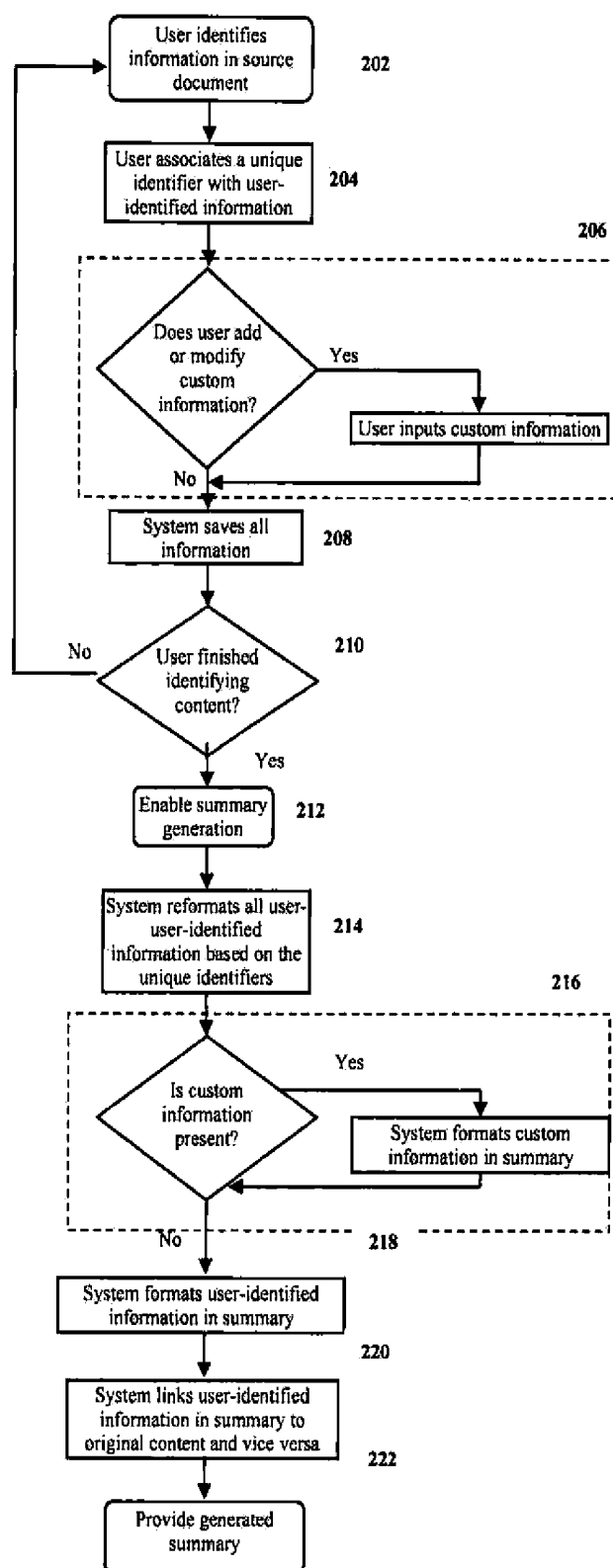


FIGURE 3



The issue is, what is chicken? Plaintiff says "chicken" means a young chicken, suitable for broiling and frying. Defendant says "chicken" means any bird of that genus that meets contract specifications on weight and quality, including what it calls "stewing chicken" and plaintiff pejoratively terms "fowl". Dictionaries give both meanings, as well as some others not relevant here. To support its, plaintiff sends a number of volleys over the net; defendant essays to return them and adds a few serves of its own. Assuming that both parties were acting in good faith, the case nicely illustrates Holmes' remark that the meaning of a contract depends not on the agreement of two intentions, but on the agreement of two sets of external facts. The parties' having meant the same thing but on the same thing." The Path of the Law, in Collected Legal Essays. The court has concluded that plaintiff has not sustained its burden of persuasion that the contract used "chicken" in the narrower sense.

**FACT
ISSUE
PROCEDURE
REASONING
RULE
HOLDING
NOTES**

302

cing
an
the
al

300

304

The action is for breach of the warranty that goods sold shall correspond to the description, New York Personal Property Law, McKinney's Consol. Laws, c. 41, § 95. Two contracts are in suit. In the first, dated May 2, 1957, defendant, a New York sales corporation, confirmed the sale to plaintiff, a Swiss corporation, of

FIGURE 4

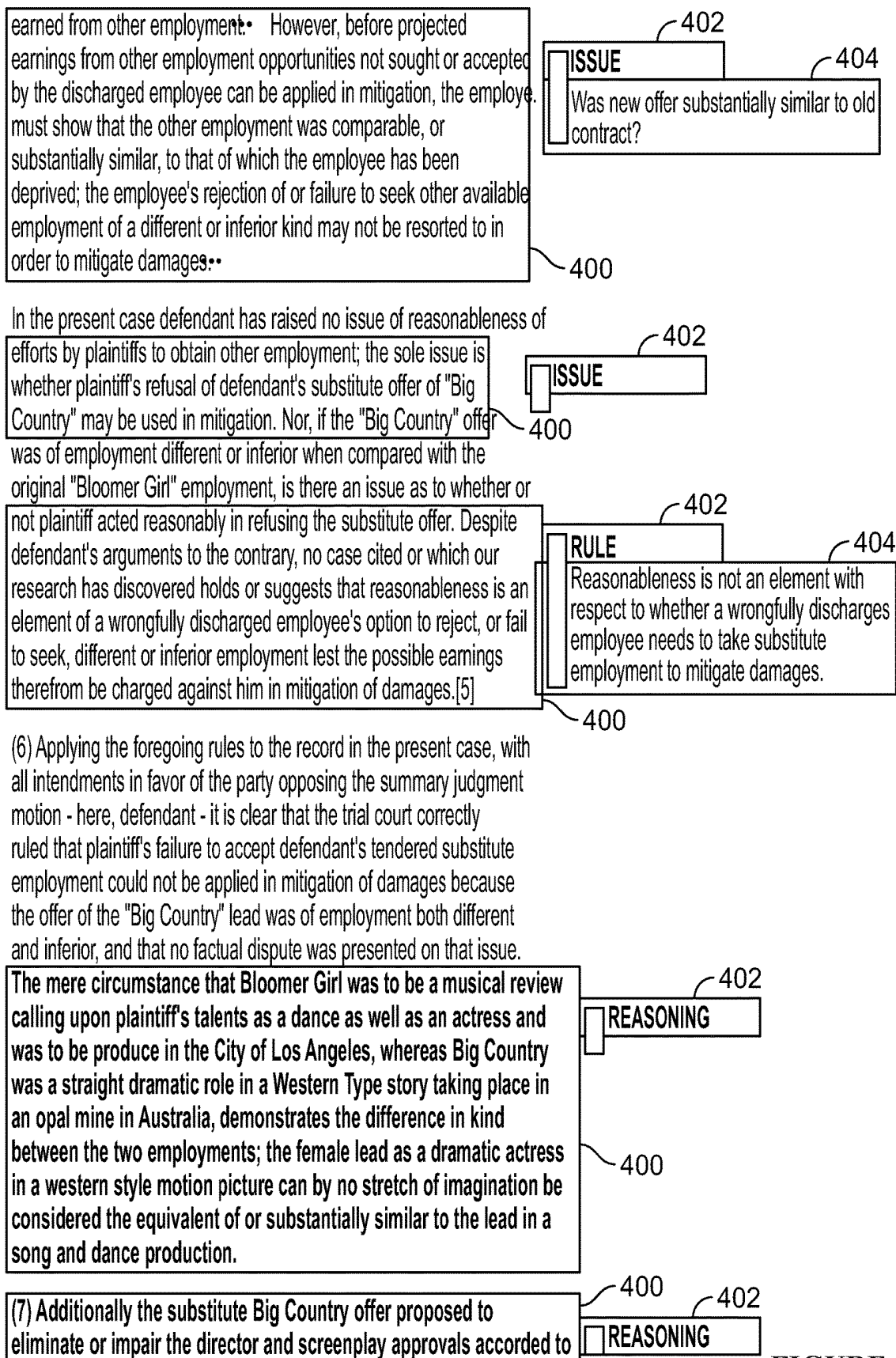


FIGURE 5

500

Issues

Was new offer substantially similar to old contract?

"However, before projected earnings from other employment opportunities not sought or accepted by the discharged employee can be applied in mitigation, the employer must show that the other employment was comparable, or substantially similar, to that of which the employee has been deprived the employee's rejection of a failure to seek other available employment of a different or inferior kind may not be resorted to in order to mitigate damages."

"The sole issue is whether plaintiff's refusal of defendant's is substitute offer of "Big Country may be used in mitigation."

Procedure

Reasoning

"The mere circumstance that "Bloomer Girl" was to be a musical review calling upon plaintiff's talents as a dancer as well as an actress, and was to be produced in the City of Los Angeles whereas "Big Country" was a straight dramatic role in a "Western Type" story taking place in an opal mine in Australia, demonstrates the difference in kind between the two employments, the female lead as a dramatic actress in a western style motion picture can by no stretch of imagination be considered the equivalent of or substantially similar to the lead in a song-and-dance production"

"Additionally, the substitute "Big Country" offer proposed to eliminate or impair the director and screenplay approvals accorded to plaintiff's under the original "Blommer Girl" contract (see in 2 anle), and thus constituted an offer of inferior employment"

Rule

Reasonableness is not an element with respect to whether a wrongfully discharged employee needs to take substitute employment to mitigate dama

"Despite defendant's arguments to the contrary, no case cited or which our research has discovered holds or suggests that reasonableness is an element of a wrongfully discharged employee's option to reject, or fail to seek, different or inferior employment test the possible earnings therefrom be charged against him in mitigation of damages.[5]"

Holding

FIGURE 6

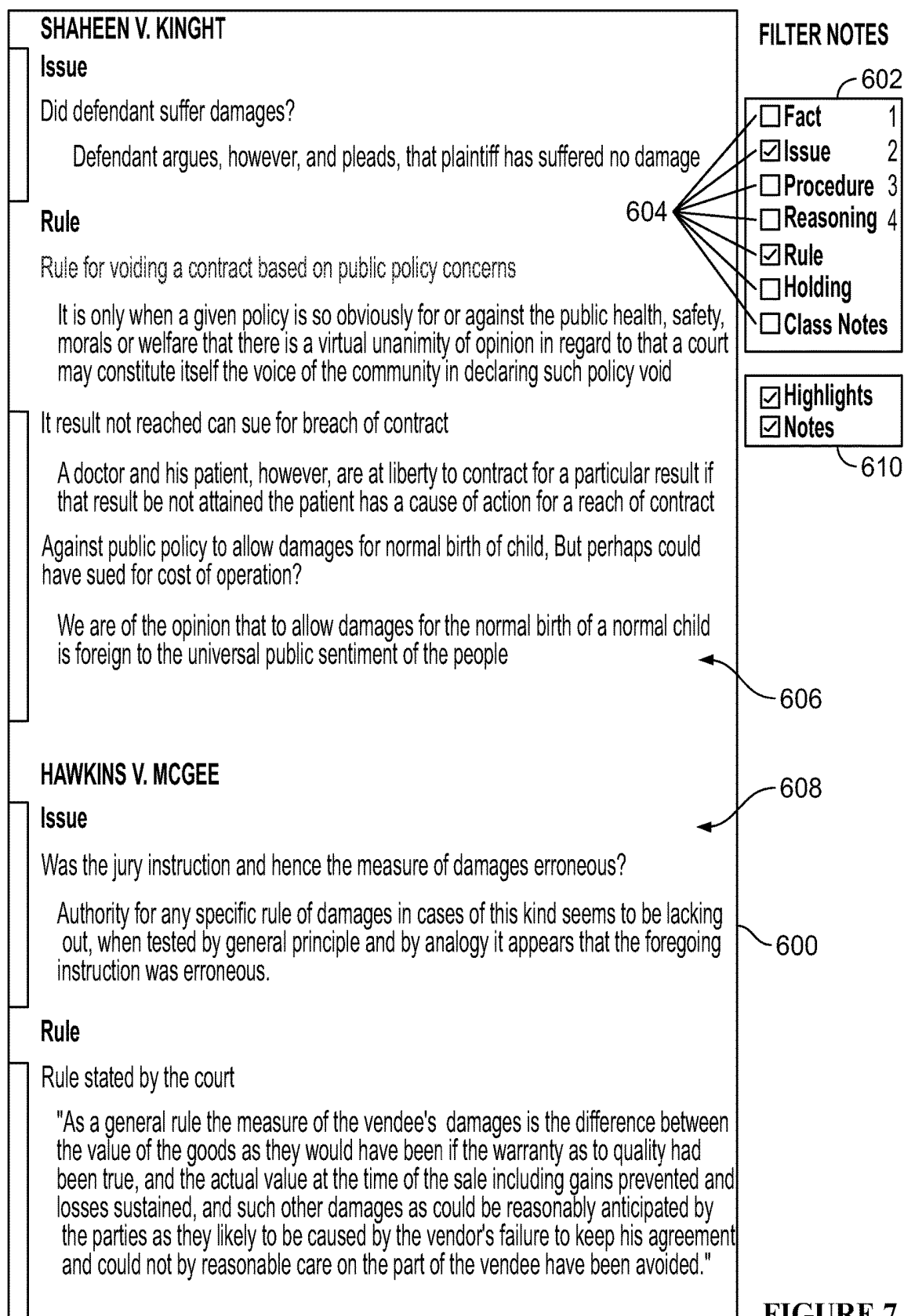


FIGURE 7

700

♥ **** *
706

Enter class notes

* ***** *

SAVE NOTE

704

Issues

Was new offer substantially similar to old contract?
"However before projected earnings from othe."
"the sole issue is whether plaintiff's refuse."

Procedure

Reasoning

"The mere circumstance that "Bloomer Girl" wa..."
"Additionally, the substitute "Big Country" o.."

Rule

Reasonableness is not an element with respect to whether a wrongfully discharged employee needs to take substitute employment to mitigate damages.
"Despite defendant's arguments to the contrar ."

Holding

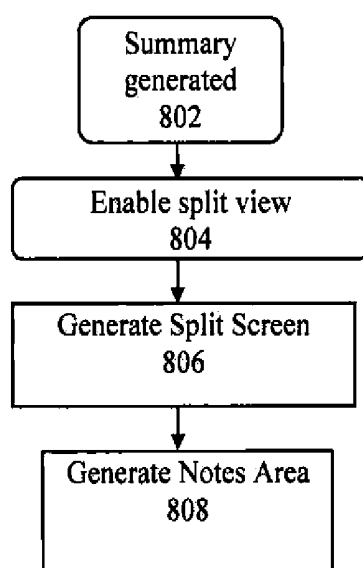
702

earned from other employment. ... However before projected earnings from other employment opportunities not sought or accepted by the discharged employee can be applied in mitigation, the employer must show that the other employment was comparable, or substantially similar, to that of which the employee has been deprived; the employee's rejection of or failure to seek other available employment of a different or inferior kind may not be resorted to in order to mitigate damage ...

In the present case defendant has raised no issue of reasonableness of efforts by plaintiffs obtain other employment; the sole issue is whether plaintiff's refusal of defendant's substitute offer of "Big Country" may be used mitigation. Nor, if the "Big Country" offer was of employment different or inferior when compared with the original "Bloomer Girl" employment, is there an issue as to whether or not plaintiff acted reasonably in refusing the substitute offer. Despite defendant's arguments to the contrary, no case cited or which our research has discovered holds or suggests that reasonableness is an element of

FIGURE 8

FIGURE 9



METHOD & SYSTEM FOR LABELING AND ORGANIZING DATA FOR SUMMARIZING AND REFERENCING CONTENT VIA A COMMUNICATION NETWORK

RELATED APPLICATIONS

[0001] This application is a continuation of U.S. patent application Ser. No. 14/215,843 filed on Mar. 17, 2014, currently pending, which claims priority to U.S. Provisional Application Ser. No. 61/793,717, filed on Mar. 15, 2013, the disclosure of which is hereby incorporated by reference in its entirety.

FIELD OF THE INVENTION

[0002] The present technology generally relates to computer-aided data summary systems and methods that assist users with analyzing data and, more particularly, to a method and system for obtaining, analyzing, labeling, organizing, combining, modifying, and/or summarizing data using single- and multi-action operations.

DESCRIPTION OF RELATED ART

[0003] Professionals and students often require the analysis and manipulation of data and information. Such individuals may practice or study in, without limitation, the fields of law, business, finance, medicine, legal education, business education, medical education, graduate education, undergraduate education, and K-12 education.

[0004] Existing technology generally requires users to: (i) either obtain pre-sorted data in excess of usability requirements or manually search for, gather, and organize multiple documents from various sources, (ii) obtain and utilize several tools to label and annotate such data, and/or (iii) manually organize, integrate, and summarize the original data with user-created data.

[0005] A need exists for a system and/or method that organizes information from multiple sources, allows users to easily and efficiently label such information, allows users to modify such information, allows users to combine and separate such information with other information, and displays all such information in a user-friendly format. Such a system, by automating and streamlining many labor-intensive steps involved in the above processes, would reduce the amount of work for and increase the efficiency of, users.

[0006] The systems and methods of the present technology can be used to automate and streamline many of these steps involved in these processes, through, but not limited to, single-action functionality.

SUMMARY OF THE INVENTION

[0007] Examples of the present technology can include components of a cloud-based (or other Internet-based) or device-based (such as a personal computer, tablet, smartphone, or other device) platform that provides a user with a streamlined interface/display with the functionality to display, label, organize, integrate, modify, and/or summarize information. Additional functionality includes, but is not limited to, simplified information selection, copying, cutting and pasting labeling/tagging, separation of annotations from information, integration of information, sorting through and displaying similar information or distinguishing dissimilar

information, and other functions, either through single-action operation or otherwise, to provide an efficient user experience.

[0008] One general aspect includes a legal data summary system which may include a system server coupled to a communication link, the system server including a system memory configured to store program instructions, and a system processor configured to execute the program instructions, which, when executed, cause the system processor to receive a first data set via the communication link, the first data set including a user identifier, user-identified information, and a unique identifier, the user-identified information including text selected from a legal source document, and the unique identifier being associated with the text, being selected from a unique identifier selection box, and being a category related to legal research, storing the user-identified information and the unique identifier in the system memory at a location associated with the user identifier. The legal data summary system may also include determine that a user has completed identifying user-identified information, reformat the user-identified information based on the unique identifier to generate a summary in response to a user command received via the communication link, the summary including the unique identifier and the user-identified information, and provide the summary via the communication link.

[0009] Additionally or alternatively, implementations may include one or more of the following features. The legal data summary system where reformatting the user-identified information includes causing the text of the user identified information in the summary to be displayed in italics and to be displayed between quotation marks. The legal data summary system further including a website database that includes a plurality of web pages, where the program instructions, when executed further cause the system processor to process hyper-text transfer protocol (http) requests received via the communication link, and provide requested web pages of the plurality of web pages included in the website database via the communication link, the requested web pages being defined in the http requests. The legal data summary system where the first data set further includes custom information associated with the unique identifier, where the program instructions, when executed, further cause the system processor to format the summary to include the custom information listed under the unique identifier, where a location at which the custom information is listed in the summary is prioritized relative to the user-identified information, such that the custom information is listed above the user-identified information in the summary. The legal data summary system where the category is selected from a group including fact, issue, procedure, reasoning, rule, and holding. The legal data summary system where the program instructions, when executed, further cause the system processor to link the user-identified information of the summary to corresponding original information of the legal source document. The legal data summary system where the program instructions, when executed, further cause the system processor to receive, via the communication link, an indication that a single mouse click has been performed at a client device associated with the user identifier, where the summary is generated by the system processor in response to the single mouse click.

[0010] One general aspect includes a method which may include steps for receiving, by a processor of a system server

via a communication link, a first data set including a user identifier, user-identified information, and a unique identifier, the user-identified information including text selected from a legal source document, and the unique identifier being associated with the text, being selected from a unique identifier selection box, and being a category related to legal research, storing, by the processor, the user-identified information and the unique identifier in a system memory of the system server at a location associated with the user identifier, determining, by the processor, that user has completed identifying user-identified information, reformatting, by the processor, the user-identified information based on the unique identifier to generate a summary in response to a user command received via the communication link, the summary including the unique identifier and the user-identified information, and providing, by the processor, the summary via the communication link.

[0011] Additionally or alternatively, implementations may include one or more of the following features. The method where reformatting the user-identified information includes causing the text of the user identified information in the summary to be displayed in italics and to be displayed between quotation marks. The method further including receiving, by the processor via the communication link, hyper-text transfer protocol (http) requests, processing, by the processor, the http requests to identify requested web pages included in a website database of the system server, and providing, by the processor via the communication link, the requested web pages in response to the http requests. The method where the first data set further includes custom information associated with the unique identifier, the method further including formatting the summary to include the custom information listed under the unique identifier, and prioritizing the custom information relative to the user-identified information, such that the custom information is listed above the user-identified information in the summary. The method where the category is selected from a group including fact, issue, procedure, reasoning, rule, and holding. The method further including linking the user-identified information of the summary to corresponding original information of the legal source document. The method further including receiving, via the communication link, an indication that a single mouse click has been performed at a client device associated with the user identifier, where the summary is generated by the processor in response to the single mouse click.

[0012] One general aspect includes a legal data summary system which may include a system server coupled to a communication link, the system server including a system memory configured to store program instructions, and a system processor configured to execute the program instructions, which, when executed, cause the system processor to receive a plurality of data sets via the communication link, the plurality of data sets including a user identifier, a plurality of sets of user-identified information, and a plurality of unique identifiers, a first data set of the plurality of data sets including the user identifier, a first set of user-identified information of the plurality of sets of user-identified information, and a first unique identifier of the plurality of unique identifiers, the first set of user-identified information including first text selected from a first legal source document, and the first unique identifier being associated with the first text, being selected from a unique identifier selection box that includes the plurality of unique identifiers, and being a

category related to legal research, store the plurality of sets of user-identified information and the plurality of unique identifiers in the system memory based on the user identifier. The legal data summary system may also include determine that identification is complete, reformat the plurality of sets of user-identified information based on the plurality of unique identifiers to generate a first summary in response to a user command received via the communication link, the first set of user-identified information being listed under the first unique identifier in the first summary, and provide the first summary via the communication link.

[0013] Additionally or alternatively, implementations may include one or more of the following features. The legal data summary system where a second data set of the plurality of data sets includes the user identifier, a second set of user-identified information of the plurality of sets of user-identified information, and a second unique identifier of the plurality of unique identifiers, the second set of user-identified information including second text selected from a second legal source document, and the second unique identifier being associated with the second text and being selected from a second unique identifier selection box that includes the plurality of unique identifiers. The legal data summary system where the program instructions, when executed, cause the system processor to reformat the second set of user-identified information to generate a second summary in response to the user command, the second set of user-identified information being listed under the second unique identifier in the second summary, aggregating the first summary and the second summary to produce an aggregated summary, and providing the aggregated summary via the communication link. The legal data summary system where the first data set further includes custom information associated with the first unique identifier, and where the program instructions, when executed, cause the system processor to format the first summary to include the custom information listed under the first unique identifier, and prioritize the custom information relative to the first set of user-identified information when generating the first summary such that the custom information is displayed above the first set of user-identified information in the first summary. The legal data summary system where the category is selected from a group including fact, issue procedure, reasoning, rule, and holding. The legal data summary system where the program instructions, when executed, further cause the system processor to link the first set of user-identified information of the first summary to corresponding original information of the first legal source document.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014] Specific examples have been chosen for purposes of illustration and description, and are shown in the accompanying drawings, forming a part of the specification.

[0015] FIG. 1 illustrates a diagram of hardware system components for one example of a data summary system of the present technology.

[0016] FIG. 2 illustrates a diagram of one example of software system components that can be used in the system of FIG. 1.

[0017] FIG. 3 illustrates a flow chart of one embodiment of a data summary method of the present technology.

[0018] FIG. 4 illustrates one embodiment of how a user could associate user-identified information with a unique identifier.

[0019] FIG. 5 illustrates one example of user-identified information before single-action operation.

[0020] FIG. 6 illustrates one example of a summary created by single-action operation.

[0021] FIG. 7 illustrates one embodiment of how a user can aggregate, manipulate, and distinguish information from multiple summaries.

[0022] FIG. 8 illustrates a merged view showing one embodiment of the linking between FIGS. 1 & 2.

[0023] FIG. 9 illustrates a flow chart of one embodiment of the merging summary data with original content.

DETAILED DESCRIPTION

[0024] FIGS. 1 and 2 illustrates an embodiment of the system architecture of one example of a data summary system 100 of the present technology. As shown in FIGS. 1 and 2, the data organizing system includes at least one user device 102 and at least one system server 104, and a communication link 106, through which the user device 102 and system server 104 are operatively connected. The term “operatively connected” as used herein means that the at least one user device 102 and the at least one system server 104 can send and receive data to one another through the communication link 106. The communication link 106 can include a wired connection or a wireless connection, and can be via a local area network (LAN) connection, or via a wide area network (WAN) connection, such as the Internet.

[0025] The hardware components of each of the user device 102 and the system server 104 include at least one processor, and at least one non-transient computer readable storage medium, such as a memory. As shown in FIG. 1, the user device 102 includes at least one user device processor 108 and at least one user device memory 110, and the system server includes at least one system processor 112 and at least one system memory 114. The user device 102 can also include an input device 134, that the user can use to input commands to the user device 102, such as selection of user-identified information. The input device 134 can include one or more of a keyboard, a mouse, an electronic pointing device, a touch screen, or any other suitable device.

[0026] Examples of system architecture components of each of the user device 102 and the system server 104 are illustrated in FIG. 2.

[0027] As shown, the user device 102 includes a browser 116, which is assigned a unique user identifier 118. The user identifier 118 is an identifier that identifies the user device 102 to the server system 104. The user identifier may be stored in a file referred to as a “cookie.” In one example, the server system 104 can assign and send the user identifier 118 to the user device 102 once, when the user device 102 first interacts with the server system 104. Once the user device 102 receives the user identifier 118, the user device stores the user identifier 118, such as in the user device memory 110, and provides it with every message sent to the server system 104. The server system 104, which can have communication links with any number of user devices, can thus identify the user device 102 as being the source of the message.

[0028] As also shown in FIG. 2, the system server 104 can include a server engine 120, user profile database 122 including a stored user profile 132 associated with each user

identifier, a website database 124 including web pages, a unique identifier database 126, a user-identified information database 128, and a custom information database 130. Each database of the server system 104 can be stored in the system memory 114. The server engine 120 can include at least the system processor 112.

[0029] In the illustrated example, the server system 104 associates, and operatively links, each user identifier 118 with a user profile 132 stored in the user profile database 122, unique identifiers, user-identified information identified by the user device 102, and custom information received from the user device 102.

[0030] In some examples, the server engine 120 receives HTTP requests from a user device 102 to access web pages identified by URLs, and provides the requested web pages from the website database 124 to the user device 102. However, one of ordinary skill in the art would appreciate that the summary techniques described herein could be used in various environments other than the Internet. For example, the summaries could also be in an electronic mail environment. Also, a server system 104 or user device 102 may include any combination of hardware or software suitable for performing the methods described herein. For example, a user device 102 may comprise any combination of hardware or software that can interact with the server system 104. Either the user device 102 or the server system 104 can include one or more handheld devices, personal computers or television- or tablet-based systems, or various consumer computer products through which information may be generated.

[0031] FIG. 3 illustrates one example of a data summary method of the present technology. At step 202, a user uses the user device 102 to identify user-identified information in a source document. The source document can be any electronic document, such as digital text, media, web pages, or other content to be included in the summary. The user can identify the user-identified information by selecting a subset of text in a source document, such as by highlighting the text and clicking a mouse to select the text. Once the user has selected the user-identified information, the user may be prompted to associate such user-identified information with a unique identifier. At step 204, the user associates a unique identifier with the user-identified information. The user can select the unique identifier associated with the user-identified information from a plurality of predefined unique identifiers, such as a listing of unique identifiers presented in a drop-down menu. Alternatively, the unique identifiers can be customized by the user. At step 206, the summary system allows the user to input or modify additional custom information under the unique identifier. The user device 102 then sends a first data set to the server system 104 via the communication link 106. The first data set includes the user identifier 118, the user-identified information, and the unique identifier associated with the user-identified information. The server system 104 receives the first dataset, recognizes the user identifier, and stores the user-identified information in the user’s profile. If the user did not input or modify additional custom information, the server system 104 saves the user-identified information under the unique identifier chosen by the user. However, if the user did modify or input custom information, the server system 104 saves the new or modified custom information, along with the user-identified information.

[0032] At step 210, the summary system determines whether the user is done creating user-identified information. If the user is not done, the method repeats steps 202 through 208, with the user selecting user-identified information from any one or more source documents, which can be the same or different from the original source document and each other. If the user has completed identifying user-identified information, the method progresses to step 212.

[0033] At step 212, the server system enables the generation of a data summary. In some examples, the user can cause the server system 104, or the user device 102, to generate the summary with a single action. For example, the user can enter a single command, such as a mouse click on a button in a selection box displayed by the user device 102, which the user device 102 sends to the system server 104, in order to cause the server system 104 to create a summary of the user-identified information and any custom information. In an example where the user device is being used offline, the user can enter the single command in order to cause the user device 102 to create a summary of the user-identified information and any custom information. If the user enters the command, summary system 100 proceeds to step 214, where the summary system 100, either through the server system 104 or the user device 102, reformats the user-identified information and the associated unique identifier, and reformats the information in order to generate a summary. Generating a summary includes prioritizing and formatting the information received from the user. For example, if custom information is present under a unique identifier, the summary system 100 can format the custom information in the summary, including prioritizing the custom information relative to, such as above or below, the user-identified information for any given unique identifier at step 216. Additionally, at step 218, the system formats the user-identified information into the summary. In generating the summary, the summary system 100 can link the summarized information to the original information and vice versa at step 220. Finally, at step 222, the summary system 100 can provide the generated summary to the user. If the server system 104 generated the summary, the summary can be provided to the user by sending the generated summary to the user device 102 via the communication link 106.

[0034] In examples where the user can cause the summary to be generated with a single action, once the user initiates the single-action (via a mouse click, keystroke, or otherwise), the system generates a formatted summary. The summary system 100 can also allow the user to modify the generated summary. In general, the user need only enter the user-identified information, and any custom information, under each unique identifier, and then execute the single action in order to generate a summary. Such systems can allow the user to efficiently and reliably generate a summary of digital text, media, web pages, or other content.

[0035] In alternative examples, a multi-action operation can be used, which can include defining a custom summary format through multiple mouse clicks and/or keystrokes prior to generating the summary.

[0036] FIG. 4 illustrates one embodiment of how the user can assign a unique identifier 300 to a selection of user-identified information 302. In this example, once the user-identified information 302 is selected by the user, a unique identifier selection box 304 is displayed by the user device 102, providing the user with various predefined or common unique identifiers 300. These may include, as examples only,

Fact, Issue, Procedure, Reasoning, Rule, Holding, Notes. The user can then select the unique identifier 300 to be associated with the user-identified information 302 through a single action (via a mouse click or otherwise) or multiple actions. Once a particular unique identifier 300 is selected by the user (via a mouse click or otherwise), the summary system 100 associates that unique identifier 300 with the particular selection of user-identified information 302, and saves such association for possible further processing. The summary system 100 can allow the user to modify and customize the unique identifiers 300 and the unique identifier selection box 304.

[0037] FIG. 5 illustrates one example of content where the user has identified user-identified information 400, and associated the user-identified information 400 with unique identifiers 402 and the system has saved such information under the user's profile. In some instances, the user has input custom information 404 under the unique identifiers. The summary system 100 stores the user-identified information 400, the unique identifiers 402, and the custom information 404 in the system server 104 in association with the user identifier 118.

[0038] FIG. 6 illustrates one example of a summary 500 generated by the system server 104 of the summary system 100. The summary 500 includes a listing of each unique identifier 502. Under each unique identifier 502, any custom information 504 and user-identified information 506 associated with the unique identifier 502 are listed. The custom information 504 is illustrated as being prioritized over the user-identified information 506 for each unique identifier 502. The system can, however, be customized to prioritize custom and user-identified information in any order.

[0039] The summary system 100 can also aggregate multiple summaries, and allow the user to view user-identified information and/or custom information by unique identifier across multiple summaries. FIG. 7 illustrates one example of how multiple summaries 600 can be aggregated and how unique identifiers can be used to filter and/or distinguish information from multiple summaries. In this example, a unique identifier activation box 602 is used to allow the user to select desired unique identifiers 604, by checking the indicated selection box(es). The summary system 100 generates a display of all of the user-identified information 606 and custom information 608 associated with the selected unique identifiers 602. An information type selection box 610 can be used to display user-identified information 606, custom information 608, or both. In this example, the information type selection box 610 differs from the unique identifier activation box 602 in that the information type selection box 610 allows the user to choose to display either user-identified information 606 or custom information 608, or both, across multiple summaries, while the unique identifier activation box 602 allows the user to choose to display information (either user-identified information 606 or custom information 608, or both) associated with certain unique identifiers (and exclude others) across multiple summaries 600. Thus, a compilation of summaries can be created by the summary system 100.

[0040] FIG. 8 illustrates an integrated view of source document 702 and a summary 704 in a split-screen display 700. The source document 702 includes the original digital text, media, web pages, or other content. As shown, the summary system can link the summarized information to the

original information in the source document, and vice versa. This example also includes a notes area **706**, for the user to input additional information.

[0041] FIG. 9 is a flow diagram of a process for generating the split screen view of FIG. 8. At step **802**, the server system **104** generates the summary, which can be in accordance with the process illustrated in FIG. 3. At step **804**, the summary system enables the user to select a split view, such as by displaying a selection box to the user via the user device and sending a user selection to the system server **104**. At step **806**, the split screen is generated by the server system **104**, and sent to the user device **102** to be displayed to the user. In generating the split screen, the summary system **100** can maintain the links between the source document and the user-identified information. Additionally, at step **808**, the summary system can also generate the notes area **706**, which can be sent to the user device **102** to be displayed to the user, where the user can input additional information.

[0042] From the foregoing, it will be appreciated that although specific examples have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit or scope of this disclosure. It is therefore intended that the foregoing detailed description be regarded as illustrative rather than limiting, and that it be understood that it is the following claims, including all equivalents, that are intended to particularly point out and distinctly claim the claimed subject matter.

What is claimed is:

1. A legal data summary system comprising:
 - a system server coupled to a communication link, the system server comprising:
 - a system memory configured to store program instructions; and
 - a system processor configured to execute the program instructions, which, when executed, cause the system processor to:
 - receive a first data set via the communication link, the first data set including a user identifier, user-identified information, and a unique identifier, the user-identified information including text selected from a legal source document, and the unique identifier being associated with the text, being selected from a unique identifier selection box, and being a category related to legal research;
 - storing the user-identified information and the unique identifier in the system memory at a location associated with the user identifier;
 - determine that a user has completed identifying user-identified information;
 - reformat the user-identified information based on the unique identifier to generate a summary in response to a user command received via the communication link, the summary including the unique identifier and the user-identified information; and
 - provide the summary via the communication link.
2. The legal data summary system of claim 1, wherein reformatting the user-identified information comprises:
 - causing the text of the user identified information in the summary to be displayed in italics and to be displayed between quotation marks.
3. The legal data summary system of claim 1, further comprising:

a website database that includes a plurality of web pages, wherein the program instructions, when executed further cause the system processor to:

- process hyper-text transfer protocol (HTTP) requests received via the communication link; and
 - provide requested web pages of the plurality of web pages included in the website database via the communication link, the requested web pages being defined in the HTTP requests.
4. The legal data summary system of claim 1, wherein the first data set further includes custom information associated with the unique identifier, wherein the program instructions, when executed, further cause the system processor to:
 - format the summary to include the custom information listed under the unique identifier, wherein a location at which the custom information is listed in the summary is prioritized relative to the user-identified information, such that the custom information is listed above the user-identified information in the summary.
 5. The legal data summary system of claim 1, wherein the category is selected from a group consisting of: fact, issue, procedure, reasoning, rule, and holding.
 6. The legal data summary system of claim 1, wherein the program instructions, when executed, further cause the system processor to:
 - link the user-identified information of the summary to corresponding original information of the legal source document.
 7. The legal data summary system of claim 1, wherein the program instructions, when executed, further cause the system processor to:
 - receive, via the communication link, an indication that a single mouse click has been performed at a client device associated with the user identifier, wherein the summary is generated by the system processor in response to the single mouse click.
 8. A method comprising:
 - receiving, by a processor of a system server via a communication link, a first data set including a user identifier, user-identified information, and a unique identifier, the user-identified information including text selected from a legal source document, and the unique identifier being associated with the text, being selected from a unique identifier selection box, and being a category related to legal research;
 - storing, by the processor, the user-identified information and the unique identifier in a system memory of the system server at a location associated with the user identifier;
 - determining, by the processor, that user has completed identifying user-identified information;
 - reformatting, by the processor, the user-identified information based on the unique identifier to generate a summary in response to a user command received via the communication link, the summary including the unique identifier and the user-identified information; and
 - providing, by the processor, the summary via the communication link.
 9. The method of claim 8, wherein reformatting the user-identified information comprises:
 - causing the text of the user identified information in the summary to be displayed in italics and to be displayed between quotation marks.

10. The method of claim 8, further comprising:
 receiving, by the processor via the communication link, hyper-text transfer protocol (HTTP) requests;
 processing, by the processor, the HTTP requests to identify requested web pages included in a website database of the system server; and
 providing, by the processor via the communication link, the requested web pages in response to the HTTP requests.
11. The method of claim 8, wherein the first data set further includes custom information associated with the unique identifier, the method further comprising:
 formatting the summary to include the custom information listed under the unique identifier; and
 prioritizing the custom information relative to the user-identified information, such that the custom information is listed above the user-identified information in the summary.
12. The method of claim 8, wherein the category is selected from a group consisting of: fact, issue, procedure, reasoning, rule, and holding.
13. The method of claim 8, further comprising:
 linking the user-identified information of the summary to corresponding original information of the legal source document.
14. The method of claim 8, further comprising:
 receiving, via the communication link, an indication that a single mouse click has been performed at a client device associated with the user identifier, wherein the summary is generated by the processor in response to the single mouse click.
15. A legal data summary system comprising:
 a system server coupled to a communication link, the system server comprising:
 a system memory configured to store program instructions; and
 a system processor configured to execute the program instructions, which, when executed, cause the system processor to:
 receive a plurality of data sets via the communication link, the plurality of data sets including a user identifier, a plurality of sets of user-identified information, and a plurality of unique identifiers, a first data set of the plurality of data sets including the user identifier, a first set of user-identified information of the plurality of sets of user-identified information, and a first unique identifier of the plurality of unique identifiers, the first set of user-identified information including first text selected from a first legal source document, and the first unique identifier being associated with the first text, being selected from a unique identifier selection box that includes the plurality of unique identifiers, and being a category related to legal research;

store the plurality of sets of user-identified information and the plurality of unique identifiers in the system memory based on the user identifier;
 determine that identification is complete;
 reformat the plurality of sets of user-identified information based on the plurality of unique identifiers to generate a first summary in response to a user command received via the communication link, the first set of user-identified information being listed under the first unique identifier in the first summary; and
 provide the first summary via the communication link.

16. The legal data summary system of claim 15, wherein a second data set of the plurality of data sets includes the user identifier, a second set of user-identified information of the plurality of sets of user-identified information, and a second unique identifier of the plurality of unique identifiers, the second set of user-identified information including second text selected from a second legal source document, and the second unique identifier being associated with the second text and being selected from a second unique identifier selection box that includes the plurality of unique identifiers.

17. The legal data summary system of claim 16, wherein the program instructions, when executed, cause the system processor to:

reformat the second set of user-identified information to generate a second summary in response to the user command, the second set of user-identified information being listed under the second unique identifier in the second summary;
 aggregating the first summary and the second summary to produce an aggregated summary; and
 providing the aggregated summary via the communication link.

18. The legal data summary system of claim 15, wherein the first data set further includes custom information associated with the first unique identifier, and wherein the program instructions, when executed, cause the system processor to:

format the first summary to include the custom information listed under the first unique identifier; and
 prioritize the custom information relative to the first set of user-identified information when generating the first summary, such that the custom information is displayed above the first set of user-identified information in the first summary.

19. The legal data summary system of claim 15, wherein the category is selected from a group consisting of: fact, issue procedure, reasoning, rule, and holding.

20. The legal data summary system of claim 15, wherein the program instructions, when executed, further cause the system processor to:

link the first set of user-identified information of the first summary to corresponding original information of the first legal source document.

* * * * *