



(12) 发明专利

(10) 授权公告号 CN 101084496 B

(45) 授权公告日 2012. 11. 21

(21) 申请号 200580021670. 0

(51) Int. Cl.

(22) 申请日 2005. 05. 03

G06F 17/00 (2006. 01)

(30) 优先权数据

60/567, 997 2004. 05. 04 US

(56) 对比文件

(85) PCT申请进入国家阶段日

2006. 12. 28

US 2003/0149660 A1, 2003. 08. 07, 全文 .

US 2004/0083422 A1, 2004. 04. 29, 全文 .

US 2004/0064438 A1, 2004. 04. 01, 全文 .

US 6499026 B1, 2002. 12. 24, 说明书附图

(86) PCT申请的申请数据

11-19、176, 说明书摘要, 说明书 26-28 栏 .

PCT/US2005/015346 2005. 05. 03

审查员 白莉

(87) PCT申请的公布数据

W02005/107405 EN 2005. 11. 17

(73) 专利权人 波士顿咨询集团公司

地址 美国马萨诸塞州

(72) 发明人 拉尔夫·W·埃卡特

小罗伯特·G·沃尔夫

亚历山大·夏皮罗

凯文·G·里韦特

马克·F·布莱克希尔

(74) 专利代理机构 中国国际贸易促进委员会专

利商标事务所 11038

代理人 董莘

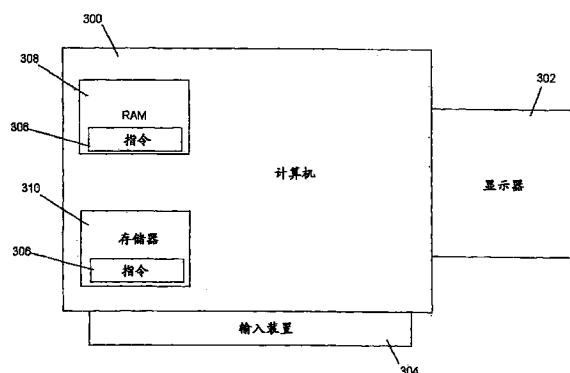
权利要求书 2 页 说明书 39 页 附图 33 页

(54) 发明名称

用于选择、分析以及将相关数据库记录可视化为网络的方法

(57) 摘要

一种用于选择和转换相关文档的数据库记录或集合为网络数据并在网络可视化系统中表示所述数据的方法和设备，所述网络可视化系统使得用户能够在其中选择，并通过选择将被表示为网络的节点和链接的数据的一个或多个属性在各种网络显示之间移动。



1. 一种用于提供两个或更多个数据库记录的网络图形表示的方法，包括：

根据一个或多个描述性标准选择所述两个或更多个数据库记录，其中所述两个或更多个数据库记录中的每一个是一个公共记录类的成员；

识别所述数据库记录的两个或更多个公共属性，使第一组网络节点与来自所述数据库记录的第一个所述公共属性的实例相关联，并使第二组网络节点与来自所述数据库记录的第二个所述公共属性的实例相关联；

利用指定所述网络节点之间的关联性的网络链接，将所述第一组网络节点的一个或多个成员连接到所述第二组网络节点的一个或多个成员，以便形成第一网络图形表示；以及

在修正所述一个或多个描述性标准的同时，迭代地执行识别和连接步骤，以便改变所选择的两个或更多个数据库记录。

2. 根据权利要求 1 的方法，其中所述公共记录类包括专利记录。
3. 根据权利要求 2 的方法，其中所述专利记录是从 LexisNexis 数据库提取的。
4. 根据权利要求 2 的方法，其中所述专利记录是从 Thomson 数据库提取的。
5. 根据权利要求 2 的方法，其中所述专利记录是从 USPTO 数据库提取的。
6. 根据权利要求 2 的方法，其中所述专利记录是从 EPO 数据库提取的。
7. 根据权利要求 2 的方法，其中所述专利记录是从 Derwent 数据库提取的。
8. 根据权利要求 1 的方法，其中所述公共记录类包括学术期刊文章。
9. 根据权利要求 8 的方法，其中所述学术期刊文章是从 PubMed 数据库提取的。
10. 根据权利要求 2 的方法，其中所述一个或多个描述性标准选自由以下内容构成的组：(i) 每个所述专利记录的主体字段内的一个或多个关键字；(ii) 每个所述专利记录的标题字段内的一个或多个关键字；(iii) 每个所述专利记录的发明人字段中的一个或多个发明人；(iv) 每个所述专利记录的受让人字段中的一个或多个受让人；(v) 摘要字段内的一个或多个关键字；以及它们的组合。
 11. 根据权利要求 2 的方法，其中所述公共属性包括发明人。
 12. 根据权利要求 2 的方法，其中所述公共属性包括受让人。
 13. 根据权利要求 2 的方法，其中所述公共属性包括申请日。
 14. 根据权利要求 2 的方法，其中所述公共属性包括颁证日。
 15. 根据权利要求 2 的方法，其中所述公共属性包括 IPC 码。
 16. 根据权利要求 2 的方法，其中所述公共属性包括 USPC 码。
 17. 根据权利要求 2 的方法，其中所述公共属性包括搜索字段。
 18. 根据权利要求 1 的方法，其中所述网络链接包括描述在所连接的节点之间出现的公共实例的数量的特性。
 19. 根据权利要求 18 的方法，其中所述特性包括链接厚度。
 20. 根据权利要求 18 的方法，其中所述特性包括链接颜色。
 21. 根据权利要求 18 的方法，其中所述特性包括链接结构。
 22. 根据权利要求 1 的方法，其中至少一组网络节点是一个元节点组。
 23. 根据权利要求 22 的方法，其中所述元节点组描述两个或更多个数据库记录的特性。
 24. 根据权利要求 23 的方法，其中所述一个或多个描述性标准包括日期范围。

25. 根据权利要求 2 的方法,还包括从除专利记录的公共记录类以外的记录类选择附加数据库记录,并且使网络节点、网络链接或二者与来自所述附加数据库记录的一个或多个属性的实例相关联。

26. 根据权利要求 25 的方法,其中除专利记录的公共记录类以外的所述记录类描述与所述专利记录相关联的许可历史。

27. 根据权利要求 25 的方法,其中除专利记录的公共记录类以外的所述记录类描述与所述专利记录相关联的诉讼历史。

28. 根据权利要求 25 的方法,其中除专利记录的公共记录类以外的所述记录类描述与所述专利记录相关联的维持费历史。

29. 根据权利要求 8 的方法,还包括从除学术期刊文章的公共记录类以外的记录类选择附加数据库记录,并且使网络节点与来自所述附加数据库记录的一个或多个属性的实例相关联。

30. 根据权利要求 29 的方法,其中除学术期刊文章的公共记录类以外的所述记录类描述与所述学术期刊文章相关联的医生联系数据。

31. 根据权利要求 29 的方法,其中除学术期刊文章的公共记录类以外的所述记录类描述与所述学术期刊文章相关联的脚本数据。

32. 根据权利要求 29 的方法,其中除学术期刊文章的公共记录类以外的所述记录类描述与所述学术期刊文章相关联的参考数据。

33. 根据权利要求 1 的方法,还包括根据用户提供的需求识别所述记录类的一个或多个属性。

用于选择、分析以及将相关数据库记录可视化为网络的方法

[0001] 相关申请的交叉引用

[0002] 本发明要求 2004 年 5 月 4 日申请的美国临时专利申请序列号 No. 60/567,997 的优先权，其全部内容在此结合作为参考。

技术领域

[0003] 本发明一般涉及数据挖掘和分析领域。更为特别地，本发明涉及以网络图形表示表示相关数据库记录的方法和系统。

背景技术

[0004] “信息时代”和“知识经济”不过是通常用于描述我们的时代的数字信息爆炸特征的两个术语。无论你如何称呼它，毫无疑问的是被创造出的信息量正以史无前例的速度增长。已经有许多用于量化新知识发展的速度的尝试并已产生有关其爆炸性增长的各种各样的估计。各种各样的来源描述并试图量化这种信息爆炸。经常引用的几个这种统计的例子有：

[0005] ●通常每 5 ~ 10 年总的人类的知识增加一倍；

[0006] ●通常每 3 ~ 5 年科学知识增加一倍；

[0007] ●通常每 2 ~ 8 年医学知识增加一倍；

[0008] ●在过去的 7 年内公布的美国专利的数量大约增加了一倍；

[0009] ●每天大约有 1,500,000 个网页增加到环球网；

[0010] ●1999 年内数字化存储的原始内容的世界常量需要大约 635,000 ~ 2,100,000 兆千兆字节以存储。

[0011] 不管这些估计的可靠性如何，它们都指出了新的信息的不可否认的爆炸性增长。计算机技术使得易于创建和存储新的信息。用于存储 这些信息的数据库的数量和规模正在爆炸性地增长。

[0012] 尽管可用信息的快速增长，人类消化吸收和理解信息的思维能力并没有显著提高。可用信息的迅速增长和我们不能消化吸收它导致信息超负荷。信息的大量存储使得越来越难于找到正确的信息而且甚至于更加难于理解巨大数量的可用的新知识。

[0013] 知识经济内的工作者工作在充满信息但却不能提取出理解的环境中。这些工作者经常需要寻找并理解有关特定主题或感兴趣的领域的信息以便能够提高他们的行为和 / 或决策。然而，尽管信息的可用性能够传达信息给他们并提高他们的决策，却没有寻找或消化吸收它的实用方式。

[0014] 为数众多的公司已投入巨资以帮助信息工作者在他们正搜索的大量数据“大海”中查找信息“针”。用于信息检索的主要范例可称为“搜索和筛选”。“搜索和筛选”方法总是以返回大量匹配的搜索结果的逻辑搜索开始。搜索者然后在结果之内筛选以找到他们寻找的信息。因特网用户和其它大型数据库的用户可能对这种方法非常熟悉。

- [0015] 信息检索领域的大多数投资集中于改进“搜索和筛选”处理。改进的例子包括：
- [0016] ● 查询精炼 - 查询精炼试图确定搜索者的查询背后的意图并精炼查询以便捕获更多与搜索有关的文档或者从结果集中排除更多的无关文档。查询精炼的一个例子是“同义词扩展”，其中查询项被扩充以包括搜索项的同义词以期捕获更多的相关文档。
- [0017] ● 结果排序 - 改进“搜索和筛选”方法的第二种方法是结果排序。结果排序试图基于搜索结果与搜索者意图的相关性排列搜索结果。相关性排序已经以各种方法进行估算，包括：搜索项的使用频率，搜索项在文档内的位置，以及察觉到的结果集内的文档的“重要性 / 有用性”。结果分类的最佳实例也许就是 Google 的基于链接到搜索结果页面的其它网页的数量的页面排序度量。
- [0018] ● 结果过滤 - 改进“搜索和筛选”方法的最后一个方法的例子是结果过滤。结果过滤试图基于某种分类方案分类结果集内的文档。其 希望是这将允许搜索者缩小他 / 她的“筛选”到最为接近与感兴趣的领域有关的结果集的子集。结果过滤的例子包括：Northern Light 的“results folders”（参见，例如图 1），其基于文档分类的固定分类法。Vivisimo 的文档群集工具，其基于文档的语义内容将文档分类为分层树形结构（参见，例如图 2），以及 Grokker，其类似于 Vivisimo 将文档分类为动态分层结构，并还利用其的“冒泡显示”提供每个分类的相对大小的可视显示（参见，例如图 3）。
- [0019] 所有这些方法都是对“搜索和筛选”方法有用的改进，然而，他们都假定一种特定类型的信息需求，即搜索者正在寻找特定的信息块（PIECE），并且被查找的信息能够在结果集中的文档之内被找到。这种类型的信息检索目标在于发现问题的答案，例如：
- [0020] ● 谁杀了 Bobby Kennedy ?
- [0021] ● 世界上第二高的山是什么？
- [0022] ● Palo Alto, CA 明天的天气预报是什么？
- [0023] ● IBM 的当前股票价格？
- [0024] 虽然在此描述的实施例代表了对“搜索和筛选”方法的进一步改进，它们的主要贡献旨在满足不同类型的信息需求。这些实施例的主要目的是通过提供一种用于消化吸收集合内的文档之中的信息模式帮助信息用户理解搜索结果、或者巨大的文档集合。这种类型的信息在此称为“元数据”，因为其代表了比包含在数据库或搜索结果内的任何特定文档或记录的信息更高级别的信息。这种类型的信息检索旨在回答诸如以下的问题：
- [0025] ● 有多少文档与我所感兴趣的领域有关，并且这个数字增长得有多快？
- [0026] ● 谁是有关这个主题的信息的主要作者？
- [0027] ● 是哪些公司正在制造有关这个主题的信息？
- [0028] ● 工作在这个领域的公司 / 作者之间的关系如何？
- [0029] 所描述的实施例利用先进的可视化技术揭示与文档集或搜索结果有关的元数据。为了理解本发明的新颖贡献，回顾一下本领域内的 其它系统和技术是有益的，特别是在两个研究领域：1) 表示元数据的现有方法，2) 用于理解大数据集的可视化方法学。
- [0030] 表示元数据的现有方法
- [0031] 之前的分析和表示与大数据集有关的元数据的努力可划分为多种类别。为区分本发明的目的，以下提供了每一种类别的简短描述及现有技术发展水平的实例。
- [0032] 统计分析

[0033] 一种最为简单且使用最广的分析文档集的方法是统计分析。统计分析可以与根据日期、作者 / 发明人、作者 / 发明人从属关系、国家、分类、或其它属性计算文档的数量一样的简单。其还可包括与被检查的特定类型数据有关的统计计算。例如，在专利数据领域，有时计算类似于引用的数量、引用 / 专利 / 年度、自提交申请到授权的时间、最近的引用的年份、最近的学术引用的年份的统计，以及其它统计。这些统计方法被广泛采用，并在某些情况下被自动化于诸如由专利领域中的 Delphion、Micropatent 和 CHI Research 和其它领域中的许多其它组织提供的商业应用中。

[0034] 统计分析可以提供对等待评估的文档集的一些有用的理解，但是显然受限于能够获得的理解度。最为公知的这种类型工具提供原文报告或简单的条形图，原文报告或简单条形图显示每个属性值（例如，就公司 A、公司 B、公司 C 来说有多少文档）的文档数量或者与整个文档集（例如，从提交申请到授权的平均时间）有关的统计。它们没有提供有关各种文档相互之间是如何关联的信息，并且它们没有提供用于以一种方式与元数据交互的方法，该方式允许用户探索是文档的各种什么属性揭示了有关整个文档集。本发明的一个或多个实施例的目的是为用户提供一种理解文档组之间的关系的手段，以及提供一种用于深度探索与文档集或搜索结果有关的元数据的手段。

[0035] 群集

[0036] 另一种用于揭示有关大文档集的元数据的方法是群集。已经开发了多种分组文档为群集的工具。这些工具中的一些基于各类别的固定分类法将文档分隔为群集，而其它的利用文档内的语法信息将他们群集为动态类别组。固定分类法群集工具的两个例子是 Northern Light 搜索引擎和 The Brain (<http://www.thebrain.com/>) 的 web 搜索工具。固定分类法群集方法以两种方式的其中之一实现。首先，各类别可以基于文档的显式属性。例如，基于因特网搜索结果的域扩展，如“.com”、“.net”、“.edu”，或者它们的国家域，如“.sp”、“.ge”、“.jp”等将它们划分为各个类别。其次，各个类别可以基于数据仓库中的文档先前已经被指定到其内的分类法。这通常可通过人工回顾文档或那些文档落入的域并将它们指定到固定分类法内的一个或多个类别来实现。

[0037] 第二种群集文档或搜索结果的方法基于动态分类法的创建。这些群集技术使用文档内的语法数据，然后将文档集群集到较小的组并基于它们公共的单词或短语“命名”那些分组。该群集方法实质上创建了能够提供对集合中的文档的性质的理解的自动化分类方案。这种技术已经被应用到广泛种类的文档类型并且有许多执行这种功能的商业软件应用是可用的。专利领域内使用群集技术的例子包括 Vivisimo 和 Themescape 工具 (<http://www.micropat.com/static/advanced.htm>)，它们被结合到 Micropatent 的 Aureka (<http://www.micropat.com/static/index.htm>) 工具集，并且在 Delphion 的工具集中可得到 Text Clustering 工具 (<http://www.delphion.com/products/research/products-cluster>)。Vivisimo 的工具可被配置以运行在任何文本文档集之上，如同 Inxight 开发的语义分析工具 (<http://www.inxight.com/products/smartdiscovery/>) 能够做到的那样。

[0038] 利用这些群集工具，可以表示有关文档集或搜索结果的基本元数据。由上面提及的工具采用的方法可以自动显示落入每个类别的集合或搜索结果中的文档的数量，使得可能更迅速地在结果之内“筛选”以找到正被查找的信息块。它们还提供有关文档集或搜索

结果的内容某些有价值的信息。

[0039] 最著名的群集工具的价值在两个重要方面受到限制。首先，所提供的有关文档集的内容的元数据实际上仅是其被群集到其内的分类法。这是固定和动态分类法群集技术的固有限制。

[0040] 由于诸如因素，固定分类法在它们的有用性方面受限：

[0041] ●分类法是基于其创建者的优先级，而不是搜索者的。分类法的创建需要作出有关信息的哪些属性是最为重要的选择。例如，在鸟类分类法中的第一个分支可能以多种可能的方式建立；迁徙的与不迁徙的、水鸟与陆鸟等等。通常，分类学家的优先级与信息用户的需求并不一致，由此使所提供的群集元数据的价值受限。

[0042] ●固定分类法无法轻易地调整为数据库演化的内容。一旦分类法已经建立并且用户已经开始使用，则改变将变得严格和困难。如同内容的演化，不可避免地存在增加新的类别、细分类别、以及重新组合类别的需求。这使得随着时间的过去难于比较结果。作为例子，考虑由 WIPO 创建的被称为国际专利分类系统 (IPC) 的分类法技术。IPC 现在是其的第七个版本。在每个版本中，类别被增加、移动、细分和删除。然而，上百万的专利文档在它们被授权时存在的原始分类方案之下的所分类的修订保留之前提出申请。当基于固定分类法时这使得群集元数据的表示存在问题。

[0043] ●另一个与固定分类法有关的问题是数据集中的文档典型地没有落入单一的类别。这引起分类问题，该问题典型地已经通过将文档分配到分类法内的多个类别来解决。这种多重分配引出当许多文档落入多个类别时如何显示群集的结果的挑战。它们的典型的解决方案是仅在单个（主要的）分类中计数每个文档，或者对每种分类类别计数一次，而文档被多次计数。这两种解决方案都有问题。第一种方案忽略了与次级分类有关的重要信息，第二种表示了每个文档的多个实例。

[0044] ●固定分类法的另一主要限制是难于分配文档到各个类别。典型地，这是一个或者由文档的作者或者由负责分类的经专门训练的个人或多人完成的手工处理。再次，这两种选择都有问题。由作者分类遭受一致性缺乏的问题，而当必须分类巨大数量的文档时集中式分类非常耗时。

[0045] 已经创建了动态分类法以便克服固定分类法的某些限制。然而，它们有自身的限制，这些限制减小了在提供有关大文档集的元数据时的有用性。与动态分类法相关的挑战描述如下：

[0046] ●发明人所知的所有动态分类法系统都是基于语义数据的。简单地说，文档的分类是基于文档内包含的单词的相似性的。这种方法的问题是当其涉及表示概念时，所有的语言都是极端不精确的。基于语义相似性的任何文档分类都将遭受同义词（多个单词表示同一个意义）和多义词（单词有多种意义）的问题。尽管语法群集无疑是有价值的，但发明人的经验显示所创建的群集使人联想到内容，而与精确相去甚远。

[0047] ●与语义群集有关的第二个语言学的问题是多语言。当数据集中包含不同语言的文档时语义群集工具将完全失效。由于全球化的趋势在持续，这个问题的重要性也在持续增长。已经作出了一些努力使用多语言的主题词表以允许多语言文档集的语言学比较，但这种研究仍处于其初期阶段。

[0048] ●动态分类法的最后一个限制是缺乏从一个文档集或搜索结果到另一文档集或

搜索结果之间的可比较性。因为分类法是针对文档集而特定创建的，没有针对不同的文档集或不同搜索结果而创建的两个分类法可以比较。

[0049] ●动态分类法还遭受上述的多重分类问题。

[0050] 群集技术的第二个限制是任何的分类法仅描述了涉及单个属性的文档集或搜索结果。多数分类法打算描述它们所分类的文档的主题或题目。虽然这种信息是有用的，但没有任何被发明人所知的允许用户同时利用群集信息以及描述文档集或搜索结果的元数据的各种其它可来源的系统。本发明的一个或多个实施例的目的是为用户提供一种迭代或同时利用固定和动态分类法中包含的信息以及广泛种类的其它元数据来源以便提供有关满足用户的特定信息需求的文档集或搜索结果的深层次的理解。

[0051] 用于理解大数据集的可视化方法论

[0052] 获得对有关大文档集或搜索结果的元数据的理解的最为先进的方法是可视化技术。由于计算机处理器变得功能足够强大以执行显示复杂数据关系所需的成百万次的计算，数据可视化领域在过去的几年内得到快速的发展。有许多与考虑关于本发明相关的数据可视化工具。它们可以被划分为下面将描述的几个类别。还将为每个类别提供相关实例。

[0053] 分层显示 - 一种已被采用的可视化方法是分层显示。在其最简单的形式中，文档或搜索结果是以与用于显示分类的数据的众所周知的隐喻 (metaphor) 的目录结构类似的树形结构的形式表示的。所设计的分层显示的一个例子用于揭示包括如上所述的 Vivisimo 的群集工具的元数据。由于显示和理解巨大的分层结构的困难，已经开发出了几种可选的方法以显示这些分层结构。一个例子是鱼眼透镜 (fisheye lens)，其用于显示 Micropatent 的 Aureka 工具集内的专利引用的巨大的分层结构。鱼眼显示允许用户放大分层结构的一部分同时仍然了解在它们在整个分层结构中的位置。

[0054] 分层显示的另一个复杂例子是在美国专利 6,879,332B2 中描述的由 Grokis 公司开发的 Grokker 工具。很象 Vivisimo 工具，Grokker 工具基于语义算法以分层结构群集文档。不同于 Vivisimo，Grokker 工具以程式化的 marimekko 图呈现信息给用户。Grokker 可视化以二维空间表示文件夹，基于群集内的文档数量将估计每个文档群集的大小。屏幕上的空间表示全部的搜索结果。在这个空间范围内，基于在那些文档内找到的公共单词显示（以圆或正方形表示）和标记文档群集。在每个群集之内，是进一步的“子群集”，再次可视表示并利用关键字标记。该分层结构由高及低最到文档自身最后在分层结构的最低层次被发现。

[0055] 分层数据可视化的这些主要的实例的每一个都基于包含在文档内的潜在信息并同样地遭受在描述固定和动态分类法的章节中的如上所述的语义分析的限制。

[0056] 空间可视化 - 用于揭示大文档集内的元数据的第二种类型的可视化是空间可视化。空间可视化使用地图隐喻以在二或三维空间中排列文档记录。尽管各种空间可视化工具稍有不同，对发明人已知的那些是都遵循类似的用于创建地图的方法学。这种方法需要四个步骤：1) 计算每个文档的语义向量 - 针对数据集中的每个文档，计算一个向量以表示文档的语义内容（典型地基于单词或概念使用的直方图）。2) 创建相似性矩阵 - 利用每个文档的语义向量，计算每个文档对的相似性矩阵并由此创建文档相似性矩阵。3) 基于该相似性矩阵创建二或三维投影 - 利用主要成分分析或类似的方法（例如，多维缩放），计算每

个文档在文档集内的位置以便各文档之间的距离最好地反映如由相似性矩阵所描绘的各文档之间的相似性。以及 4) 绘制可视化信息空间 - 利用二或三维投影, 将文档绘制为文档空间内的点。

[0057] 一些空间可视化工具采取进一步的步骤覆盖信息空间上的地形覆盖以揭示群集的程度。有些甚至可以基于群集之内的公共单词识别和标记群集的分组。

[0058] 空间可视化工具的一个例子是 Themescape 地图, 其是由 AuriginSystem 开发的专利分析工具包的一部分, 并且现在是通过其让受人 The Thomson 公司通过其的附属 Micropatent 提供的提供物 (offering) 的一部分。Themescape 可视化工具使用有关专利标题、摘要、或全文 (以用户的判断力) 的语义分析基于上述的方法创建信息空间的二维投影。如图 4 所示, Themescape 使用地图隐喻并利用表示信息空间的最为高度群集的部分的山覆盖信息空间之上的地形。Themescape 地图的用户通过在信息空间搜索公司名称和其它关键字 或通过选择文档群集以阅读或者输出回一个文档列表用于进一步的回顾或分析来探索地形。

[0059] Themescape 工具的基础技术来自西北太平洋国家实验室所作的研究, 该实验室还有已知为 SPIRE (Spatial Paradigm for Information Retrieval and Exploration) 的空间可视化工具。如图 5 所示, Spire 有两种可视化模拟, 其一, “星域 (Starfield)” 在看起来非常象布满星星的天空的视图中以三维显示文档图。其二, “主题视图” 是非常类似于利用 Aurigin 的 Themescape 地图的实现的地形隐喻。

[0060] 虽然在开发包含在大数据集内的信息的一般理解方面的确有用, 但对发明人已知的空间可视化工具使它们的可视化独自基于包含在文档之内的潜在语义信息, 并同样遭受如上面在动态分类法的章节中描述的语义分析的限制。

[0061] 网络可视化 - 最后一种可视化技术是网络可视化, 其有时被应用于增进与大数据集有关的元数据的理解。在其最简单的形式中, 网络图 (数学家可能会称其为图形) 仅仅是由链路 (也称为边或带子) 连接的一组节点 (典型地表示为圆点)。网络图不是新事物, 某些网络概念至少可追溯到古希腊。社会网络分析在二十世纪三十年代显著发展。具有强大的处理器的现代计算机的发展使得可能创建计算机化的网络可视化工具。

[0062] 网络范例是应用于分析大数据集的非常有价值的方法。网络透镜 (network lens) 如此有价值有两个特殊的原因。首先, 大多数可视化工具被设计为注意力放在被分析的实体上 (典型地为文档、个人或机构)。当网络可视化显示有关单个实体的信息时, 它们还重点强调那些实体之间和它们之间的关系。网络显示不仅示出了实体, 而且示出了这些实体在其中操作的系统。最近几年中, 不同的科学和学术研究人员认识到反论证法分析 (例如, 集中于将问题分解为其组成部分并充分分析每个成分的分析法) 是受限的。类似生物学、遗传学、生态学、社会学、物理学、天文学、信息科学和许多其它学科领域基于系统分析法都经历了发展。系统分析法并不集中于最小的元素 (例如, 基因、原子 - 或可能为夸克和比特) 之上, 而是集中于这些元素之间或之中的相互作用。网络工具在其本质上是系统可视化工具。其因此可以导致与现有技术范围内的其它可视化工具完全不同类型的理解和结论。

[0063] 网络可视化工具适合于分析大数据集的第二个原因是网络具有从各种视点观察同一信息集合的潜力。现有技术的网络可视化系统没有抓住这个事实的显著优点, 但网络

具有从一种透视 (perspective) 转换到另一种的潜力, 每一种透视都提供了有关被分析的数据的不同理解。下面对网络可视化系统的描述将描述如何实现这种潜力以便戏剧性地改进能够获得的有关巨大和复杂数据集的理解。

[0064] 然而, 首先需要了解网络可视化的当前技术发展水平, 并识别现有工具的关键局限性。有多种计算机化的网络可视化工具, 包括:

- [0065] ● aiSee (www.aisee.com)
- [0066] ● Cyram NetMiner- (www.netminer.com)
- [0067] ● GraphVis (www.graphvis.org)
- [0068] ● IKNOW
- [0069] (<http://www.spcomm.uiuc.edu/projects/TECLAB/IKNOW/index.html>)
- [0070] ● InFlow- (www.orgnet.com/inflow3.html)
- [0071] ● Krackplot
- [0072] (www.andrew.cmu.edu/user/krack/krackplot/krackindex.html)
- [0073] ● Otter (www.caida.org/tools/visualization/otter/)
- [0074] ● Pajek (<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>)
- [0075] ● UCINET & NetDraw (www.analytictech.com)
- [0076] ● Visone (www.visone.de/)

[0077] 这些工具中的每一个都能够创建网络图。更高级的包 (例如 UCINET/NetDraw, NetMiner) 提供了可视化能力范围, 如:

- [0078] ● 选择可选的布局算法;
- [0079] ● 显示多节点类型;
- [0080] ● 基于属性值排序 / 改变颜色 / 选择节点的形状;
- [0081] ● 显示多链接类型;
- [0082] ● 基于链接的类型排序 / 改变颜色 / 选择链接的线型。

[0083] 所有这些工具都是通用的网络可视化工具。换句话说, 它们被设计用于显示以网络的节点和链接都被定义了的方式构造的任何数据的网络图。这些工具的每一个都使用特定的 (并且通常是唯一的) 文件格式以捕获有关节点和节点属性以及链接和链接属性的信息。节点信息是通过节点列表捕获的, 每个节点在该列表中由一个节点记录表示。节点记录包含至少一个对于该节点唯一的标识符的字段, 但还可包含提供有关该节点的信息的其它属性字段。链接信息是通过链接列表 (或链接矩阵) 捕获的, 该列表至少识别了是哪两个节点被链接了, 但也可捕获类似于链接强度、链接方向和链接类型的信息。

[0084] 尽管各个工具在它们的细节上各不相同, 但与它们一同工作的处理过程遵循如图 6 所示的共同的模式。任何已知的现有技术系统的用户从被利用的任何的来源收集数据。用户然后选择是数据内的什么实体将表示节点的定义以及选择她将使用什么信息创建节点之间的链接。然后数据必须被格式化以符合该网络可视化工具的特定文件结构。在所有的情况下, 这都需要用户创建节点列表和链接列表或链接矩阵。一旦被适当地格式化了, 网络数据文件于是可以被输入到网络可视化系统并被分析和可视化。用户可以在该工具的范围内与数据一同工作, 选择不同的布局算法或显示属性, 并利用任何提供的分析工具分析网络结构。

[0085] 如果用户想要利用节点和 / 或链接的不同定义开发可选的数据可视化,他必须从头开始,重新定义节点和链接,重新格式化数据为节点和链接列表并重新引入该新文件到可视化系统中。该系统于是可以基于节点和链接的该新定义显示网络图。这些现有技术系统的一些固有限制包括:

- [0086] ●由于数据库记录不包含可由系统使用的节点和链接信息,来自任何数据源的数据库记录都不能被可视化;
- [0087] ●存取和格式化数据的过程并没有集成到网络可视化工具中;
- [0088] ●用户必须格式化数据为节点 / 链接列表以适应系统;
- [0089] ●用户必须在格式化数据用于在系统中使用之前选择是什么构成了节点以及是什么构成了链接的稳定定义;
- [0090] ●当在网络可视化系统之内工作时没有改变节点和链接的定义的方式;
- [0091] ●如果选择了新的节点 / 链接定义,则即使是两个网络基于相同的基础数据,也没有结合或连接基于第一种定义的网络到基于第二种定义的网络的方式;
- [0092] ●没有用于指定与来自特定来源的数据重复地使用特别有用的节点和链接定义的方式。每一次来自该来源的数据将被可视化时,用户必须从头开始并指定每个节点和链接定义并操作该数据以适应该可视化系统。

发明内容

[0093] 在一方面,一种提供两个或多个数据库记录的网络图形表示的方法,包括根据一个或多个描述性标准选择所述两个或多个数据库记录。所述两个或多个数据库记录的每一个都是公共记录类的成员。所述方法还包括识别所述记录类的一个或多个属性,并使网络节点与来自所述数据库记录的所述一个或多个属性的实例相关联。所述方法还包括利用指定具有所述一个或多个属性的公共实例的网络节点的网络链接连接所述网络节点。

[0094] 所述公共记录类可包括来自数据库,如 LexisNexis 数据库、Thomson 数据库、USPTO 数据库、EPO 数据库或 Derwent 数据库,的专利记录。

[0095] 所述公共记录类可包括来自数据库,如 PubMed 数据库,的学术期刊文章。

[0096] 所述描述性标准可以包括例如 (i) 每个所述专利记录的主体字段内的一个或多个关键字; (ii) 每个所述专利记录的标题字段内的一个或多个关键字; (iii) 每个所述专利记录的发明人字段中的一个或多个发明人; (iv) 每个所述专利记录的受让人字段中的一个或多个受让人; (v) 摘要字段内的一个或多个关键字; 以及它们的组合。

[0097] 所述属性可包括例如发明人,受让人,申请日,颁证日,IPC 码,USPC 码,或搜索字段。

[0098] 所述网络链接可包括描述在所连接的节点之间出现的公共实例数量的特性。所述特征可包括例如链接厚度,链接颜色或链接结构。

[0099] 所述网络节点可包括元节点,其描述了两个或多个数据库记录的特性。

[0100] 所述方法进一步还包括在修改所述一个或多个描述性标准的同时迭代执行所述识别和连接步骤,以便改变所述被选择的两个或多个数据库记录。所述一个或多个描述性标准可包括例如日期范围。

[0101] 所述方法进一步还包括从除专利记录的所述公共记录类以外的记录类选择附加

数据库记录，并使得网络节点、网络链接或二者与来自所述附加数据库记录的一个或多个属性的实例相关联。所述其它的记录类可描述例如与所述专利记录相关的许可历史，与所述专利记录相关的诉讼历史或与所述专利记录相关的维持费历史。

[0102] 在另一方面，一种提供两个或多个数据库记录的网络图形表示的方法，包括根据一个或多个描述性标准选择所述两个或多个数据库记录。所述方法还包括识别所述数据库记录的两个或多个属性，并使网络节点与来自所述数据库记录的其中第一个所述公共属性的实例相关联。所述方法还包括利用指定具有所述两个或多个公共属性的其中一个的公共实例的网络节点的网络链接连接所述网络节点，以便形成第一网络图形表示。所述方法还包括通过使所述网络节点与来自所述数据库记录的第二个所述公共属性的实例相关联将所述第一网络图形表示变换为第二网络图形表示，并利用指定具有所述第二属性的公共实例的网络节点的网络链接连接所述网络节点。

[0103] 在另一方面，一种提供两个或多个数据库记录的网络图形表示的方法，包括根据一个或多个描述性标准选择所述两个或多个数据库记录。所述方法还包括识别所述数据库记录的两个或多个公共属性，使第一组网络节点与来自所述数据库记录的第一个所述公共属性的实例相关联，并使第二组网络节点与来自所述数据库记录的第二组所述公共属性的实例相关联。所述方法还包括利用指定所述网络节点之间的关联性的网络链接将所述第一组网络节点的一个或多个成员连接到所述第二组网络节点的一个或多个成员，以便形成第一网络图形表示。

[0104] 在另一方面，一种用于提供两个或多个数据库记录的网络图形表示的方法，包括根据一个或多个描述性标准选择所述两个或多个数据库记录。该方法还包括识别所述数据库记录的两个或多个公共属性，使第一组网络节点与来自所述数据库记录的第一个所述公共属性的实例相关联，并使第二组网络节点与来自所述数据库记录的第二组所述公共属性的实例相关联。该方法还包括在网络配置中表示的所述第一组网络节点的一个或多个网络节点内包含在网络配置中表示的所述第二组网络节点，其中所述第二组网络节点中的每一个与所述第一属性的网络节点共用一个公共属性实例，在所述第一属性的网络节点内含有所述第二组网络节点。该方法还包括使第三组网络节点与所述数据库记录的第三个所述公共属性相关联，在网络配置中表示的所述第二组网络节点的一个或多个网络节点内包含在网络配置中表示的所述第三组网络节点。所述第三组网络节点中的每一个与所述第二属性的网络节点共用公共属性实例，在所述第二属性的公共节点内包含所述第三组网络节点。该方法还包括使附加的一组或多组网络节点与来自所述数据库记录的其它多个公共属性相关联，并在其它网络节点内分组所述附加的多组网络节点的一个或多个成员，使得每组网络节点成员的特征在于与所分组的网络节点相关联的属性。

[0105] 在另一方面，一种包括计算机可读介质的网络可视化系统，计算机可读介质具有适用于提供两个或多个数据库记录的网络图形表示所存储的指令。所存储的指令实现了在此描述的一个或多个方法的步骤。

附图说明

[0106] 图 1 示出了利用结果过滤生成的图形表示的一个现有技术实例；

[0107] 图 2 示出了利用结果过滤生成的图形表示的另一个现有技术实例；

- [0108] 图 3 示出了利用结果过滤生成的图形表示的再一个现有技术实例；
- [0109] 图 4 示出了空间可视化工具的现有技术实例；
- [0110] 图 5 示出了空间可视化工具的另一个现有技术实例；
- [0111] 图 6 示出了与现有技术可视化工具一同使用的公共过程；
- [0112] 图 7 示出了在所描述的实施例中数据库记录是如何能够转换为链接数据的；
- [0113] 图 8 示出了基于少量数据库记录的简单网络图；
- [0114] 图 9 示出了在一个描述的实施例中的网络链接表示；
- [0115] 图 10 示出了其中具有相同受让人的文档被群集在一起的专利文档的简单网络；
- [0116] 图 11 示出了一个描述的实施例的按照年代顺序排列的网络图的实例；
- [0117] 图 12 示出了由表示来自 PubMed 数据库的单个文章的一个描述的实施例生成的网络；
- [0118] 图 13 示出了不使用元节点由一个描述的实施例生成的网络；
- [0119] 图 14 示出了利用元节点由一个描述的实施例生成的网络；
- [0120] 图 15 示出了图 7 的记录是如何转换为元节点的；
- [0121] 图 16 示出了一个网络实例，其中受让人节点和发明人节点之间的链接是基于发明人是否使受让人持有某专利的发明；
- [0122] 图 17 示出了具有表示 IPC 码的元节点的图；
- [0123] 图 18 示出了由一个描述的实施例产生的分形网络图；
- [0124] 图 19 示出了由一个描述的实施例产生的用户接口，以及图 19A 示出了在表示分组的单个受让人之下分组多个受让人姓名的结果；
- [0125] 图 20 示出了专利数据的属性和学术文献的属性之间的关系；
- [0126] 图 21 示出了基于 PubMed 数据并利用在此描述的 MNVS 可视化的典型网络；
- [0127] 图 22-25 示出了图 21 所示的可选网络视图；
- [0128] 图 26-27 示出了由一个描述的实施例产生的不同合作群集的网络；
- [0129] 图 28 示出了由一个描述的实施例产生的受地理限制的搜索的网络；
- [0130] 图 29 示出了由一个描述的实施例产生的受组织限制的搜索的网络；
- [0131] 图 30 示出了由一个描述的实施例产生的示意了跨越组织的研究协作或代替的网络；
- [0132] 图 31 示出了由一个描述的实施例产生的示意了研究强度的区域基础的网络；
- [0133] 图 32 示出了所描述的实施例的计算机实现。

具体实施方式

[0134] 如同在此描述的实施例中所使用的那样，网络可视化系统 (NVS) 是用于通过提供数据库记录的网络图形表示理解有关数据库记录或文档的集合的系统和 / 或方法。这些系统和 / 或方法可被应用于具有记录的数据库，其中可以建立记录之间和之中的关系。NVS 可以应用于其中的一些领域的例子包括但不限于专利文档、学术文章 / 论文 / 期刊、医学 / 科学文章 / 论文 / 期刊、文献、网页、消费者 / 产品 / 供应商 / 销售的公司数据库、公司知识管理数据库、零售数据库、人口普查信息 / 经济数据数据 / 等的官方数据库、会员资格 / 签约用户 / 机构人事关系的机构数据库以及许多其它领域。实际上，任何是或者可以被结构化

为具有两个或多个信息字段的信息表的信息都可以利用本发明可视化为网络。

[0135] 一个重要的理解是数据库记录 / 文档通过可以表示为网络的各种 属性相互关联。可以被用于创建记录 / 文件之中的连接关系的属性的类型可包括但不限于引用链接 (例如,由于文档 A 引用了文档 B 所以链接文档 A 和 B)、共同引用链接 (例如,由于文档 C 引用了文档 A 和 B 二者所以链接文档 A 和 B)、数目连结 (例如,由于文档 A 和 B 都引用了文档 C 所以链接文档 A 和 B)、公共原作者、公共从属关系 (受让人,公司,期刊等)、某些静态或动态分类法内的公共分类、公共关键字、语义相似性、以及许多其它可能的链接。这些连接使得可能将数据库记录或文档集合表示为网络,该网络使得能够使用各种网络统计信息和可视化工具以帮助用户理解所选择的信息。

[0136] 每个数据库记录的特征在于它们的属性的特定实例。例如,对于专利数据库记录中的“发明人关系”属性,该属性的“实例”可以是“John Smith”,即,某特定的发明人。于是例如共享属性“发明人”的同一个实例“John Smith”的两个专利可被认为是链接的,并由此可视化为网络的一部分。

[0137] 下面阐明 NVS 的详细描述,其描述了用于转换数据库信息或文档为网络信息的方法,并接着描述了用于创建多重网络可视化的方法。进一步公开的是对特定数据库 : 专利数据库和医学期刊数据库,的 NVS 的两个所选择的应用实例。应理解的是,这些仅是示例性实施例,并且本领域的技术人员将理解的是在各个实施例中描述的特定的方法可应用于其它实施例,也可应用于具有记录 / 文档的任何其它数据库,其中可经由如下面描述的各种连接类型建立文档之间的关系。

[0138] NVS 内的主要可视化范例是网络。网络是以某种方式相互连接的对象的集合。利用“图形”可视表示网络是常见的方式。网络图是一种可视表示,其中网络中的每个对象由称为节点的图标或符号 (emb1em) 表示,而且对象之间的每个连接表示为可视连接该节点的链接 (也称为边或带子)。这些节点和链接可以以这样的一种方式展示,以便提供网络中的不同的对象之间的关系的可视表示。

[0139] 找到一种适当的方式展示网络图以揭示对象当中的关系并不是一件无关重要的任务。图形理论和网络布局算法是非常确实的研究领域。已经开发出了多种布局算法以创建网络的有用的可视表示。网络可视化系统的目的是并不是要对现有图形布局方法加以改进。根据 NVS 的不同的实施例,可以利用任何网络布局方法作为可视化相关大专利文档或数据库记录集合的各种重要属性的方式。NVS 利用了现有布局算法以便显示数据库记录或相关文档的集合的网络图。

[0140] 由于将理解的有关大文档 / 数据库记录集合的关键属性之一是存在与它们当中或之间的关系,网络范例已经被选择作为用于可视化的基础。网络可视化在其本质上设计用于揭示相互关系并因此是理解大文档集合的非常有用的工具。

[0141] 数据的收集

[0142] 利用 NVS 的第一个步骤就是收集要检查的文档 / 数据库记录。有几种方法可用于得到用于分析的文档 / 数据库记录集。存储于电子数据仓库 (或者在同一计算机系统范围内或者在一个或多个远程在现场的或不在现场的服务器之内) 的数据可以以其整体或者作为记录的子集被 NVS 存取。这可经由计算机实现的或辅助搜索通过基于一个或多个描述性标准电子提交用户查询来实现。查询可利用常规的、非常确实立的逻辑语法提交,并可通

过在数据库记录的一个或多个字段范围内或者在数据库记录的整个“全文”范围内搜索用户指定的各个项来执行。然后可以分析全部数据或查询结果，并通过网络可视化系统可视化。在一个实施例中，电子数据仓库内的数据库记录是公共记录类的所有成员，例如来自 USPTO、EOP、Aureka、Micropatent、Thomson、Lexis Nexis 或 Derwent 数据库的专利，或包含在类似于例如 PubMed 数据库的许多学术、科学、工程或医学文档数据库的其中之一内的学术期刊文章。在其它的实施例中，数据库记录是两个或多个记录类的成员。

[0143] 也可通过 NVS 分析没有存储在电子数据仓库内的数据，然而，该数据首先必须经过数据输入、OCR（光学字符识别）、或其它适当的技术转换为电子格式。一旦数据被转换为电子格式，则利用 NVS 可将其分析作为任何其它的数据库。

[0144] 变换数据库记录为网络数据

[0145] 如上所述从数据仓库中提取出来的数据仅仅是一组记录。这个数据不能被认为是依据现有技术范围内的任何已知网络可视化工具的“网络”数据。这是因为其并没有被构建为节点列表和链接列表（或矩阵）。该数据仅仅是记录的收集，每条记录具有表示该记录的属性的两个或多个字段。例如，公司用户数据库可能有用户 ID、姓名、街道地址、城市、州（state）、邮政编码、国家、电话号码、e-mail 地址字段、以及许多其它字段。虽然这种数据作为其中每条记录被视作一个节点的节点列表是有用的，但是没有链接列表，因而数据不能被表示为网络图。

[0146] 网络可视化系统通过为每个记录链接属性创建链接列表将这种数据转换为网络数据。这是通过在共用同一属性值的每个记录之间创建链接来完成的。图 7 中示出的该图表提供了数据库记录是如何能够被转换为链接数据的简化例子。

[0147] 在这个例子中，我们使用了非常简化的专利数据库记录集。对于类似于受让人或发明人的属性，为共用同一属性值的每对记录创建链接。由于它们仅仅是基于属性值的共同出现的，故这些链接没有方向性。如果提供作为列表，则引用链接必须进行语法分析（在此情况下基于逗号划界），然后如该实例中所示的那样指定引用和被引用专利之间的方向链接。

[0148] 这种方法可用于转换任何的数据库记录集为网络数据，其中原始记录为节点列表并如上所述基于一个或多个属性的公共实例创建链接列表。一旦数据库记录已被转换为网络数据，则 NVS 允许用户可视化该网络。

[0149] 基本的数据库网络

[0150] 在一个基本的数据库网络中，每条记录可以由节点表示，而节点 可通过表示一个或多个如上所述的各种类型的连接的链接相互连接。图 8 示出了基于少量的数据库记录的非常简单的网络图。

[0151] 表示节点 - 节点可以被显示为基本形状（例如，圆、椭圆、矩形等）或图标（例如，文档图片）。节点的颜色可以改变以表示节点的某一属性。节点还可以使用识别该记录或显示该数据库记录的一个或多个属性的文本来标记。作为一个实际的问题，长的节点标记趋向于使网络显示难于操纵。NVS 以若干种方式致力于这个问题。首先，针对想要在节点内显示多少的属性值，为用户提供了多个选项。选项包括：全部（整个值），短的（首字或头“n”个字符），要点（仅两位数字的年度）和无（不标记）。通过无论何时用户指向或选择节点（利用诸如鼠标或跟踪球的电子指示装置）时允许用户看见完整的节点标记可进一步

解决该问题。

[0152] 表示链接 - 可由一条线或箭头表示的链接。参见例如图 9。通过附加一个箭头, 利用顶点指向被引用的文档的三角形形状, 或者通过利用不同的颜色或线型 (例如, 点线、实线) 表示前向和反向引用, 链接的显示可用于揭示连接的方向。此外, 通过变化的线的厚度或者通过改变其颜色或线型可直观地描述节点之间的连接的强度。链接的强度也可通过显示与靠近网络图上的链接的每个链接的强度有关的值来描述。节点之间的各种类型的链接可如下所述建立。

[0153] 作为一个实际的问题, 当多个链接连接到同样的两个节点时, 对于用户来说很难区分两个节点的各种相关的方式。NVS 以多种方式解决了这个问题。首先, 显示了不同的链接类型以便它们在视觉上不同。这是通过以不同的颜色、线型 (例如, 点线、实线、虚线)、线厚度等显示不同的链接类型实现的。节点之间的多个链接被并排排列以便可以显示多个链接而不会重叠。

[0154] 用于解决多链接类型的问题另一种技术是将这些链接压缩 (collapse) 成单个“复合链接”, 以及将表示不同类型的链接和约束强度的图标附加到该链接。图 9 示出了如何将这多链接类型压缩为具有图标的单个链接的实例。

[0155] 两个节点之间的这种复合链接的强度可以以多种方式来计算, 其可以简单地基于链接的数量、所组合的链接的强度之和、或者所组合的链接的加权平均值。如果使用加强平均值, 则可以基于每个链接类型的相对重要性的估计值来选择权重因数。

[0156] NVS 的另一个特征是用户具备用于选择哪些链接类型有效或无效以及哪些链接类型可见或不可见的工具。通过定义, “有效”链接是那些影响网络图的布局的链接。换言之, 它们具有将所链接的节点连接在一起的力 (类似于弹簧或弹性带)。然而, 不是所有的有效链接都需要以可视化显示。当网络高度群集时 (即, 链接高度集中) 或同时使用多链接类型时, 网络图可能变得是利用链接群集的。通过允许用户使这些链接不可见, 其允许用户将这些链接从网络图中去除, 同时继续使这些链接影响网络图布局。

[0157] 导航网络 -NVS 提供了用于导航网络的多种手段, 包括但不限于 :

[0158] ● 选择节点周围的半径 - 导航网络的一种方式是选择围绕被选节点的半径。在此环境中, 半径是被选节点和另一节点之间的链接数。例如, 如果该半径被设置为 3, 则将在网络图中显示从被选文档可通过少于 3 个链接到达的所有文档。

[0159] ● 扩展 - 网络可以被扩展为向网络图中增加附加节点。用户可以例如选择一个或多个节点 (其表示文档或数据库记录), 以及选择“扩展”, 并且可经由单个链接从该 / 那些节点 (尚不可见到达的) 到达的所有节点都被添加到该网络图中。

[0160] ● 收缩 - 网络可以收缩以从网络图中去除节点。用户可以例如选择一个或多个节点 (其表示文档或数据库记录), 并选择“收缩”, 从该 / 那些节点可在单个链接中到达并且不再以任何其它方式链接到网络的所有节点都被从网络图中去除;

[0161] ● 隐藏 - 节点可从网络图中隐藏。通过选择一个或多个节点并选择“隐藏”, 所选节点从网络图中被去除。

[0162] 过滤网络 -NVS 的另一有用特征是过滤在网络中表示的节点的能力。这可以通过以下几种方式实现。

[0163] 1) 滤波器可以通过指定要显示的专利文档的属性值的最小、最大或范围来应用。

例如,文档可以被过滤以便仅表示满足一组特定标准的那些记录,特定的标准有:

[0164] ○日期在特定日期之前或之后;

[0165] ○仅显示特定属性在记录集内出现多于或少于某一指定的最小次数的节点(例如,仅显示在数据集中有至少5篇文档的作者的文档节点)。

[0166] 2)可以通过指定要显示的节点属性的值来应用滤波器。例如,可以过滤这些文档以便仅表示如下文档:

[0167] ○与一个或多个公司有关的;

[0168] ○由一组特定的一个或多个作者撰写的;

[0169] ○基于某个固定或动态分类法在一组特定的一个或多个话题内分类的。

[0170] 3)可以通过为用户提供属性值列表并允许用户选择或不选择要显示的属性值类应用滤波器;

[0171] 4)可以通过为用户提供利用计算机指示设备(如鼠标或跟踪球)从网络可视化中选择一个或多个节点的手段并从指示应该过滤所选节点的菜单中选择一个命令来应用滤波器。

[0172] 通过单独或组合利用这些过滤方法,用户就可能动态过滤数据集以仅显示感兴趣的文档。例如,用户可以指定她只想看到以类别A1、C3和D5分类的1999年和2005之间出版的A、B和C公司的文档。

[0173] 由于以下两个原因,这种能力是非常重要的:1)其允许用户在数据集内的不同文档子集之间前后移动,以及2)其使用户能够精练他们的查询以去除他们不感兴趣的那些文档。

[0174] 根据属性群集-揭示网络中的文档之间的不同类型的关系的另一中方法是基于它们的属性将它们群集在一起。实现这种方法的一种方式是在共用特定属性值的节点之间放置附加链接。例如,具有相同受让人的所有专利都可以通过附加链接相互链接以便它们相互吸引并形成一个群集。图10示出了专利文档的一个简单网络,其中具有相同受让人的文档被群集在一起。或者,可以引入一个附加节点到表示该属性值的图形中,具有该值的每个节点与该新的节点链接。这就有效地将所有这些节点拉到一个群集内。注意,不必显示这种新的“属性节点”或该属性节点与可视化内的其它节点之间的链接。

[0175] 识别网络内的自然群集-所链接的文档的网络自然具有比网络中的其它区域更为高度群集或紧密组合的区域。“群集度”是社会网络领域内的一个术语,而且存在众所周知的统计方法用于确定网络的一部分或整个网络内的群集度。这些群集可以利用作为社会网络分析领域的一部分研发的技术来识别。存在各种识别群集的技术。改进已知的群集技术并不是NVS的目的,然而,NVS利用各种群集技术来识别专利的有关组合以便提供对大文档集合的本质的理解。

[0176] 一旦已经识别了这些群集,就可以对它们进行标记。一种标记方法是识别在落入一个群集的文档的标题和摘要的所有或多个标题和摘要中找到的单词。通过将头几个(通常为1~5个)单词串接在一起可以为每个群集创建一个标记。这个标记可以为用户提供一个有关每个群集的内容的信号。由于该群集中最频繁使用的单词的列表不太可能是理想的群集标记,实际中是为用户提供将该群集标记更改为一个更有意义的单词组的工具。

[0177] 按时间排列的网络图-揭示有关文档集的信息的另一种方式是以通过日期分类

节点的方式显示网络图。所使用的日期可以是与该文档和数据库记录关联的任何日期。例如，专利文档可以有许多与之相关的日期：优先权日、申请日、公开日、授权日、期满止日以及其他日期。例如，网络图可以将所有的最旧的文档放在网络图的左侧，而将最新的文档放置在右侧（或反之亦然）。可以在网络图的旁边放置时间线以展示技术发展随时间的进展。或者，网络背景的背景可以按年、十年或某个其它时间分割分为时间段，并以落入出现在适当时间段内的范围的该范围的文档进行标记。图 11 示出了这种按时间排列的网络图的一个实例。

[0178] 其它梯度 - 也可以沿除了时间以外的任何数量的其它梯度来显示网络。可量化的（或者使得可以被量化的）节点或元节点的属性可用作在其之上显示网络可视化的梯度。可选的梯度的一个简单实例是根据客户的年度花费分类的客户数据网络。

[0179] 利用“元节点”变换网络

[0180] 在此描述的实施例的一个中心新颖性特征是变换网络表示的能力。现有技术网络可视化工具维持了关于什么是节点和什么是链接的一种固定的稳定定义。例如，如果专利文档集上的数据被引入到其中一个现有技术网络可视化工具中，必须精确地定义什么是节点和什么是链接。如果，例如你选择使每个专利被表示为一个节点，而且共有发明人表示链接，则该可视化工具将在分析期间保持该节点 / 链接的定义而不会变化。NVS 根本上的不同在于其使得用户能够通过当在他们使用数据时重新定义什么是节点和什么是链接的定义来变换网络。

[0181] 网络可视化系统工作于以下原理：数据库记录的任何属性都可以表示为节点、链接或二者。作为一个简单的实例，如果会议组织者有各个会议专题组以及每个专题组的参会人员的列表，则他可以将它们可视化为根据共同参会者链接的专题组的网络，但是他可能只是将他们视为根据他们一起参与的专题组所链接的参会者的网络。

[0182] 在极端情况下，甚至于单个数据库记录可以被视为具有由节点所表示的每种属性以及具有基于其它公共属性所链接的各种属性的网络。图 12 示出了表示来自于 PubMed 数据库的单个文章的相当复杂的网络。中心节点表示该文章本身，而该文章的各种属性由其它节点表示并且基于诸如共同作者链接和其它公共出现的链接相互连接，因为这些属性都出现在所选择的选文章中。

[0183] 网络可视化系统不仅将数据库信息变换为网络信息，而且允许用户创建他自己的节点和链接定义，在单个网络上组合任意数量的节点，以及在分析期间随意改变它们的定义。借助先前已知的方法不可 能以这种方式变换网络可视化。

[0184] 创建可选节点定义的能力是简化网络显示和开发有关数据集的理解的强有力的工具。通过在网络显示中重新定义节点和链接，用户可以将他的注意力集中到其感兴趣的实体上。例如，分析专利数据的研究人员可以关注公司、企业或发明人而不是专利。这些节点表示比单个文档或数据库记录的节点更为高级别的实体。我们将这些更高级别的节点称为“元节点”，因为它们代表文档或数据库记录组而不是单个记录。这些元节点之间的链接我们称为“元链接”，因为它们代表由元节点所表示的文档或数据库记录的集合之间的链接的聚集。这种将网络抽象化为“元级别”的能力使得用户能够以比使用任何其它已知可视化方法可能的更高级别地回答问题和通知决定。

[0185] 网络变换方法的能力可以以实例来论证，设想 > 1000 个专利文档的复杂网络，其

中节点是专利文档而链接是引用连接。该网络图看起来可能有些类似于图 13 中的图片。

[0186] 很难确定从该网络图中可以理解到什么。然而,如果你通过重新定义节点的定义来变换网络使得每个节点是一个公司,则你最后得到的网络图就类似于图 14 所示的网络图。

[0187] 与特定摄影技术有关的专利文档的这种网络图使得很容易识别在该技术领域领先的公司并且了解它们之间的联系。通过变换该网络图,其已经被大为简化,因此能够实现更好的理解。

[0188] 之前,我们描述了将数据库记录变换为网络数据的实施例。该实施例依赖于稳定的节点定义,即每个数据库记录是一个节点。另一个实施例从数据库记录中创建了元节点数据和元链接数据。图 15 所示的实例利用早先在图 7 中所示的数据库记录的相同简单集合来论证它是如何实现的,用于创建受让人元节点和元链接。

[0189] 该处理的第一步是创建元节点列表,这是通过简单地列出特定记录属性的每个唯一值并记录该值在数据集中出现的次数实现的。然后基于在其它属性字段(例如,发明人、IPC 分类和引用)中的共同出现的值为这个属性(在本例中为受让人)创建一个或多个元链接列表。除了两个区别之外,该方法与上述的用于创建链接列表的所使用的方法一致。首先,在此实例中的“记录”并不是来自于数据库的实际记录,其是来自于刚刚创建的元节点列表的记录,其次,该元链接具有指示在该链接中聚集的共同出现(或引用)的数量的链接强度值。

[0190] 从数据库记录创建链接列表、元节点列表和元链接列表使得可能在文字上从差不多任何数据库看到数据库信息,作为使用在此描述的网络可视化系统的一个网络。作为实际的问题,所描述的 NVS 的实施例实际上并不将每一个可能的属性转换为链接列表,也不会将每一个属性转换为元节点列表或元链接列表。只有那些对用户的目的最有用的属性才被转换为网络数据。

[0191] 对本领域的技术人员显而易见的是,存在用于选择要转换哪些属性以及在什么分析步骤作出该选择的可选方法。例如,有时希望提前定义对于感兴趣的特定数据库将哪个属性转换为链接和元节点数据作为计算机程序的一部分。这允许用户访问节点、元节点、链接和元链接的标准集合以便在他借助该工具进行研究期间与之一同工作。网络可以被过滤、变换以及利用多个节点、元节点和链接来展示,但是仅仅在为该分析之下的特定数据集建立的属性的界线之内。

[0192] 或者,可能给予用户从数据库记录中选择属性以便在其分析期间转换为链接、元节点和元链接的能力。这可以通过简单地允许用户从列表中选择属性(字段)以转换为网络数据来实现。一旦选择了属性,就可以根据上述的方法生成链接、元节点和元链接,并将它们加入到用户可用的网络可视化资源的集合中。

[0193] 还存在从数据库记录中创建元节点和元链接信息的其它方式。以下的实例示出了创建这种信息的两种可选方式,尽管除了这两种方式之外还可以使用其它方式。

[0194] 实例 1- 可以基于属性值的范围创建元节点和元链接。例如,如果数据库记录属性是数值(例如,在客户数据库中记录年度销售额的字段),则可以基于该属性字段内的值的范围来创建元节点(例如,< \$200 = 低消费者, \$200-\$1,000 = 中等消费者, > \$1,000 = 高消费者)。

[0195] 实例 2- 元节点和 / 或元链接可以基于多记录属性的组合。例如，市场调查结果的数据库可以被转换为网络数据，其中特定的客户类别可以基于它们对一组问题的共同回答组合在一起。通过这种方式就可能基于 $> \$50,000$ /每年的收入、孩子数 ≥ 2 ，以及车型=SUV 或小型货车为“Soccer Moms”定义一个元节点。

[0196] 图 14 中的网络图还示意了网络可视化系统的两个附加属性：元节点大小估计和元链接聚集。

[0197] 元节点大小估计 - 在图 14 的网络图中，每个元节点不是表示单个专利而是表示共享受让人属性的公共值的所有专利。换言之，每个节点表示由同一受让人提交的所有专利。在此图中，节点的大小是基于转让给该特定受让人的专利数量，而且将该数量附加到这些元节点以显示与元节点大小相关的值。

[0198] NVS 的另一个特征是为用户提供基于所表示的文档的各种属性估计元节点大小的能力。例如，在具有年度消费的客户数据库中，可以基于年度消费的总和（或平均值）估计由元节点表示的所有客户的大小。也可以基于任何数量的网络统计计算，类似于中心性 / 本征向量中心性 / 中间中心性之和，为所表示的节点对元节点估计大小。基于这些各种度量估计节点大小的能力使得用户能够对类似节点值的事物以及对用户感兴趣的其它重要度量得出结论。

[0199] 用于元节点大小估计的许多可能的属性可以应用于单个文档节点以及元节点。特别感兴趣的是引用属性（前向引用，后向引用，总引用）和社会网络的中心性统计（中心性，本征向量中心性，中间中心性）。基于这些和其它统计估计节点大小可以提供节点在网络内的值的信号。

[0200] 元链接聚集 -NVS 的另一特征是将网络从二元（关 / 开）链接网络变换到元链接网络（具有不同程度的强度的组合链接）的能力。这种将链接聚集为元链接通过揭示元节点之间的关系的强度和性质还为用户提供了进一步的理解。

[0201] 在上述的例子的情形中，链接表示受让人之间的引用。示出了多链接是因为引用可以在公司之间的任一个方向上流动。在本例中链接的值是基于一个公司的专利和另一个公司的专利之间的引用总数。这就揭示了在这个“技术革新网络”中谁是领先者，以及谁是跟随者。箭头被附加到链接以示意链接的方向，附加数字以示意与链接强度相关的值。此外，在一个优选实施例中，当用户指向（利用诸如鼠标的电子指示装置）一个特定节点时，输入和输出链接以不同颜色突出显示以提供关于所选择的公司是领先者（高度被引用）还是跟随者（引用别人的）的直观线索。

[0202] 如同节点大小估计一样，链接强度也可基于各种不同连接属性。一些例子包括引用数量、被引用的唯一文档的数量、引用的文档的数量、平均引用年度、最近的引用年度、以及其他属性。还要提醒的是，也可通过上述的更多种多样的链接类型连接元节点。这些链接也可以被聚集，而且它们之间的联系的强度可以基于与在此描述的度量类似的度量来确定。

[0203] 多节点、元节点和链接类型的同时显示

[0204] 元节点概念的下一扩展是同时将多个节点和链接类型放置在同一个图上。例如，在专利环境中特别是揭示了看到包含表示受让人和发明人二者的节点的图。图 16 示出了受让人和发明人的网络的一个实例，其中受让人节点和发明人节点之间的链接是基于受让

人是否持有发明人的专利。在图 16 所示的网络中,可直观看出哪个发明人为哪个公司工作,以及在所审查的技术领域范围内哪个发明人为多个公司工作。

[0205] 作为另一个例子,可以在同一个图之上显示表示专利的节点和表示 IPC(国际专利分类)、USPC(美国专利分类)、和 / 或 Derwent 分类的元节点。图 17 示出了一个图,其中元节点表示 IPC 码而专利被分组为特定 IPC 的成员。如果过滤器设置为仅查看来自特定受让人的专利,则这个实施例允许用户直观地确定受让人过去一直在投资什么技术,以及这些优先权是如何改变的。

[0206] 表示连接的不同连接属性或类型的节点和链接相互之间可直观地区分以便于提高系统的可用性。节点可根据形状、颜色、边框类型、填充模式、或根据每个表示以特定图标,例如一个人表示发明人而一个文档图片表示专利的特定图标,加以区分。链接可以通过形状、颜色、线型(例如实线、虚线)或其它方式区分。

[0207] NVS 允许用户选择在图上显示什么节点和元节点,以及选择使用哪个链接属性作为链接基础。这就提供了用于阐明大专利集合并理解它们的内容以及它们当中的关系的有力的工具。

[0208] 分形网络

[0209] 元节点概念的另一种扩展是分形网络的概念。分形网络在此定义为包含在每个元节点内的元节点网络,其它节点或元节点的网络,如图 18 所示。这种分形节点表示可以具有所期望的多个层。

[0210] 使用分形节点的一个实例可能是表示受让人的元节点网络,其中每个节点是通过其表示的专利数量估计大小的。在每个受让人元节点之内,可以显示表示 IPC 分类的元节点网络。这种表示将显示专利集中的每个受让人公司正在就哪个 IPC 分类进行研发。此外,在每个 IPC 元节点之内,可以显示表示发明人的网络。并且在发明人元节点之内,可以显示表示专利的网络。

[0211] 这种网络表示允许用户提出和回答有关技术领域内的专利文档的更广泛的问题。其使得能够在其它情况下不可能的方式深入研究文档集合中的专利的属性和它们之间的关系。分形节点的一种理想实现为用户提供了选择由分形网络图的每一级的网络所表示的属性和链接属性的能力。

[0212] 分形节点尤其还在显示诸如各种分类方案的分层属性中有用,分类方案包括专利领域内的 IPC 和 US 专利分类,医学数据领域内的医学主题标目(MeSH)以及诸如 Vivisimo 类别的分类。该分层结构的较低级别可以在表示该分层结构的较高级别的节点内表示。这种直观 表示为用户提供了理解每个类别和子类别的相对大小以及在它们之间的关系的直观方式。

[0213] 直到目前为止,已经通过假设在该分层结构的每一级只可以显示单个节点或元节点类型描述了分形网络。通过为用户提供在该分层结构的每一级放置多个节点和链接类型的手段可以生成其它附加的理解。用户于是可以更深入地研究各种属性是如何彼此相关的。作为这种能力的一种实例,用户可以显示受让人元节点网络并且在每个元节点内显示表示发明人和 IPC 分类的元节点。通过这样做,用户可以快速理解每个公司当前正在致力于什么技术领域以及什么那些技术领域内的关键发明人是谁。

[0214] 使用分形网络的一个挑战是每个元节点内的“子网络”在网络显示内可能非常小

的事实。为了解决这个问题，该系统允许用户放大和缩小该网络以便以任何的详细程度显示他们所选择的分形网络。这是以两种方式的其中之一实现的。第一，用户可以从工具条按钮或菜单选择放大级别。第二，用户可以简单地通过从网络显示中选择元节点在特定元节点内的分形网络上放大。通过选择元节点（利用鼠标和其它电子指示装置），系统可以自动以该元节点为中心并放大，使得该分形网络的下一级可以被清楚地看到。为了重新缩小回去，用户可以或者从工具条按钮或菜单选择选择一个新的放大级别或者可以点击元节点的外部以返回到前一放大级别。

[0215] 对元节点和元链接的隐式过滤

[0216] 如上所述，对用户而言有各种可用的过滤待审数据库记录的手段。这种过滤对于元节点和元链接的使用有重要的暗示，即，每次应用过滤时元节点列表、元节点大小、元链接列表和元链接链接强度都将遭受改变。值得注意的是，每次对数据应用过滤时，元节点和元链接信息都必须更新以便保持待审记录集与网络显示中的元节点和元链接相关联的值之间的一致性。

[0217] 提供有关网络的统计信息。

[0218] 帮助用户理解大文档集的 NVS 的另一要素是清楚地表示关于考虑之下的文档集的统计信息。之前描述的用户界面允许用户与网络交互，扩展和缩小该网络以创建表示用户感兴趣的区域的网络。在 NVS 中，提供一个当考虑之下的网络改变时动态更新有关网络的统计信息的界面。

[0219] 通过该系统提供有关网络的各种统计信息，包括但不限于：

[0220] ● 文档 / 记录计数；

[0221] ● 元节点计数（例如，所显示的网络中的受让人数量）；

[0222] ● 节点属性值之和（例如，网络中的所有客户的总销售额）；

[0223] ● 根据元节点类别的文档计数（例如，每个作者的文章列表）；

[0224] ● 每年的文档图（例如，根据出版年份的每年的文章）；

[0225] ● 其它网络统计 - 也可以提供其它网络统计，包括但不限于：有关网络的统计（例如，密度、直径、集中化、健壮性、传递性），有关网络内的群集的度量（例如，小集团、自我网络、密度），以及有关节点的度量（例如，中心性（例如中间中心性、本征向量中心性）、等效性），以及其它网络统计。

[0226] 上述的统计信息也可根据用户请求的菜单和工具条选择来提供，或者可以在界面内的单独窗口或窗格中提供。在优选实施例中，提供一个单独的窗格，其具有标签以允许用户访问所期望的有关当前网络的信息。如图 19 所示，这种窗格可以按照用户的意愿扩展、缩小或关闭。

[0227] 通过允许用户选择一个或多个类别并突出显示网络可视化中的相关节点该界面还有利于进一步的理解。

[0228] 在用户使用该系统时，可以基于环境敏感提供附加信息。具体而言，弹出窗口可用于提供有关网络图中的各个节点、元节点、链接以及元链接的附加信息。在每个弹出窗口中提供的信息与从该图中选择的对象有关，注意可以每次选择一个以上的同一类型的节点的对象（节点、基于相同属性的元节点、相同链接类型的链接）。

[0229] 解决歧义的属性值

[0230] 在使用上述的网络变换方法中遇到的一个问题是需要解决歧义术语。数据库管理员或用户将认识到在数据库记录中包含的数据经常是混乱的和不精确的。由于文本中的细微差别,表示同一值的属性值经常不同。我们已经发现发明人姓名和受让人姓名(以及其他属性)经常以不同形式出现在专利数据库中。例如,受让人“IBM”可以表现为 IBM、IBM 有限公司、国际商用机器、国际商用机器公司以及其它变化形式。

[0231] 这就在使用上述的元节点和元链接方法时产生了一个问题,因为这些形式上的小的区别导致系统在实际上组合这些元节点 / 元链接时产生多元节点 / 元链接。因此,该工具为用户提供了将属性值组合为单个值的多种手段。

[0232] 该系统通过允许用户在单个值之下将属性值组合到一起为用户提供了解决模糊属性值的手段。图 19A 示出了在表示该组的单个受让人姓名之下分组归类多个受让人姓名的结果。可以提供多种手段来实现这种方法。第一种方法是允许用户从列表中选择属性值并将它们结合到新的名称或属性值之下。例如,给用户提供一个按字母表顺序的受让人列表并从该列表中选择 IBM、IBM 有限公司、国际商用机器、国际商用机器公司。用户于是可以利用工具条按钮或菜单选择将所选择的选项归类到一起,并且或者为组名(例如 IBM)选择系统建议,或者键击出他自己的组名。系统于是在该新组名之下组合所有的这些名称并为整个的分析的目的将其显示为单个受让人。

[0233] 第二种方法为用户提供要组合成组的属性值的建议集合。系统比较这些属性值的相似性并建议在单个属性值之下将这些组群集在一起。除了利用正在考虑的属性(例如,受让人),该工具还对于属性值应当组合的线索考查其它属性值。例如,如果 IBM 和 IBM 公司都位于 Armonk, NY, 或者它们共有相同的发明人,该工具建议它们应当可能组合。用户可以回顾每个建议的组,并在选择接受该分组之前从该列表中增加或去除值。

[0234] 解决模糊属性值的最后一种方法是使用网络图本身。用户可以直接在网络图上选择元(利用诸如鼠标的电子指示设备)节点。其然后能将多个元节点组合成单个组。这是通过选择多个元节点并选择工具按钮或菜单选择来组合这些值来实现的。或者,用户可以将一个元节点“拖放”到另一个元节点之上从而建议系统它们应该组合。该系统将提示用户确保组合那些项是真正的意图,然后为分析目的将这些属性值组合为单个组。

[0235] 系统使得一旦它们已经被创建也可能取消属性组的组合。用户可以从列表选择元节点或选择组名并回顾他们已经组合了哪些属性值。用户于是可以选择解除特定属性值的组合,然后选择一个工具条按钮或菜单选择以取消所选择的值的组合。

[0236] 除了上述方法,还可能通过与外部数据源的比较去除属性值中的歧义。例如当考虑受让人时,可以参考外部的公司名等同物的列表。这些列表可以包括子公司和获取的公司,它们可以被建议给用户作为可能用于组合的组。在医学领域中,可以利用医生的 DEA 号来解决医生名的问题。

[0237] 这种将多个属性组合为单个属性值的处理也可以在另一方面上有益。通过将值组合为组,创建了值的分层结构。这种信息接着可根据上述的方法用于显示在该分层结构的不同级之上的数据之间的关系。具体来说,在该分层结构的每一级的属性值可以表示为元节点而且可以被显示为网络显示的一部分,或者作为网络图内的独立节点,或者作为利用上述分形网络方法的分层网络。这种方法尤其对于类似于有关受让人的起源 / 辅助信息的分层信息特别有价值。

[0238] 动画网络

[0239] 到目前为止我们描述的工具允许用户以各种方式变换他的网络视图。然而,到目前为止的描述假设每个网络图是特定时间点的快照。那样,我们所描述的可视化到此为止是静态的。

[0240] 网络可视化系统的另一重要要素是使网络图有生气以揭示随时间的过去它们是如何变化的能力。系统有几种不同的使得用户能够审查动态网络出现的能力。

[0241] 用于揭示网络动态的第一种方法是基于感兴趣的时间周期限制图表之内显示的数据的能力。用户可以为将要显示的数据建立日期范围的最小和最大日期。使用的实际日期可以基于与基础数据库记录有关的任何数据信息。就专利数据来说,可以选择多种日期,包括但不限于优先权日、申请日、公开日、以及授权日。一旦用户选择了日期类型和日期范围,系统于是过滤数据并仅基于该符合特定参数的数据显示网络图。

[0242] 第二种方法根据这种能力建立。系统为用户提供以非常简单的方式改变日期范围的能力。用户可以选择改变日期范围的“步长”(例如,一个月、一年)并且接着可以点击单个按钮以根据该增量向前或向后移动日期范围。此外,提供单独的工具条按钮以便可以通过单次点击调整该最小日期、最大日期、以及二者。一旦用户点击以更改日期范围,系统迅速调整数据集以反映最新选择的范围并重画网络图。这允许用户以指定增量逐步贯穿数据的时间周期。有效地,系统使得可能可视化随时间的过去网络是如何显现的。

[0243] 第三种方法是网络发展的真实动画的创建。系统为用户提供了一种方法以键入总的日期范围、初始日期范围(其可能根本就没有范围-例如如果最小和最大日期设置为同一个值),该日期(最小、最大、或二者)将被改变,以及键入动画的日期增量大小。系统使用这些输入以基于所提供的增量自动逐步贯穿特定的日期范围并显示网络随时间的过去的出现的动画。

[0244] 这些动画方法在揭示网络发展中令人无法置信地有用,然而,它们造成一些必须克服的挑战以便使得系统实用。首先,当正显示的网络比较大(其中有许多节点和/或链接,和/或大量的基础记录),高度的计算复杂性使得在除了最强大的计算机系统外的所有计算机系统上的动画较慢而且不平稳。为了克服这种限制,为系统提供了一种手段以批量处理网络图序列然后将它们保存作为一系列快照或作为视频剪辑。快照或视频剪辑接着可以以用户所选择的速度回放而无需系统必须在动画中的每一个点重新计算基础数据。这使得用户可能反复回顾动画并且还可随意暂停、倒带、以及快进该动画。

[0245] 与网络图动画有关的第二个挑战是难于彻底了解图内发生了什么。当一个图被动画后,由于各个节点和元节点之间的吸引力随时间流逝而改变,出现了新的节点和链接、元节点增长和缩小、以及节点在图内改变了位置。所有这些同时发生的改变使得用户难于理解当动画展开时发生了什么事情。为了使事情对于用户更为简单,系统提供了一种减少动画期间改变的参数的数量的工具。特别地,系统允许用户在动画期间保持各种参数恒定不变。能够在动画期间保持恒定不变的参数包括但不限于:

[0246] ●不变的节点位置-要跟随的动画的一个最难部分是在动画运行时网络显示内的节点的位置改变。因此,系统给用户提供了在动画期间保持节点位置不变的选项。为了实现这个功能,系统首先计算每个节点在动画结束将保持的最终位置,然后随着节点出现、改变大小、以及新的链接出现、生长并消失,节点的位置在这个位置保持不变。

[0247] ●呈现所有节点 - 系统提供的另一个选项是用户在整个动画期间保持所有节点呈现在该图上的能力。利用所选择的这个选项，系统在整个动画期间保持每个节点可见，但是对有区别的节点提供一个可视信号，这些有区别的节点不表示落入该动画中的每个点处的特定日期范围内捕获的日期范围内的数据。该可视信号可以在颜色（例如，通常不可见的节点为灰色）、大小、形状、边界和其它可视属性上有所区别。这允许用户能够在该动画的整个过程期间连续跟踪每个节点的路径。

[0248] ●在动画期间可以保持不变的其它参数包括链接存在，不变的节点大小，以及不变的链接大小。

[0249] 在动画期间保持这些参数的任一组合不变的能力给予用户对所 显示的动画的大力控制以及增加了他们彻底了解网络如何显现的能力。

[0250] 系统提供的另一能力是提供有关在动画期间与各种参数相关的变化率的可视信息的能力。虽然网络显现的动画提供了有关网络如何发展的有力的可视信息，但是很难精确地比较和评估它们出现时的变化。例如，虽然很容易看到有多种技术和公司文件夹在增长，用户很难或不可能评估在任一时间点哪个公司文件夹增长最快。

[0251] 为此，该系统为用户提供了可视化动画期间各种参数的变化率的工具。在动画期间用户可能有兴趣的一些参数是：节点增长率，链接附加率，以及网络中的节点中心性的各种度量的变化率。系统为用户提供了跟踪动画期间的这些变化率以及在表格、图中或直接在网络图中显示有关这些值的信息的能力。用户可以选择跟踪哪些变量以及为哪些节点和链接跟踪这些变量（如果需要的话包括所有节点和链接）。该数据可以显示在与网络图相邻的表格和图中（柱状图或线图）并在当显示动画时更新。此外（或可选地），数据可以用于在显示动画时改变网络显示中节点和链接的外观。我们发现基于时间改变率或中心性统计改变节点或链接的颜色以便显示网络图的哪些部分“最热”是有用的（如根据如上所述的参数度量的）。或者，数据可以用于在显示网络动画时改变节点或链接或某一其它可见特性的大小。

[0252] 提供了有关于网络动画的一个其它有用的能力。通常，用户对网络图的特定部分感兴趣，并且想要对网络的该部分在整个过程中是如何显现的深入理解。为此目的，提供一种手段借此用户能够在动画期间放大、和 / 或保持集中于特定节点或各个节点是有益的。例如，公司可能对他们自己的专利文件夹随时间过去是如何显现的特别感兴趣。提供一种手段以便用户能够选择在动画期间要放大的特定节点。提供一个或多个“画中画”显示是有益的，以便用户能够观察整个网络的显现，以及观看网络的一个或多个“放大的”部分是如何显现的。这在当被动画显示的网络为分形网络时以及用户对观察较大网络之内的“子网络”是如何显现的感兴趣时特别有用。

[0253] 网络动画是用于揭示网络之内显现模式的强有力的工具。在专利数据环境中，动画揭示了随时间的过去技术是如何出现的、公司的位置是如何改变、发明人的职业经历和合作者是如何改变的，以及许多其它的特征。这种能力为用户理解大数据集的能力提供了重要的贡献。

[0254] 链接倒外部数据

[0255] 到目前为止，已经基于可从单个数据源，在此情况下为单个数据库记录集，生成的网络分析和理解描述了网络可视化系统。显而易见的是，根据之前的描述，通过利用这种内

生数据可以简单地生成极大的理解。然而,当连同专利数据使用其它外生数据源时可以生成附加的理解。

[0256] 通过链接到外生的数据源,可以获得有关由数据库属性表示的实体的附加的信息。选择链接到哪一个外部数据以及链接到该数据的值依赖于正在审阅中的数据源的环境。数据库记录的每一个属性创建潜在链接到可扩展有关感兴趣的主题的可用信息的附加外部数据源。这种附加的数据可被用于完整地创建新的元节点分类,附加另外的属性到一个或多个数据库或元节点记录,提供有关特定节点或元节点的信息,以及提供节点或元节点之间的附加链接信息。在本申请中随后将在某些优选实施例中描述这种类型的有用的外生数据的特定例子。

[0257] 网络可视化系统的优点

[0258] 这些不同工具和技术的组合在使得用户能够迅速理解大数据集内包含的文档或记录方面提供了戏剧性的改进。首先,通过动态过滤快速识别和精练文档集合的能力使得用户可能更加快速、有效、精确地识别有关于他们感兴趣的领域的文档集合,并且需要更少的对于特定技术的知识。

[0259] 其次,用户具有探索大的文档和数据库记录集合使得能够快速产生对该领域内的活动的特性的理解的能力。这是通过提供有关来自多种不同观点的有关该领域的总结信息实现的。网络透镜与元节点组合以表示文档的各种属性提供了不仅理解该领域内固有的分组而且理解那些分组之间的关系的直观方式。

[0260] 用户可以以所期望的任何详细程度有利地探索他们感兴趣的领域,而且在摘要级信息和详细信息之间无缝地前后移动。

[0261] 实例 1 :用于理解专利数据库的方法和装置

[0262] 本发明的一个和多个实施例涉及用于理解专利数据库中的专利的改进的方法。专利数据库的一个属性是很容易基于它们的引用关系建立文档之间的相关度。以下讨论的关于利用引用作为建立专利之间的相关度的基础同样适用于具有引用的数据库,包括但不限于学术 / 科学 / 医学文献以及在万维网页面上嵌入的超链接,这些超链接也可以认为是一种引用。

[0263] 本发明的各种实施例也可以用于帮助商人、工程师、科学家、代理人、专利审查员和其它有兴趣的一方来理解巨大的专利集合。这种挑战是选取一个大的专利文档集合并找到不用必须阅读它们就理解它们所描述的技术发展的方式。为了实现这一点,提供一种方法,通过这种方法,用户可以可视化文档的各种属性和它们之间的关系。

[0264] 可以利用本发明的各个实施例解决的一些问题包括但不限于 :

[0265] ●有关本专利组的是什么技术?

[0266] ●各种技术发展有多快?

[0267] ●在此领域中什么是最热门的技术领域?

[0268] ●最近的发展是什么?

[0269] ●哪个公司在开发这些技术上最为活跃?

[0270] ●哪个发明人在开发这些技术上最为活跃?

[0271] ●哪个专利是最为重要的?

[0272] ●各个公司参与这个技术领域的专利文件夹有多重要?

- [0273] ● 哪个公司领先开发,哪些公司跟随其后?
- [0274] ● 什么样的其它领域与此技术相关?
- [0275] ● 这些公司在该技术领域的投入多大?
- [0276] ● 这些专利对申请它们的公司有多重要?
- [0277] ● 这些公司已经放弃了哪个技术领域并且在哪个技术领域上继续投资?
- [0278] ● 什么公司 / 发明人正合作开发这种技术?
- [0279] ● 哪些发明人已经更换了公司?
- [0280] ● 什么技术领域是第一次被跨接 (bridge) 的,哪些公司拥有正在跨接他们的专利?
- [0281] ● 我应该引用哪个专利作为我当前的专利申请的现有技术?
- [0282] ● 哪个专利可潜在地用于使我的专利和我的竞争者的专利无效?
- [0283] ● 哪个公司最有可能违反我的专利?
- [0284] ● 哪个公司最有可能对授予我的专利许可感兴趣?
- [0285] ● 学术研究多快可转化为可取得专利的技术?
- [0286] ● 这个领域内的哪些技术将要成为专利?
- [0287] ● 这些专利的公司 / 发明人在建立该项技术和扩展他们的专利保护上有多积极?
- [0288] ● 公司在哪些技术上增加他们的投资以及在哪些技术上放弃他们的投资?
- [0289] ● 在我的业内正在投资什么技术?
- [0290] ● 什么行业正在利用这种技术?
- [0291] 专利网络的可视化

[0292] 上述的网络可视化系统能够容易地应用到专利数据库以产生巨大的影响。世界上的专利数据库特别可按照这种类型的分析处理,因为他们包含提供了专利之间的自然链接信息的引用信息。专利环境中的 NVS 的值是特别相关的,因为其使得除了专利代理人和 R&D 工程师之外的感兴趣的各方都能够使用专利数据。

[0293] 专利数据可从包括世界上的各个专利局 (USPTO、EPO 等) 多 个来源以及从类似 Thomson(包括其客户和 acquirees, Aureka, Micropatent, IHI 和 Delphion) 和 Lexis Nexis 的专利数据提供者得到。这种专利数据富含可被转换为网络数据的信息。能够被转换为节点和链接的这种类型的数据的一些实例包括但不限于 :

[0294] ● 节点 / 元节点 - 专利,发明人,受让人,IPC 分类,US 分类,Derwent 分类,优先权 / 申请 / 公开 / 授权 / 期满年份,语义群集,状态 (申请 / 授权 / 过期 / 放弃),审查人,发明人城市 / 州 / 国家,受让人城市 / 州 / 国家,申请的管辖权 (US/EPO/WIPO 等),优先权号,以及其他的信息 ;

[0295] ● 链接 / 元链接 - 引用,共同引用,著录连结,共同 IPC/US/Derwent 分类,共同优先权 / 申请 / 公开 / 授权 / 期满年份,共同语义群集,共同状态 (申请 / 授权 / 过期 / 放弃),共同审查人,共同发明人城市 / 州 / 国家,共同受让人城市 / 州 / 国家,共同申请管辖权,共同专利权号,以及其他的信息。

[0296] 可以在 NVS 之围内使用这些节点 / 元节点和链接 / 元链接定义以及上述的范围和组合的任意组合以审查专利数据的集合。

[0297] 此外,可以估计这些节点和链接的大小以如上所述为用户提供附加的信息。专利环境中能够用于估计节点和链接大小的一些特别有用的属性包括:

[0298] ●节点 / 元节点大小估计 - 在专利环境中,有几个特别的度量与节点 / 元节点大小估计有关。一些例子包括,可基于如下的度量估计元节点的大小:专利数,优先权数(例如,唯一同族专利数),专利被引用的次数(前向引用),由所表示的专利引用的专利数(后向引用),总引用(前向加上后向),自公开 / 授权以来的引用 / 年份,剩余的专利年份(例如,所表示的专利所剩余的年度之和),每专利的平均引用,平均专利年代,IPC/US/Derwent分类的平均 / 总数,发明人数,以及许多其它的属性度量。

[0299] 如之前所提及,也可基于任何数量的类似所表示的专利的中心性 / 本征向量中心性 / 中间中心性之和的网络统计计算估计节点和 元节点大小。基于这些各种度量估计节点大小的能力使得用户能够作出有关类似专利价值、技术革新的多样性、发明人集中度、以及用户对专利数据感兴趣的其它重要度量标准的判断。

[0300] 专利价值的度量标准是特别感兴趣的,并且专利数据(或者可以被链接到的其它的外生数据)之内有各种各样的可以给出专利价值的信号的属性。一些特定的例子包括但不限于:被引用的文档,引用的文档,学术引用的数量,最近的引用年代,中心性 / 本征向量中心性 / 中间中心性,专利说明书的长度,权项数,独立权限数,最短独立权限长度,覆盖宽度(申请的国家),维持费用支付,授权后异议(欧洲),维持费用支付,许可,专利诉讼,就受让人来说的 R&D 成本 / 专利,业内平均 R&D 成本 / 专利。这些度量标准的一些或者全部可以利用一个加权平均聚集到一起以在网络中提供专利价值的信号。可以合计这些值以提供专利文件夹的值的估计,并且这个值可以被用于估计表示专利文件夹的专利节点和元节点的大小。这可为用户提供有关哪个专利和文件夹在感兴趣的特定领域内是最为重要的极大的洞察。

[0301] ●链接 / 元链接大小估计 - 链接和元链接也可基于各种属性以与节点大小估计相同的方式估计大小。链接强度可以基于各种不同连接属性。一些例子包括引用数,被引用的唯一专利数,专利引用数,平均引用年代,最近引用年代,以及其它属性。

[0302] 专利和专利文件夹之间的链接可告知两件重要的事情:相关性和相似性。相关性是一个专利或专利文件夹是如何相互依赖或者是如何由另一专利或专利文件夹构建的度量标准。告知相关性的度量标准提供了有关潜在的侵权并因此在专利分析中是关键性重要的重要信号。一些告知相关性的度量包括引用次数,引用次数减去被应用的次数,由引用了你引用的同一个专利引用的次数,等等。

[0303] 相似性是专利分析中的另一重要的链接属性。两个专利文件夹之间的相似性暗示两个公司的 R&D 计划之间的接近平行以及或许是冗余。战略上,较高程度的相似性暗示潜在合资或者潜在的某 些其它程度上的成本共享。告知两个专利之间的相似性的度量标准包括总相互引用,结构上的等同物(一个网络分析术语,意思是它们在网络之内持有相同的构造位置),共同引用,著录连结,学术相似性,等等。

[0304] 上面描述的 NVS 的所有特征都与包括网络变换、多节点和链接的使用、分形网络、网络动画、统计信息以及链接到外部数据源的专利数据的分析有关。下面描述了特别用于专利数据分析的 NVS 的优选实施例的某些元素。

[0305] 识别未转让的专利

[0306] 与利用和链接 NVS 中的专利数据有关的一个唯一的挑战是专利申请典型地没有受让人与它们关联。由于其意味着数据库中的最新专利（最前沿的专利）不能够容易地被公司识别，故此是很不幸的。这些专利典型地还不包含引用，并且因为它们是全新的而没有被引用。专利数据的 NVS 的实现的一方面是通过创建适当地将它们连接到网络的可选链接来解决这个问题。这是通过比较未转让的专利和专利申请的属性以作出有关哪个公司已提交该专利申请的“最佳猜测”。使得这一比较变得可能的几个属性包括发明人姓名，发明人地址，IPC/USPC 分类，被引用的专利，执行处理的法律事务所，语义数据以及其它属性。

[0307] 通过比较搜索结果中的未转让专利和其它专利之间的这些字段，有可能根据同一的公司创建显示数据库中的哪一个其它的专利是最为可能的链接。作为一个例子，考虑一个未转让专利，其具有地址相同的三个共同发明人并在同一 IPC 分类中提出申请并且由同一个法律事务所执行处理作为数据库中的另一个专利。这些专利极可能是由同一个受让人提出的。

[0308] 系统回顾所有的未转让申请并在每个专利和数据库中最为相关的其它专利之间创建链接。每个相关性引用可被给予一个分数并使用加权平均以确定两个文档的总相关性的。用户于是可以选择将相似性 超过所选择的阈值的专利“转让”给该高度相关度的文档的受让人。或者，用户可以回顾每个链接并选择接受和拒绝所提议的“转让”。这些转让在 NVS 之内被标记为“计算机转让的”以便用户能够讲述有关那些专利实际上是否被转让给该特定受让人仍然有某种不确定性。这些在未转让的专利和最高度相关的专利之间创建的链接是可按用户判断而打开和关闭的不同链接类型。一种采用这些链接的特别有用的方式是可视化与未转让专利的网络结合的受让人网络图。这允许用户在单个网络视图中根据公司回顾所有的“计算机转让的”专利。

[0309] 统计信息

[0310] 有多种类型的统计信息与用于专利数据分析有关，包括：

[0311] ●受让人 - 根据受让人由最高至最低排序的所选择的网络中的专利号。

[0312] ●发明人 - 根据发明人受让人由最高至最低排序的所选择的网络中的专利号。

[0313] ●分类 - 根据分类码受让人或根据分类类别由最高至最低排序的所选择的网络中的专利号。数据可被提供给包括 IPC、USPC、Derwent 分类以及其它分类的几种分类方案的每一种。由于许多分类方案是分层结构的，可利用树形结构显示数据，每个分类之内有专利号，并且在树的每个分支并排地显示子类别。

[0314] ●单词使用 - 专利号包含关键字，短语或单词分组。存在几种识别文档集内的公共单词使用的工具。它们包括 Micropatent 的 Thescape 产品，Vivisimo 的群集工具，Grokker 的群集工具以及其它工具。这些单词群集工具可以容易地被结合到系统中以提供对所考虑的专利数据集的附加的理解。

[0315] ●引用 - 可以提供有关引用的各种类型的信息。它们包括但不限于如下信息：

[0316] ○最频繁被引用的专利、受让人、发明人、或者其它的专利分组；

[0317] ○自专利、受让人、发明人、或者其它的专利分组颁发证书以来每年的最高引用数量。

[0318] 还在弹出式窗口中提供统计信息。在弹出式窗口中提供的此类信息的几个例子特别是与专利数据有关，如下面的描述：

[0319] ●专利节点弹出 - 当选择了一个专利节点,可唤醒一个弹出式窗口,其显示有关由该所选择的节点表示的专利的信息。所提供的信息包括在典型的专利的首页面上提供所有基本信息,包括专利号、标题、发明人、受让人、申请号、优先权 / 申请 / 公开 / 授权日期、IPC/USPC 分类、研究领域、引用(专利或非专利二者)、审查人和代理人以及来自专利的其它数据,类似页数、权项数(独立的和相关的)、图表数、最短独立权项中的单词数等。此外,弹出窗口中的许多字段都是超链接的,允许用户带出附加的信息。例如,专利号是到专利的全文(或.pdf文件)的超链接,引用链接是超链接(链接到唤起对所引用文档的因特网搜索的非专利引用),以及其它超链接。弹出窗口还可包括有关专利的各种统计信息(例如,中心性)以及来自外部源的包括法律状态、诉讼状态、许可状态、专利族内的其它专利、过期授权异议、文件包装信息等的其它信息。

[0320] ●受让人元节点 - 当选择了一个受让人元节点,一个弹出窗口被唤醒,其显示了可显示有关由该元节点表示的受让人和专利的不同类型的数据的菜单。菜单选项包括表格,各表格示出了由该元节点表示的专利、根据 IPC 分类的专利、根据 USPC 分类的专利、根据发明人的专利的列表,以及根据年度显示各专利的图。附加菜单选项包括可显示有关受让人元节点的网络统计信息,包括总引用、(自发布年度以来)每年度的平均引用、受让人的文件夹的本征向量中心性之和/整个网络的本征向量中心性之和(文件夹价值的度量)。另一个菜单选项提供了有关受让人的信息。这个菜单选项链接到基本公司以及有关该公司的财务情况。可以使用的这种类型的信息的各种来源包括 Hoovers (www.hoovers.com) , Bloomberg (www.bloomberg.com) , Yahoo Financial (<http://finance.yahoo.com/>) 以及包括包含公司简介 信息的公共和专有站点的许多其它的来源。

[0321] ●发明人元节点 - 当选择了一个发明人元节点,一个弹出窗口被唤醒,其显示了可显示有关由该元节点表示的发明人和专利的不同类型的数据的菜单。菜单选项包括表格,各表格示出了由该元节点表示的专利、根据共同发明人的专利、根据 IPC/USPC 分类的专利的列表,以及根据年度显示各专利的图。另一个菜单选项提供了有关发明人的信息。这个菜单选项链接到两个不同类型的信息,一个是针对发明人姓名的基本 web 搜索,第二个是来自万维网的“人员查找程序”信息。人员查找程序站点有诸如 people.yahoo.com/ , www.zabasearch.com , www.intelius.com , www.peoplefinders.com , 以及许多其它的站点可以基于可直接在专利信息中找到个人姓名、城市和州提供住址历史,生日,结婚 / 离婚 / 死亡信息,房地产记录,留置权和抵押,破产,兵役,亲戚,邻居,信用帐单和背景情况检查。当需求出现时这种信息是在寻找发明人时有用。其也可用于识别数据库之内的可能表示同一个人的发明人姓名。

[0322] ● IPC/USPC 元节点 - 当选择了一个 IPC/USPC 元节点,一个弹出窗口被唤醒,其显示了可显示有关由该元节点表示的 IPC/USPC 和专利的不同类型的数据的菜单。菜单选项包括表格,各表格示出了由该元节点表示的专利、根据受让人分类的专利、根据发明人的专利的列表,以及根据年度显示各专利的图。另一个菜单选项提供了有关受让人的信息。这个菜单选项提供了有关 IPC/USPC 分类的详细信息,包括该分类的完整描述及其在 IPC/USPC 分类层次结构中的位置,显示所选择的 IPC 分类是如何与 USPC 分类关联的重要词汇索引信息(或反之亦然),显示所选择的 IPC/USPC 分类与使用行业和制造行业有关的 SIC/NAICS 之间的链接的重要词汇索引信息。(稍后在这个实施例的描述中详细讨论 IPC/USPC 到 SIC/

NAICS 的链接)

[0323] ●元链接 - 当选择了一个元链接,一个弹出窗口被唤醒,其显示了有关由该元链接表示的连接的信息。可显示一个表格,其示出了随时间的过去由元链接表示的专利到专利链接的列表以及个别链接数 的图。如果例如该元链接是共同发明人关系链接,则该元链接弹出窗口将显示这两个发明人之间的合作历史。如果该元链接是受让人 - 受让人引用链接,则该弹出窗口将显示该两个受让人之间的引用历史。

[0324] 所有的这种弹出信息使得用户可能以所期望的任意详细程度从高级别的元数据到有关公司、发明人、技术和专利细节的最深度的级别来探究专利。这使得专利网络可视化工具成为理解大型专利文件集合的强有力的工具。

[0325] 链接到外部数据源

[0326] 下面描述特别是与专利数据有关的有用的外生数据源的一些例子,以及它们在专利网络可视化系统中的使用 :

[0327] ●行业数据 - 有关利用专利数据的一个关键的观察是对于除了专利代理人和 R&D 工程师以外的大多数决策者而言,感兴趣的实体并非专利。相反,用户典型地对了解有关公司、技术、发明人、或一些其它实体感兴趣。一个许多用户可能更好理解的特别感兴趣的实体是行业信息。用户经常想要知道类似下面的问题的答案 :

[0328] ○什么技术是这个行业的关键技术?

[0329] ○什么公司在这个行业中正引领技术开发?

[0330] ○什么行业利用了特定的技术?

[0331] 不幸的是,行业数据并没有直接附加到专利数据库中的专利。然而,可能以两种方式将专利数据库中的数据链接到行业数据。第一,专利数据库之内的受让人 / 公司可以被链接到它们参与的行业。全世界的各国政府已经作了各种尝试以开发它们的经济范围内的标准化行业分类法。其结果是类似 SIC(标准行业分类) 代码和 NAICS(北美行业分类系统) 代码。这些代码根据它们参与其中的行业分类各公司。

[0332] 各种各样的数据库包含有带有有关它们的 SIC/NAICS 行业的信息的公司目录。一个例子是全球商业目录 (www.siccode.com),其收容有各公司和它们的行业的数据库。

[0333] 通过链接专利数据库中的受让人到他们的 SIC/NAICS 代码,就 可能在网络可视化之内创建元节点,显示审查之中的专利数据中所表示的各行业。利用上述的特征,接着可能检查各行业之间和之中的关系,以及各行业、公司、技术、发明人、国家、以及专利数据内表示的其它各实体之间的关系。

[0334] 另一个通过其专利数据能够被链接到行业数据的工具是通过 US 专利分类 (USPC) 代码或国际专利分类 (IPC) 代码。这是使得可能通过各种“技术 - 行业重要词汇索引”。从 1990-1993 期间,加拿大专利局与 Statistics Canada 合作将所有的新专利申请既转让给了 SIC 使用又转让给了 SIC 制造。对总共大约 148,000 件申请作了这种转让。这个信息已经被各个政府实体和学术研究人员用于描绘 IPC/USPC 技术分类和 SIC/NAICS 行业分类之间的对应关系。

[0335] 已经建立了公开可用的显示了各行业和技术之间的连接的表格。这些表格的各种版本可从万维网上的以下来源获得,包括 :

[0336] ○ 来 自 <http://www.rotman.utoronto.ca/~silverman/ipssic/>

documentation_IPC-SIC_concordance.htm 的 PC-US-SIC 重要词汇索引

[0337] ○ OECD 技术重要词汇索引

[0338] [http://www.olis.oecd.org/olis/2002doc.nsf/linkto/dsti-doc\(2002\)5](http://www.olis.oecd.org/olis/2002doc.nsf/linkto/dsti-doc(2002)5)

[0339] ○ Yale 技术重要词汇索引

[0340] <http://faculty1.coloradocollege.edu/~dhohnson/jeps.html>

[0341] ○ USPC 到 SIC 重要词汇索引

[0342] <http://www.uspto.gov/web/offices/ac/ido/oeip/catalog/products/tafresh1.htm#USPC-SIC>

[0343] 利用这些表格,就可能将技术分类信息链接到行业分类。这使得专利网络可视化的用户可能分析有关行业信息作为他们的专利数据分析的一部分。

[0344] 也可同时使用该两个行业数据源。例如,系统可以创建 SIC 或 NAICS 元节点并在有关于数据库中的受让人的同一图形(或作为分形节点)之内显示它们。与此同时,技术 - 行业重要词汇索引数据可以被用于创建链接到那些行业元节点的 IPC 或 USPC 网络。通过这样做,就可能确定一个在该行业中公司正采用特定技术的合理的可信度。这种能力致力于专利数据研究中的公共问题,即,“对于具有与此技术相关的专利的不同的公司,其中哪一个正在采用我的行业中的技术,以及哪一个正在采用其它行业中的技术?”

[0345] ● 法律状态数据 - 与专利数据有关的有价值的外生数据的第二来源是诸如 INPADOC 的法律状态数据库。这种数据库,连同其它的数据库一道,包含有关维持费用、转让、过期授权异议等的数据。到这种数据的连接技术上简单作为可以直接链接到数据库的专利号或优先权号。到这种数据的链接的价值非常高。专利网络可视化系统可使用这种数据以识别已经被放弃(由于维持费用支付的缺乏)的专利,并重新转让。这通过显示公司的优先权在何处提供了有关法人优先权的强烈的信号。这是通过改变专利节点、受让人元节点、IPC/USPC 元节点、发明人元节点或其它节点的外观以显示哪个专利已经被放弃来实现的。外观的改变可能包括改变颜色、形状、大小、边框类型或填充类型。

[0346] 另外,法律状态数据可被用于显示哪个专利已经被转手了,给出商业或商业单位的获取或处理的证据,并给出专利价值的信号。例如,大量的专利被重新转让给一个新的公司很可能发出法人机构在改变的信号。另一个例子是专利已经在法律程序中反对更可能是有价值的专利,因为当事人可能不太可能继续一个反对,除非有显著的经济刺激去这么做。再次,这个信息可被用于改变节点或元节点的外观以向用户发出有关该专利数据网络的重要线索的信号。

[0347] ● 文件包装数据 - 另一种值得链接到的数据源是专利局文件包装数据。在美国,可在线在 www.USPTO.gov 上找到该数据。由于其包含可直接链接到专利数据库的优先权日期或专利号,这种数据在技术上易于链接。这个信息的使用是事半功倍的。首先,其对用户能够“点击通过”他们特别感兴趣的专利的文件包装数据是有用的。其次,文件包装数据提供了有关专利价值和有效性的线索。专利局审查决定书编号、权项驳回、从申请到授权后专利的权项数的改变、答复专利局审查决定书的时间以及在文件包装中发现的其它信息都提供了对专利网络可视化用户有用的信息。可以改变节点和元节点的外观以向用户发出任何这些文件包装参数的存在或价值。

[0348] ● 法律数据 - 有关专利的另一重要信息来源是相关的法律程序的法律状态。有关

专利诉讼的存在和状态的信息对理解专利前景是关键性的。通过结合上述的专利网络可视化系统的能力利用这个信息,可回答重要的问题,如 :

- [0349] ○什么专利经过审查程序后已经有效?
- [0350] ○什么专利当前正被提出诉讼?
- [0351] ○什么公司被控告专利侵权,以及他们面临多少的诉讼案?
- [0352] ○在我的行业中谁积极地申明专利?
- [0353] ○什么技术最为活跃地纠缠于专利诉讼中?
- [0354] ●许可证数据 - 有关专利的信息的另一重要来源是许可证数据。有关专利许可证的信息提供了有关单个专利和专利文件夹的价值的信号。现存的各种许可证数据库包括 www.yet2.com、www.royaltystat.com、www.royaltsource.com、IP 事务处理数据库 (www.fvgi.com)、IP 研究联合数据库 (www.ipresearch.com)、许可证版税率 (www.aspenpublishers.com)。根据专利号、根据公司、根据行业 (SIC/NAICS) 或根据其它手段可构成链接。通过结合上述的专利网络可视化系统的能力利用这个信息,可以回答重要的问题,如 :
- [0355] ○什么专利可有效的用于许可证交易?
- [0356] ○在这个行业内与专利有关的典型版税率是什么?
- [0357] ○这个公司发给它的专利颁发许可证了吗?
- [0358] ○我的竞争者许可证进 / 出的技术是什么?
- [0359] ●法人数据 - 法人数据是另一种可以结合到专利数据可视化系统内的外生数据源。通过专利数据库中受让人字段的方式可生成到公司数据的链接。无数的来源存在各种类型的法人数据。特别有用于链接到的数据类型的实例包括财务数据,以及产品数据。
- [0360] 法人财务数据的各种来源存在于从如 SEC 的 EDGAR (<http://www.sec.gov/edgar.shtml>) 的政府系统到如 Hoovers (<http://www.hoovers.com>) 和 Bloomberg (www.bloomberg.com) 的数据聚集器的范围中。类似销售、R&D 开销、市场容量 (marketcap) 和许多其它的财务信息可用于带给对专利数据的进一步理解。例如,年度 R&D 开销可除以每年的专利申请数 (具有时间滞后) 以比较相对 R&D 效率。销售除以专利数可用作将来的收入流中的法人投资的信号。市场容量除以专利数 (或者文件夹值的估计) 以发出将可能获得专利文件夹多么昂贵或便宜。这些和许多其它度量的比较可以提供有关各公司的相对表现的洞察以及他们的专利文件夹的重要性、价值和强度。这些度量可被结合到专利网络可视化系统中作为属性,用于估计相关节点和元节点的大小,或相反用于改变节点或链接的外观以便通知用户有关他们的研究的重要信息。
- [0361] 产品信息是能够为专利网络可视化系统的用户提供进一步洞察的重要的公司信息的另一来源。许多公司都有在线产品目录。这个信息常常包含通过针对在专利说明书中找到的关键术语搜索产品数据库可以链接到专利数据的技术信息。该系统可连同来自专利说明书的关键字一起使用受让人信息以创建到公司产品目录上的产品数据的链接。这些外部链接可被显示为网络图中的节点并可以允许用户了解各公司是否、以及如何应用他们已取得专利的技术。
- [0362] ●学术数据 - 另一可以告知用户对专利数据库的研究的外生数据源是学术数据。学术数据包括来自学术和行业期刊、会议论文集、科研授权、以及其它来源的信息。在专利

被颁发证书之前这种信息典型地存在于公共领域。因此其充当对于许多专利数据的用户来说是重要的刀片 (cutting edge) 研究的标杆。链接可以在专利数据和学术数据之间以各种方式建立。首先, 专利常常引用学术或行业期刊作为现有技术。其次, 专利上列表的发明人通常首先在学术文献中发表他们的研究, 因而可以通过连接发明人姓名建立链接。最后, 学术文献也可以通过被转让了该专利并且一个公布与他们有关系的机构或公司的方式连接到专利。有关围绕以专利主体的学术文献的信息可用于理解在该领域内正在进行的基础研究的来源, 识别行业和学术之间的协作, 识别它们在专利数据中出现之前潜在的突破技术, 识别已经自学术环境中出现的异军突起的公司, 以及其它的理解。

[0363] 通过以上述的方式链接到学术数据, 就可能使用所有的网络可视化系统的能力创建学术文献和专利数据的组合网络。下面详细描述的网络可视化的第二个实施例, 讨论了网络可视化系统用于可视化学术文献。此二者的组合在用户理解之前已经出现了什么之上, 人们、公司、行业和地理之中的技术发展和网络革新的能力方面提供巨大突破。

[0364] 根据专利局或商业专利数据厂商的实现

[0365] 本发明人已经发现的 NVS 的本实施例在理解专利数据方面令人难以置信的有用。USPTO、EPO 和其它专利局以及诸如 Aureka、Micropatent、Delphion (现在全部归 Thomson 所有)、Lexis Nexis、以及其它的商业专利数据厂商有巨大的专利信息数据库。不同于上面在现有技术部分中描述的一些简单的分析工具, 商业专利数据厂商的这些专利局没有任何一个为其顾客提供了用于专利数据分析的复杂工具。在此描述的实施例, 或者全部地、或者更可能以基本的网络可视化能力的简单版本, 将为他们的顾客提供非常强大的前端用于访问他们的专利数据库。

[0366] Thomson 或者 Lexis Nexis 能够选择采用非常简化的 NVS 实现作为他们的专利数据库的用户界面。该简化的实现可能允许用户使用逻辑搜索来搜索数据库然后返回一个文档列表作为他们今天所做的工作。他们接着能够使用该 NVS 将那些搜索结果转换为网络数据并且允许用户从搜索结果的一个有限的网络可视化集合中选择。该系统可以为用户提供选项以从下列选项中选择一个或多个:

- [0367] ● 引用网络 - 根据引用链接的结果集中的专利网络;
- [0368] ● 受让人网络 - 根据引用链接的结果集中的受让人网络;
- [0369] ● 发明人网络 - 根据共同发明人关系链接的发明人网络;
- [0370] ● IPC 或 USPC 网络 - 根据分配给多个分类的专利链接的 IPC 或 USPC 分类网络;
- [0371] ● 受让人 / 发明人网络 - 根据引用链接的受让人网络, 有基于他们将他们的发明转让给其的公司链接到受让人节点的发明人;
- [0372] ● 受让人 / IPC 或 USPC 网络 - 根据引用链接的受让人网络, 有基于该公司在 IPC 或 USPC 分类之内提出申请的专利号链接到受让人节点的 IPC 或 USPC 分类。

[0373] 他们可能想要包括的附加的特征可能是过滤结果集以限制可视化中的记录的能力。过滤选项应该包括利用特定的日期范围、受让人、IPC 或 USPC 分类、发明人从可视化中过滤掉记录的能力。

[0374] 虽然 NVS 的这个实现将缺乏在这个实施例中描述的许多特征, 但对于他们的用户在他们理解他们的专利搜索结果的能力方面可能是巨大的突破。其将允许他们从许多不同的观点审查他们的搜索结果, 通过 NVS 过滤能力精炼他们的搜索, 并最终直接通过 NVS 系统

审查专利列表或专利文档。

[0375] 还值得注意的是,该两个主要的商业专利数据厂商是收容有许多其它的数据源的大型组织的一部分。在本实施例中描述的 NVS 以及上面更为一般的描述均为所有的他们的各种数据类型提供了一个前端。同样,在本实施例描述的某些外生数据源实际上是通过根据该两个主要专利数据公司, Thomson 和 Lexis Nexis, 的专利使用权转让协定拥有的或提供的。NVS 作为前端用于访问他们的数据库的实现还是一种链接在他们的系统范围内的异类数据源的方式,将允许这些公司为他们的顾客提供高度适应个别差异的价值提议。这些和其它的数据厂商需要为他们的数据库的用户找到替换方式以从他们的数据中提取更多的价值以便增长和支持更高的价格和利润。NVS 能够为那些目标作出巨大贡献。

[0376] 实例 2 :用于搜索和分析医学出版数据库的方法和设备

[0377] 在第二个实例中,本发明的一个或多个实施例被应用于搜索和分析学术文献数据库中的文档。这种数据库的一个例子是被称为 PubMed 数据库的医学出版数据库。对其它学术数据库的应用同样是可能的。

[0378] PubMed 数据是出现在近 200 个医学期刊中的医学研究论文的大型数据库。其是有关医学世界中研究领域的信息内容丰富的资料档案库。PubMed 数据最为频繁地被医生或其它寻找有关特殊疾病、治疗方案或其它感兴趣的医学主题的医学专家所使用。他们的研究总是以对关键字、作者或期刊的逻辑搜索开始,其后他们被表示为匹配他们所选择的条件的论文列表。研究人员的下一步骤是扫描搜索结果列表并阅读标题和摘要直到找到引起她兴趣的文章。她然后继续阅读该篇文章的一些或所有内容。然后可以返回到她的搜索结果并继续扫描和阅读结果直到找到她正在寻找的信息。

[0379] 如果研究人员的目标是寻找类似于一篇或几篇论文中的信息,则这个方法对医学专家是有用的并且完美地适用。然而,存在无法容易地以这种方式回答的一类问题。我们称这些问题为元问题。它们是与那些文章所表示的元实体有关而非与文章中包含什么内容有关的问题。不是询问有关文章和文章中包含什么内容的问题,研究人员通常对如下的问题感兴趣 :

[0380] ● 基因治疗领域正在发生什么事情?

[0381] ● 在 Alzheimer 治疗方面谁是领先的研究者?

[0382] ● 什么机构在合作研究骨关节炎?

[0383] ● 各个癌症研究领域相互之间是如何相关联的?

[0384] ● 谁是纳米生物技术最有影响力的研究人员?

[0385] ● 在医学科学特定领域正在进行什么研究,如疾病小组(例如,Alzheimer 病)、疗法(例如,免疫疗法)、或行为的特殊机制(例如,抗斑形成)? 工作进度如何? 随时间过去是如何改变的?

[0386] ● 给定领域内的合作模式是什么? 涉及到谁? 他们是如何在一起工作的? 何处存在紧密合作的研究团体? 其在何处分段? 何处是按顺序连接各个研究组的最佳机会,例如,将科学发现转换为 MRSA 中的实践? 随时间过期合作模式是改进了、减弱了?

[0387] ● 给定领域内的智能结构是什么,即,其的课题趋于集中研究? 谁不是? 这是科学的原因或是制度和社会关系的失误? 频繁重复的强课题关系是什么? 将来会出现哪些关系?

[0388] ●一个领域的研究方式（即，随时间过去随其发展的智能结构）是如何影响出现的合作模式的？反之亦然。

[0389] ●特定公司或大学内的人们就有关特定课题是如何合作的？是在他们自己之间？与机构之外的其他人一起？

[0390] ●谁是给定治疗领域内最具影响力的医学科学家？谁是网络的最核心？什么医学科学家居以他们独立的影响模式最好地共同跨越网络？

[0391] 可以利用如上所述的 NVS 来回答这些以及许多其它的问题。也可以以与专利数据相同的方式分析和可视化学术数据和特定的 PubMed 数据。使得专利数据通过在此描述的方法是可分析的属性完全类似于在包括 PubMed 的学术数据库中找到的数据。图 20 示出了专利数据的属性与学术文献的属性之间的关系。

[0392] 如可明显看出的那样，在该两个数据源之间有直接的对应关系。这使得可能利用在前面的实施例中描述的相同的方法来分析 PubMed 数据（或者任何学术文献数据库）。然而，并不是每个专利属性都与学术数据直接类似。因此，为了特别与学术文献一同使用，需要稍微修改在专利网络可视化系统中描述的该方法。

[0393] 就像使用专利可视化系统一样，医学网络可视化系统允许用户创建和检查来自学术数据库的数据库记录作为节点和链接的网络。如同使用专利数据库，学术数据最初也没有结构化为网络信息。换言之，其并不包含节点列表和链接列表。在其能够被可视化为网络之前，首先必须结构化学术数据以便将其转换为网络信息。这是以与前面的实施例中描述的相同的方式实现的。

[0394] 一旦数据被结构化，用户就有可能基于来自 PubMed 数据库或其它学术数据库的数据库记录查看各种网络可视化。不同于现有技术的系统，医学网络可视化系统不需要网络节点和链接的稳定定义。相反，研究人员可根据她的兴趣动态改变节点和链接的定义。这种将网络从一个节点 / 链接定义变换为另一定义的能力，以及同时查看同一数据的多个连接的网络视图的能力使得用户可能快速并容易地理解大型数据库记录集合，并且回答通过其它方式不能够回答的元阶问题。

[0395] 如同专利数据一样，也可以以各种方式思考医学数据网络什么是节点和什么是链接的不同定义。下面描述了由 PubMed 数据可以创建的各种节点 / 元节点和链接 / 元链接的一些例子：

[0396] ●节点 / 元节点 - 学术数据环境中的节点 / 元节点定义包括但不限于文章、论文、授权、评论、作者、期刊、发表年度、评论者、作者城市 / 州 / 国家、机构城市 / 州 / 国家、期刊国别、Mesh 类别、以及其他信息。

[0397] ●链接 / 元链接 - 引用、共同引用、著录连结、公共 Mesh 类别、公共发表年度、公共语义群集、公共评论者、公共作者城市 / 州 / 国家、公共机构城市 / 州 / 国家、公共期刊国别、以及其他信息。

[0398] 也可在 NVS 之内使用这些节点 / 元节点和链接 / 元链接的任意组合以及上面的范围和组合以检查 PubMed 数据集。

[0399] 此外，可以估计这些节点和链接的大小以便如上所述为用户提供附加的信息。可用于估计医学数据环境中的节点和链接大小的一些特别有用属性包括：

[0400] ●节点 / 元节点大小估计 - 在医学数据环境中，有几种与节点 / 元节点大小估计

有关的特殊度量。一些例子包括：可根据文章的数量，文章被引用的次数（前向引用），文章被所表示的专利引用的次数（后向引用），总引用数（前向加上后向），自发表以来的引用 / 年份，每篇文章的平均引用，平均 / 总 MeSH 分类数，作者数，以及许多其它的属性度量估计元节点大小。

[0401] 如之前所提及，也可以基于任何数量的类似所表示的文章的中心性 / 本征向量中心性 / 中间中心性之和的网络统计计算估计节点和元节点大小。在医学研究环境中，这些度量告知研究基于对等的引用的重要程度如何。发现重要研究的能力对于当生物工艺学和制药以及其它生命科学公司试图继续保持位于研究的前沿以及接近将帮助他们引进下一鸣惊人的药品或高度有利可图的医疗装置的前沿研究是至关重要的。

[0402] ●链接 / 元链接大小估计 - 医学领域内一些用于链接 / 元链接大小估计的属性的例子包括引用数，被引用的唯一文章号，引用的文章号，平均引用年度，最近引用年度，以及其它属性。

[0403] 医学网络可视化系统的例子

[0404] 下面所示出的是示意利用用于执行搜索并分析特定医学文档的医学网络可视化系统 (MNVS) 生成的各种网络图的各种示例性屏幕快照。这些图尽管是基于有限的节点和连接组合的集合，它们还是揭示了可供网络可视化系统的用户使用的能力。对于这些简单的例子，节点和元节点二者均保持大小不变，而链接和元链接保持宽度不变。然而，如同专利网络可视化系统一样，也可以修改这些参数以便为用户提供更进一步的理解。

[0405] 图 21 示出了基于 PubMed 数据的典型网络，并利用在此描述的 MNVS 可视化。其是搜索自 2000 年 1 月至今成文的医学主题题目 (MeSH) 为 “Diabetes Mellitus Type I” 并且位于 “Boston” 的文档的结果。该工具检索到 124 篇文档以便创建网络图。任何人都可以按照上述的关键字解释图中的链接。

[0406] 该网络图显示了三种不同的元节点类型（作者元节点，期刊元节点，和 MeSH 元节点）。在医学网络可视化系统的一个优选实施例中，节点类型以不同颜色区分（作者 - 黄底黑字，期刊 - 绿底白字，和 MeSH - 白底蓝字）。尽管在黑白表示中很难看出这些颜色，它们看起来如下（作者 - 浅底深色字，期刊 - 深色底白字，MeSH - 白底深色字）。

[0407] MNVS 为用户提供了选择节点和链接定义作为她的工作的能力。由于图 22 示出了如图 21 的同一个网络的细节从而证实了这种能力，但是只显示了作者 - 作者链接，其揭示了科学团体的社会网络。根据这些类型的网络图，就有可能知道前沿研究人员处于特定的研究领域，他们与谁在合作以及哪个科学家是最具影响力的。

[0408] 医学期刊数据库的一个唯一元素是文章中的作者姓名顺序的重要性。基于访谈和我们对这种类型的分析的经验，我们已经知道医学期刊文章中的第一作者是此研究的主研究人员 (PI)。如果第二 IP 涉及到该研究（经常是这种情况），她的名字将出现在第二位。作者列表的最后一位是该研究在其内实施的实验室负责人。该“实验室负责人”可能仅轻微涉及到实际的研究项目，但很可能是该领域中的重要人物。位于作者列表的第二和最后一位之间的名字典型地是实验室助手以及其它对该篇文章有次要贡献的人员。

[0409] 为此，MNVS 允许用户选择在网络中包括哪个作者姓名。我们发现一个有用的设置是在网络中包括第一个、第二个以及最后一个作者姓名而排除所有其它的。

[0410] 图 23 和 24 进一步论证了 MNVS 的能力。在这里同一网络被变换为作者和期刊网

络（图 23）以及作者和 MeSH 分类的网络（图 24）。这些网络使得用户能够快速了解网络内研究人员感兴趣的研究领域。

[0411] 最后，图 25 再次示出了同一个网络，然而这次其显示了根据 MeSH 分类命名的医学专题领域之间的链接网络。以这种方式使用 MNVS，医学专家经常惊讶于发现表面上看上去不相关的两个医学领域之间的意料之外的联系。有关医学主题之间的意料之外的连接可导致新的思考医学问题的方式并且由于它们提供了潜在的应用在一个研究领域内的发现物挑战另一研究领域而暗示有新的研究路径。由于领域 A 的专家和领域 B 的专家不会出席同一会议、参与同一高级训练计划、阅读同一期刊或另外相互作用，所以他们很少有往来，因此医学专家趋于根据专业特性 silo-ed。将来自不同专业的适当的人放在一起有极大的价值，因为经常会提出完全新的研究路径。医学网络可视化系统使得可能同时发现意料之外的连接区域，由该区域可能出现新的医学洞察。

[0412] 医学 / 学术数据的网络可视化应用

[0413] 根据医学网络可视化系统允许的能力给予研究人员以深入理解分析和开发巨大医学数据库记录集合的能力。这些理解以各种形式出现并且因此该网络可视化系统可被用于各种环境中用于分析主题，如：

[0414] ● 合作的组织（一般而言）；

[0415] ● 公司范围内的合作组织；

[0416] ● 谁被视为某地域内临床试验或市场影响的关键意见领导 (KOL) 或关键研究人员；

[0417] ● 特定领域内群集的主题（其伴随 MeSH 分类）；

[0418] ● 跨组织的研究协同或代替；

[0419] ● 更广地域内研究强度的区域基础。

[0420] 一般而言的合作组织

[0421] 在基本级别，图 26 的网络示出了合作的不同“群”。用户能够很容易地识别在不同期刊中一起发表的作者组。该系统的另一个特征在于基于作者的组织从属关系将作者元节点以颜色区分。这提供了对合作模式的更深入的理解。

[0422] 公司范围内的合作组织

[0423] 图 27 中的网络是搜索 MeSH 分类 Diabetes Mellitus 和机构 Joslin（机构被发现作为 PubMed 中地址字段的一部分）。Joslin 是 Joslin Diabetes Center – 世界领先的糖尿病中心 – 的简写。该图识别合作卡片匣 – 该组织内就特定题目著作的共同作者的人员。该图以全文显示搜索中出现在五个或更多的文档内的期刊名称和 MeSH 术语。这使得 用户能够看到诸如 Diabetic 视网膜病和胰岛的流行的研究题目，以及这个组织自 2000 年以来已经发布的包括糖尿病、糖尿病学、移植、糖尿病照料以及其它的期刊。

[0424] 某地域内临床试验或市场影响的目标关键意见领导 (KOL) 或关键研究人员

[0425] 在图 28 中，糖尿病 mellitus 搜索受地域（澳大利亚）而不是机构的限制。在此，该网络图被限制为仅显示已撰写了 15 篇或更多篇文档的那些作者。Cooper ME 是名字在 49 篇文档中出现的作者 – 符合进一步调查的条件，如果制药或生物工艺学公司正在销售糖尿病药品的话，Dr. Cooper 可能就是该公司打算想要选择的澳大利亚的关键意见领导。

[0426] 特定领域内群集的主题（其伴随 MeSH 分类）

[0427] 图 29 示出了受组织限制的搜索产生的网络,组织被高亮以便用户还能够看见相关的 MeSH 分类。在此删除了所有其它的节点和链接,留下的揭示了共有 MeSH 分类的文档。例如,在这个例子中具有 MeSH 分类 Diabetes Mellitus、类型 II 的文档还编码为 Obesity、Islets of Langerhans、Blood Glucose、Insulin 以及其它。

[0428] 跨组织的研究协同或代替

[0429] 图 30 所示的网络由对 MeSH 分类 Cardiovascular Agents 和三个特定的组织的搜索生成。下列的组织或他们的组合已经使用颜色查询特征被高亮 :Pharmacia(黄色), Pfizer(绿色)、Warner Lambert(蓝色)、以及 Pfizer 和 Pharmacia 的组合 (紫色)。这个图使得用户能够了解在较大领域内组织的研究归入哪一个 MeSH 主题之下。这可能帮助组织思考有关某个研究计划中的资源投资的战略,特定研究领域内的竞争,和 / 或正出现的他们未曾涉及过的领域。

[0430] 更广地域内研究强度的区域基础

[0431] MNVS 还能够帮助揭示跨地域的研究“中心 (hubs) ”。在图 31 中,搜索被设计为高亮 MeSH 分类 Diabetes Mellitus 类型 I 的研究,麻萨诸塞州 (颜色编码为绿色), 加利福尼亚 (颜色编码为粉红色), 以及北卡罗来纳州 (颜色编码为蓝色)。用户可以不带地域限制运行类似的搜索并探究该数据以了解哪个区域看上去将显现为“中心”。另外,用户也许能够识别集中于广阔区域内的较小的适当位置的地域 (例如,在这个糖尿病研究例子中为 Autoantigens) 。

[0432] 如在这些例子中所证明的,上述的所有 NVS 的特征与 PubMed 数据的分析相关,包括网络变换、使用多节点和链接、分形网络、网络动画、统计信息以及链接到外部数据源。特别针对 PubMed 数据分析的 NVS 的优选实施例的一些元素在下面进一步描述。

[0433] 统计信息

[0434] 与 PubMed 数据分析相关的各种类型的统计信息包括 :

[0435] ● 机构 - 根据机构由高至低排序所选择的网络中的文章号。

[0436] ● 作者 - 根据作者由高至低排序所选择的网络中的文章号。

[0437] ● 分类 - 根据 MeSH 分类由高至低排序所选择的网络中的文章号或者根据分类类别排序。由于 MeSH 分类方案是分层结构的,数据是利用树形结构显示的,在该树的每个分支的旁边显示每个类别和子类别内的文章号。

[0438] ● 单词使用 - 文章号包含关键字、短语或单词分组。

[0439] ● 引用 - 可以提供多种类型的有关引用的信息。包括但不局限于下列信息 :

[0440] ○ 最为频繁被引用的文章、机构、作者、或其它分组 ;

[0441] ○ 自文章、机构、作者或其它分组发表以来每年的最高引用数量。

[0442] 还在弹出窗口中提供统计信息。在弹出窗口中提供的特别是与 PubMed 数据有关的信息类型的几个例子描述如下 :

[0443] ● 文章节点弹出 - 当一个文章节点被选择,可唤醒一个弹出窗口,其显示有关所选择的节点代表的文章的信息。提供的信息包括在典型的文章的摘要页面上提供的基本信息,包括 PubMed ID 号、标题、作者、机构、发表日期、MeSH 分类、引用、以及来自文章的其它数据,类似页数、图表数、字数等。此外,弹出窗口中的许多字段都是超链接的,允许用户带出附加的信息。例如,文章号是到该文章的全文 (或 .pdf) 的超链接,引用链接是超链接,

以及其它超链接。弹出窗口还可包括有关文章的类似中心性的统计信息。

[0444] ● 机构元节点 - 当选择了一个机构元节点, 可唤醒一个弹出窗口, 其显示一个不同数据类型的菜单, 其可有关于由该元节点表示的机构和文章而显示。菜单选项包括表格, 各表格示出了由该元节点表示的文章、根据 MeSH 类别的文章、根据作者的文章的列表, 以及根据年度显示各文章的图。附加菜单选项包括可显示有关机构元节点的网络统计信息, 包括总引用、(自发表年度以来) 每年度的平均引用、机构的文件夹的本征向量中心性之和 / 整个网络的本征向量中心性之和 (文件夹价值的度量)。另一个菜单选项提供了有关机构的信息。这个菜单选项链接到机构的网站或链接到有关该公司的基本公司和财务信息。

[0445] ● 作者元节点 - 当选择了一个作者元节点, 可以唤醒一个弹出窗口, 其显示了可显示关于由该元节点表示的作者和文章的不同类型的数据的菜单。菜单选项包括表格, 各表示示出了由该元节点表示的文章、根据共同作者的文章、根据 MeSH 分类的文章的类别, 以及根据年度显示各文章的图。

[0446] ● MeSH 元节点 - 当选择了一个 MeSH 元节点, 可唤醒一个弹出窗口, 其显示了可显示有关由该元节点表示的 MeSH 分类和文章的不同类型的数据的菜单。菜单选项包括表格, 各表格示出了由该元节点表示的文章、根据机构的文章、根据作者的文章的列表, 以及根据年度示出各文章的图。另一个菜单选项提供有关 MeSH 分类的信息。该菜单选项提供了关于 MeSH 类别的详细信息, 包括该类别的完整描述 以及其在 MeSH 分层结构中的位置。

[0447] ● 元链接 - 当选择了一个元链接, 可唤醒一个弹出窗口, 其显示了有关该元链接表示的连接的信息。可显示一个表格, 其示出了随时间的过去由元链接表示的文章到文章链接的列表以及由该元链接表示的个别链接数的图。如果例如该元链接是共同作者链接, 则该元链接弹出窗口将显示这两个作者之间的合作历史。如果该元链接是机构 - 机构引用链接, 则该弹出窗口将显示这两个机构之间的引用历史。

[0448] 所有这些弹出信息都使得用户可能以所期望的任意详细程度从高级元数据一直到最深级别的有关机构、作者、研究领域和文章的细节来探究文章网络。这使得 MVS 工具成为用于理解大型 PubMed 文档集合的强有力的工具。

[0449] 链接到外部数据源

[0450] 下面描述特别是与医学数据有关的有用外生数据源的一些实例, 以及它们在 MNVS 中的使用 :

[0451] ● 医生联系数据 - 链接以与 PubMed 数据库相关联的一个有价值的外部数据源是关于研究人员的从属关系的信息。PubMed 中的大多数作者都是医生, 因此有可能在公共和专有数据库中链接到有关于这些医生的信息。这些数据库包含类似于医学专业、医院特权、DEA#、上的医学院校、完成的居住计划、医学社团成员等的信息。链接到该数据使得可能创建元节点的全部新的类别以及无法单独通过 PubMed 数据构成的新型链接。

[0452] ● 脚本数据 - 另一种高价值的外生数据源是脚本数据。诸如 IMS 的专有数据库维护有关于医生的处方模式的信息。它们计算医生对他们开出的每一种药所写的处方的数量。该数据具有难以置信的价值作为生物工艺学和制药公司的信息, 以从市场观点看确定哪个医生是对于研究最为重要的。当结合 MNVS 时, 该工具使得生物工艺学和医药公司能够识别在所关心的治疗领域内与医学的最大量用户最为紧密相连的关键意见领导。通过以这些 KOL 为目标, 这些公司能够影响医生的处方模式并获取市场份额。

[0453] ● 参考数据 - 类似 LRX 的专有数据库还提供所链接到的有价值的外部数据源。LRX 数据库捕捉可用于创建专业内或跨专业的医学关系的社会网络的医生参考信息。

[0454] ● 调查数据 - 类似 Alpha Detail 的公司对成千上万个医生进行了调查以确定他们读取什么信息以及他们受其它什么医生所影响。该数据是有价值的,可作为尤其是在解决生命科学公司的市场问题中的另一外生数据源。

[0455] ● 授权数据 - 在医学文献被公开之前的一个步骤经常是提交和批准授权。在美国,大量的这些授权都到达国家健康机构 (NIH)。NIH 维护一个称为科学项目信息计算机检索 (CRISP) 的数据库。该数据库具有有关所有 NIH 提供基金的研究项目的信息。对该数据的链接使得甚至在首次医学文章公开之前就能够跟踪创新的研究。其它国家也维护有类似的数据。

[0456] ● FDA 试验数据 -FDA 在公共数据库中维持有关在 FDA 批准过程的各个阶段中的各种药品的候选信息。通过链接到该数据就可能分析对该药品流水线嵌入了多少医学研究以及评估各种药品在药品公司中的位置。

[0457] ● FDA 生产数据 - 在时标的另一端是 FDA 数据库。大部分的医学研究的目标是开发对某种疾病的治疗,这在大部分情况下必须通过 FDA 的批准 (在美国)。FDA 维护有 DRUG 数据库以及提供有关买卖双方直接交易的和处方药品以及食品添加剂和许多其它有关健康的产品的信息的许多其它数据库。通过链接到该数据,就可能跟踪包含在 PubMed 数据库中的研究的输出。

[0458] ● 机构数据 - 机构数据是可以结合到 MDVS 的另一外生数据源。可以通过 PubMed 数据库中的机构字段或者直接通过由医生 / 作者持有的机构从属关系的数据库链接到机构或公司数据。根据各种各样的来源存在各种类型的机构 / 公司数据。对该数据的链接使得可能更深入地分析公司、大学、政府、实体和研究机构在该感兴趣的研究领域内所起的作用。

[0459] ● 专利数据 - 到专利数据的链接也是相当重要的。专利数据代表已经被转换为可受到保护的知识产权的医学研究的那些部分。

[0460] 结合专利和医疗 NVS

[0461] 尽管已经分别描述了专利和 PubMed 实施例,但 NVS 也能够结合这两种数据源。这两种数据源之间的链接存在多种形式,包括从专利到学术文献的引用,可以链接文章作者和发明人,也可以链接公司 / 机构。通过将医学研究数据与专利数据以及授权、FDA 和脚本数据相链接,就可能获得一个想法从开始一直到产品认证和上市的整个生命周期的画面。

[0462] NVS 使得能够比之前可能的实现更深入地理解科学和技术开发的本质。许多不能被现有技术中已知的任何手段回答的不同类型的问题都可以得到回答。许多这些问题不仅在经济上而且对于社会产品都具有非常高的价值。

[0463] 其它实施例

[0464] 对本领域的技术人员显而易见的是,NVS 可以应用于广阔的环境中,所描述的实施例论证了 NVS 对两种不同数据源的应用性。以对所描述的实施例中所介绍的那些类似的方式对许多其它来源的应用也是可能的。

[0465] 计算机实现

[0466] 根据本发明的各个实施例分析数据库记录的方法优选在图 32 所示的通用计

计算机 300 中实现。代表性的计算机 300 是个人计算机或工作站平台, 例如基于 Intel **Pentium®**、**PowerPC®** 或 RISC 的, 并包括诸如 **Windows®**、**Linux®**、**OS/2®**、**Unix** 等的操作系统。众所周知的是, 这些机器包括显示接口 302(图形用户界面或 GUI) 以及相关的输入装置 304(例如键盘或鼠标)。

[0467] 该数据库记录分析方法优选以软件实现, 因此其中一个实施例是驻留在诸如计算机 300 的随机存取存储器 308 的计算机可读介质中的代码模块中的一组指令 306(例如程序代码)。直到计算机 300 需要时, 该组指令 306 可以被存储在另一计算机可读介质 310 中, 例如硬盘驱动器, 诸如光盘(最终用于 CDROM) 或软盘(最终用于软盘驱动器) 的可移动存储器中, 或者经由因特网或某些其它计算机网络下载。另外, 虽然所述的各种方法可以很方便地在通过软件选择性激活或重新配置的通用计算机 300 中实现, 本领域的技术人员也应该认识到这些方法也可以以硬件、固件或被构造用于执行该指定的方法步骤的更专用的设备中执行。

[0468] 其它方面、修正以及实施例都在本权利要求书的范围内。

图1（现有技术）

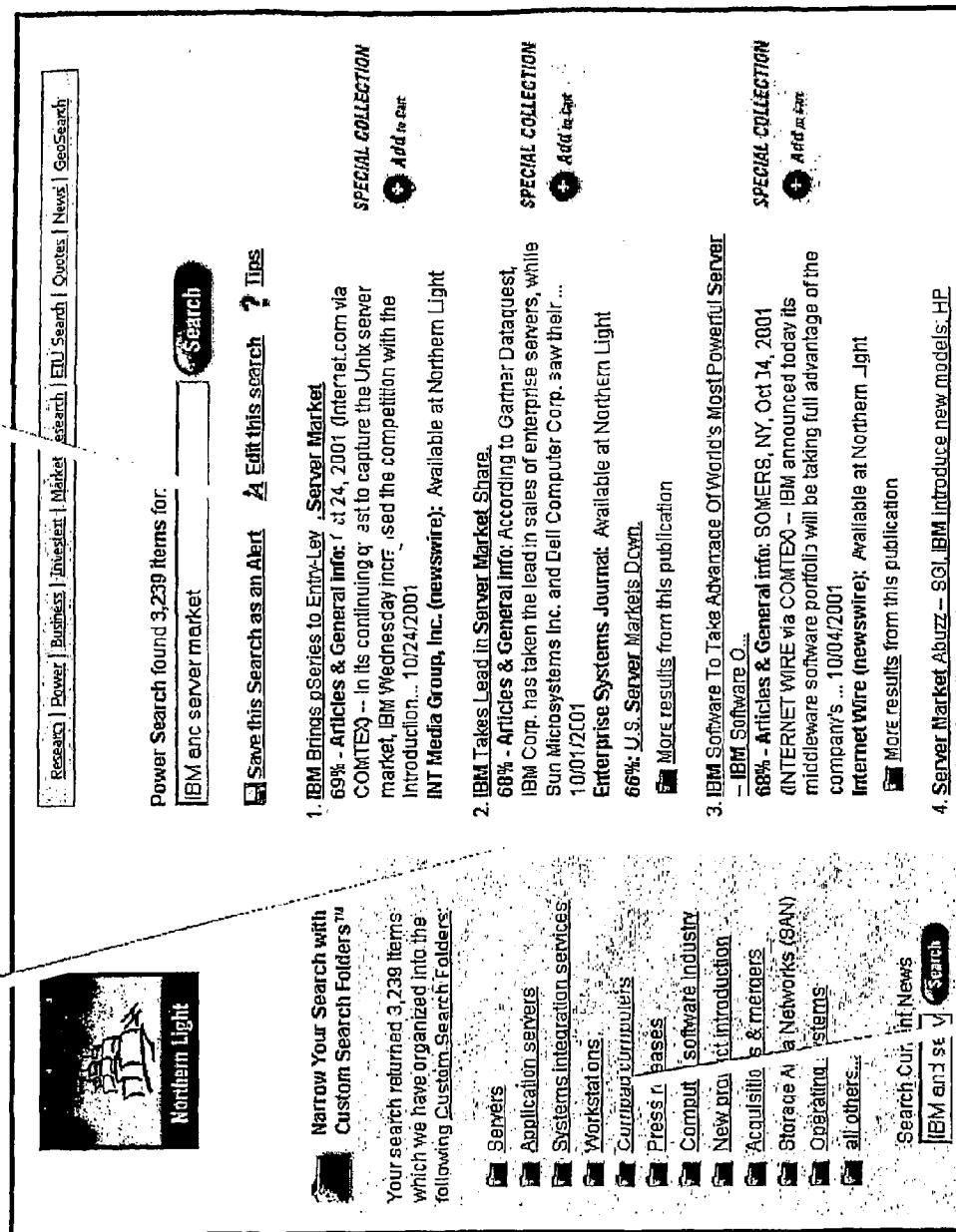


图 2 (现有技术)

The screenshot shows a search results page for 'Visualization' on Vivisimo.com. The top navigation bar includes links for company, products, solutions, customers, demos, press, Visualization, the Web, Advanced Search, Help, and a search bar. A message at the top says 'NEW search the Wikipedia at Clustify.com'. The search results are clustered under several main categories:

- Visualization (229)**
 - ▶ **Graphics** (35)
 - ▶ **VTK, Open source** (4)
 - ▶ **Solutions, Engineers** (4)
 - ▶ **Virtual reality** (3)
 - ▶ **Modelling and rendering** (3)
 - ▶ **International Conferences in Central Europe** (2)
 - ▶ **Real Time, Industry** (2)
 - ▶ **Sci** (2)
 - ▶ **Computer Interaction** (2)
 - ▶ **Art** (2)
 - ▶ **Other Topics** (12)
 - ▶ **Scientific** (24)
 - ▶ **Data Visualization** (29)
 - ▶ **Laboratory** (24)
 - ▶ **Solutions** (16)
 - ▶ **Tools** (16)
 - ▶ **Graph** (13)
 - ▶ **Resources** (12)
 - ▶ **Visualization System** (10)
 - ▶ **Visualization And Image** (9)

Below the categories, there are two sponsored links:

- Image Analysis** [new window] [preview]
 - 300+ operators, Signal Processing, Image Processing, 3D visualization
www.acusoft.com
- The Art of Visualization** [new window] [preview]
 - Using the powers of thought and imagination for major life changes.
www.tut.com/fip.htm

The main search results list includes:

- OpenDX** [new window] [frame] [preview] [cluster]
 - Check out this open source 3D visualization utility. Programmers include original image galleries, program features and specifications, and support options.
www.opendx.org - Lykos 8, Ask.Jeepes 8, Open Directory 20, Wisenut 33, Looksmart 48, MSN 51
- Collaborative Visualization** [new window] [frame] [preview] [cluster]
 - Supporting teachers in transforming science learning to better resemble the authentic practice of science.
www.covis.mn.edu - Wisenut 5, Looksmart 10, MSN 58
- Scientific Visualization Studio** [new window] [frame] [preview] [cluster]
 - Provides an understanding of science through visualization. Source for global surface temperature anomalies.
srs.gsfc.nasa.gov - Wisenut 2, MSN 34, Looksmart 45
- UC Berkeley - Ballando Information Visualization Projects** [new window] [frame] [preview] [cluster]
 - Explore methods of visualizing abstract information more effectively. Offer tutorials, a course, & material on graphs, tilesbars and user interfaces.
ballando.sims.berkeley.edu/infovis.html - MSN 3, Looksmart 14
- RasMol** [new window] [frame] [preview] [cluster]
 - Download free software that allows users to see and animate 3D molecular structures. Check out video clips of molecules in motion.
www.umass.edu/microbio/rasmol - MSN 10, Ask.Jeepes 11, Looksmart 32
- xml visualization laboratory** [new window] [frame] [preview] [cluster]
 - More

图3(现有技术)

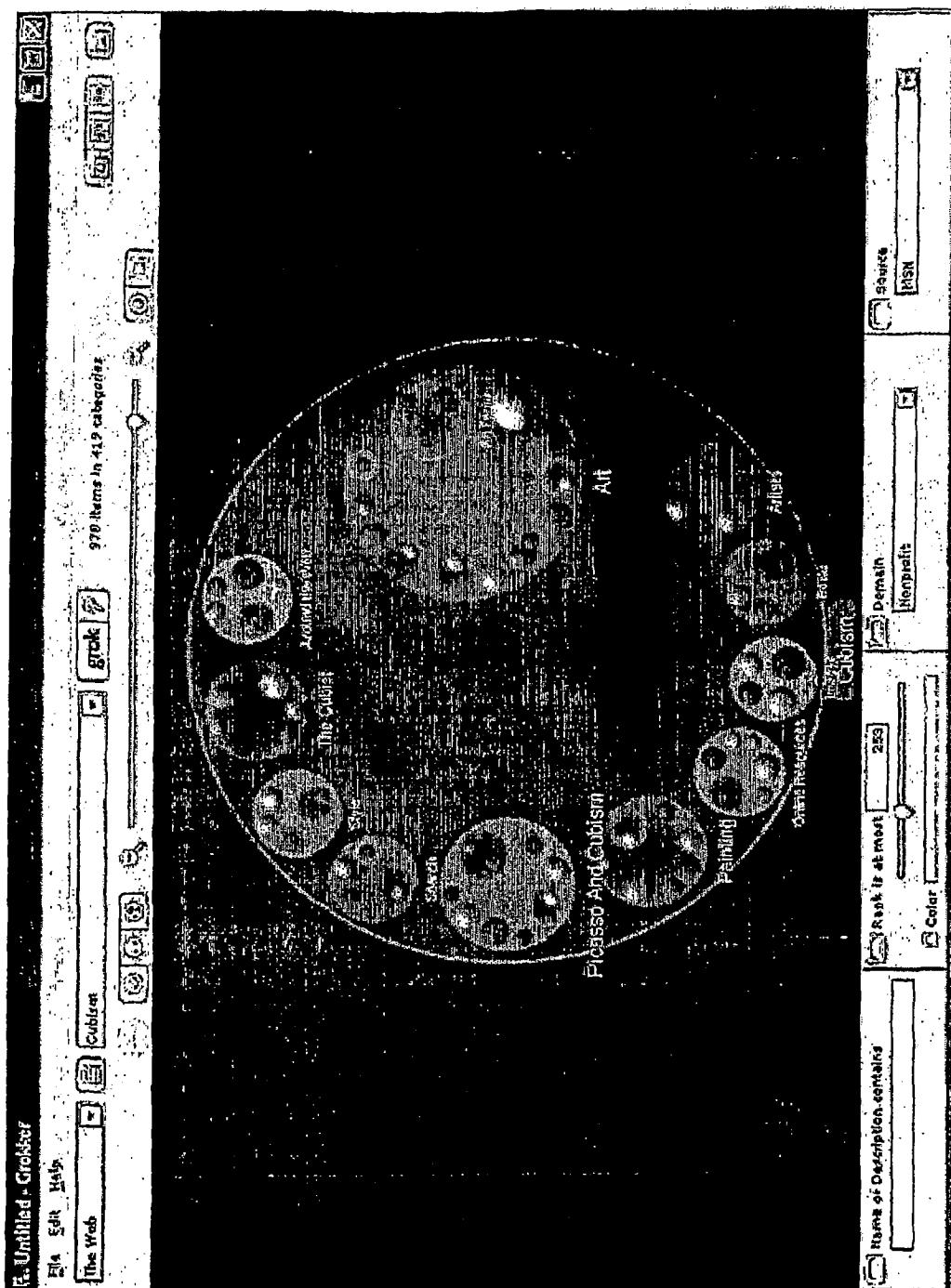


图 4 (现有技术)

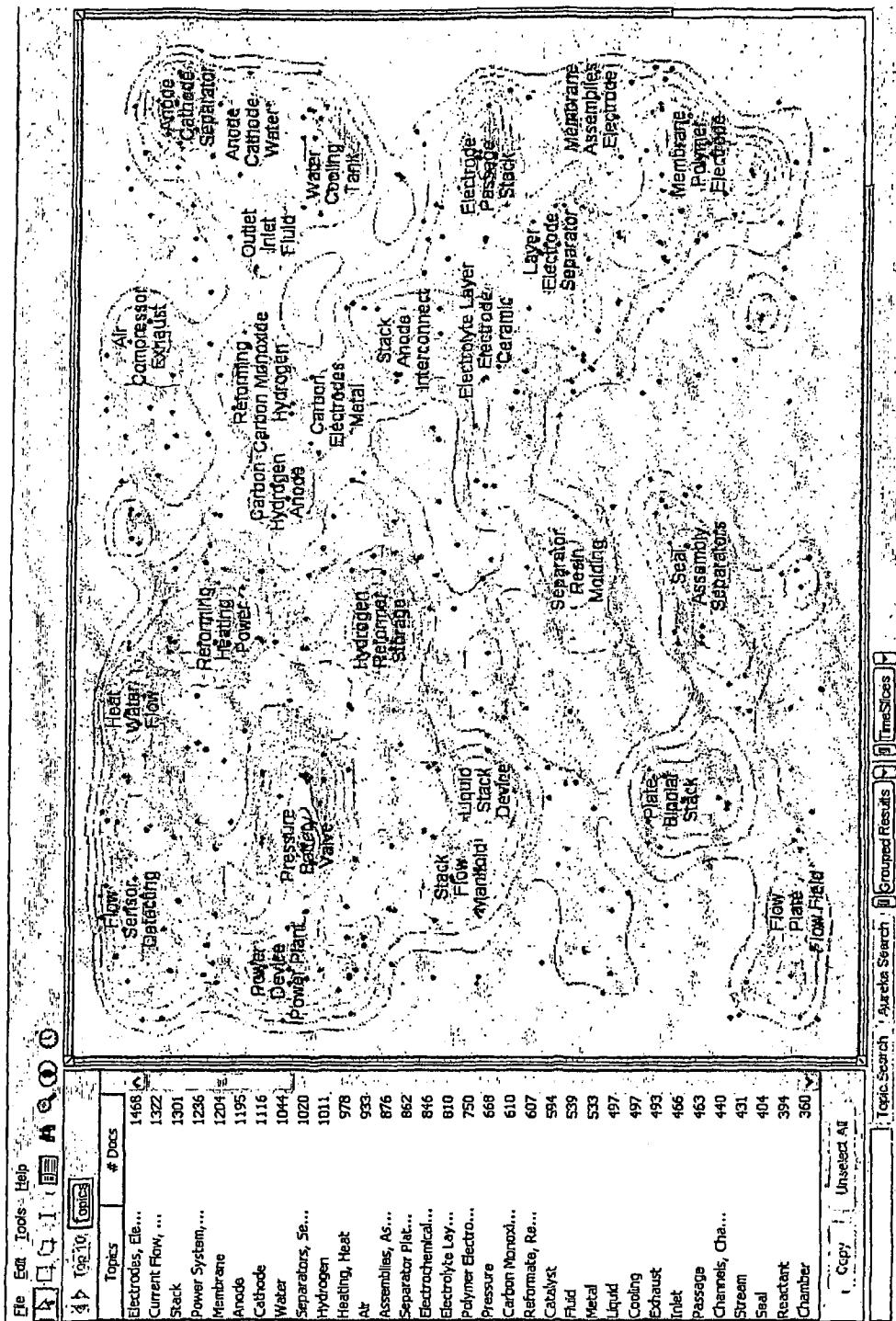


图5(现有技术)

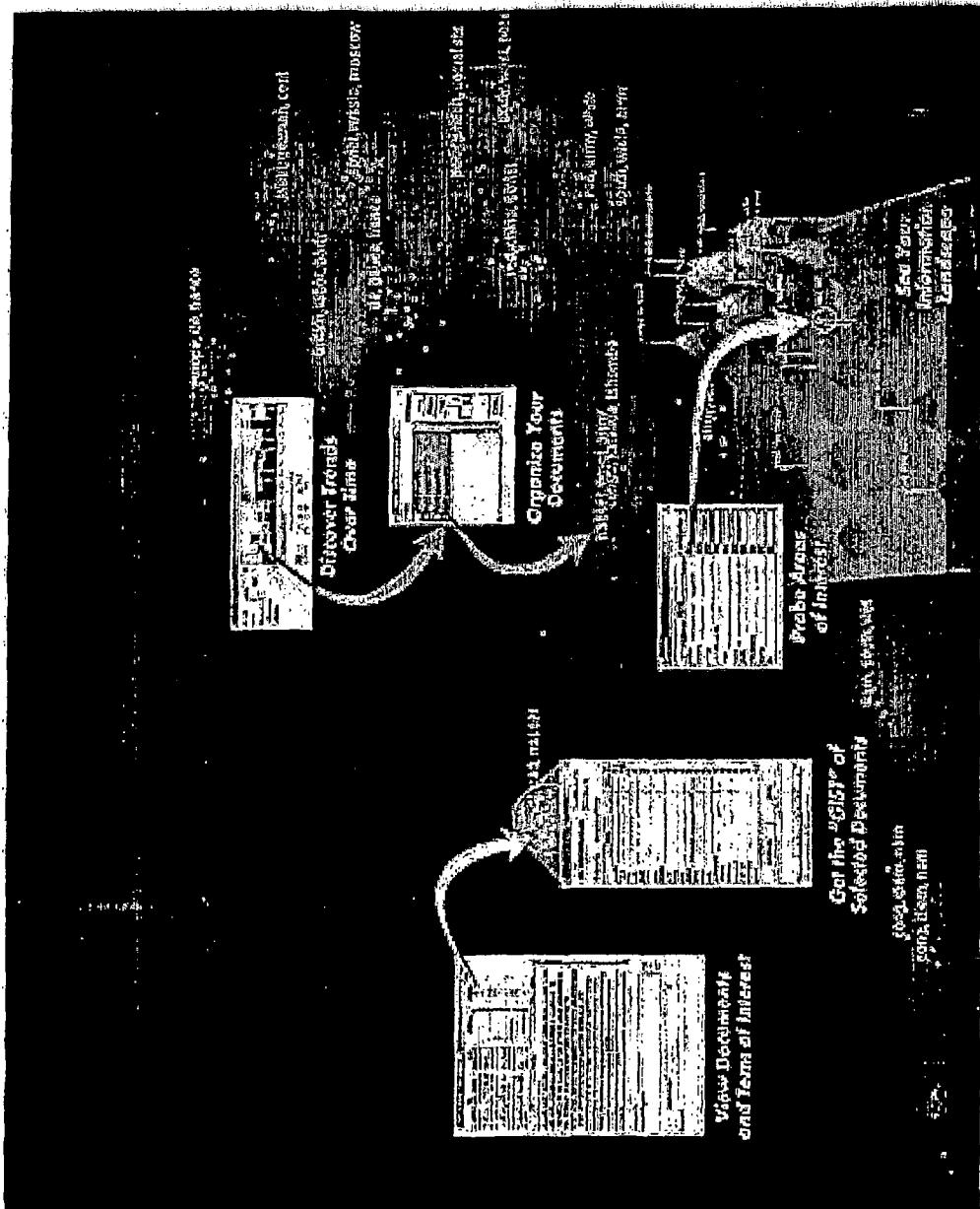


图6
(现有技术)

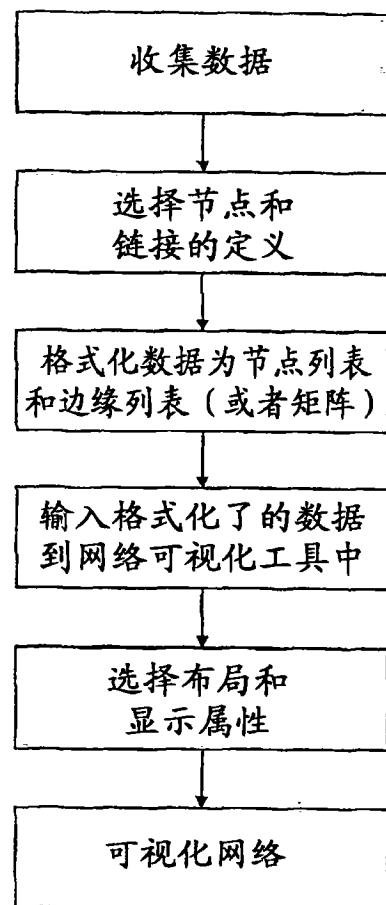


图 7

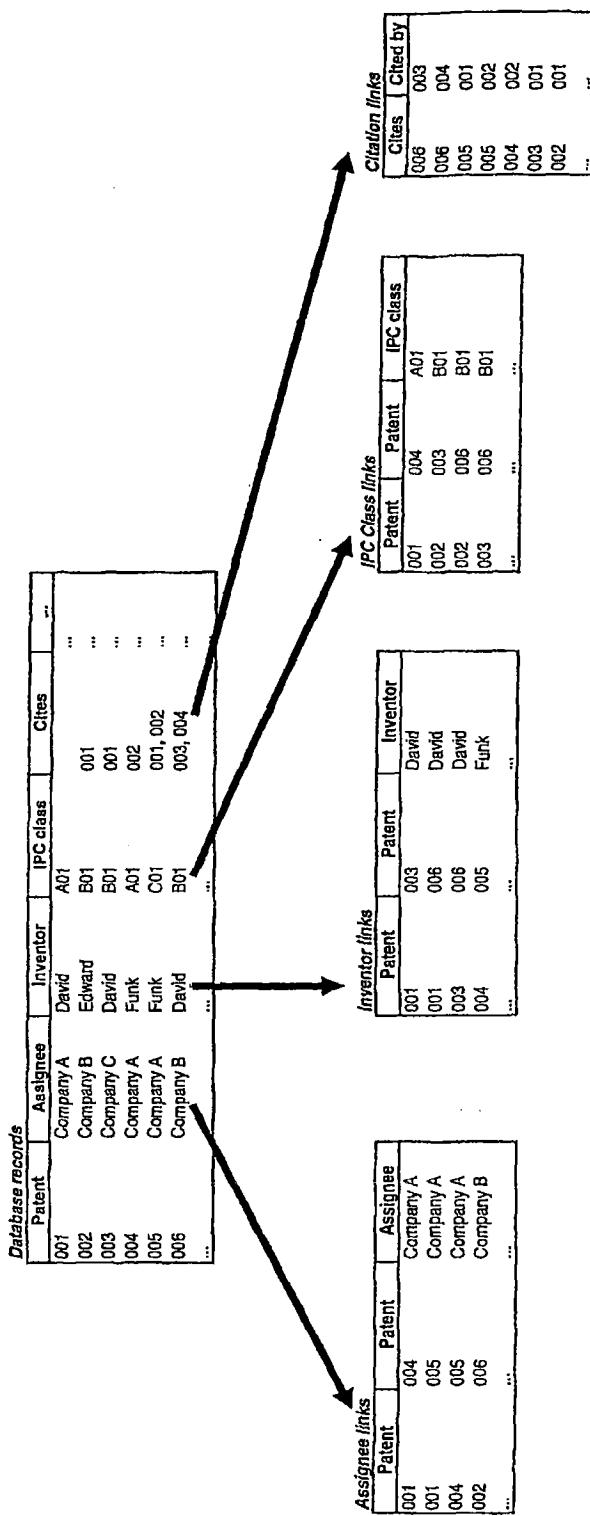


图 8

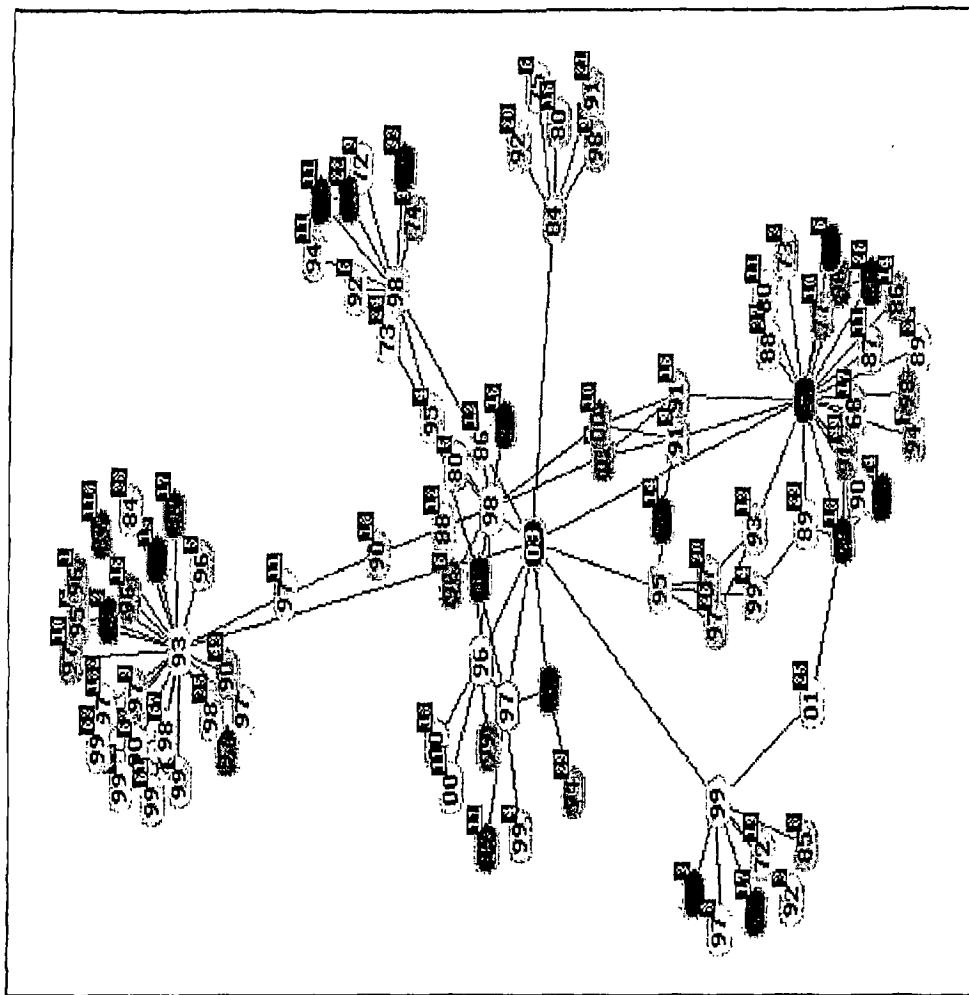


图 9

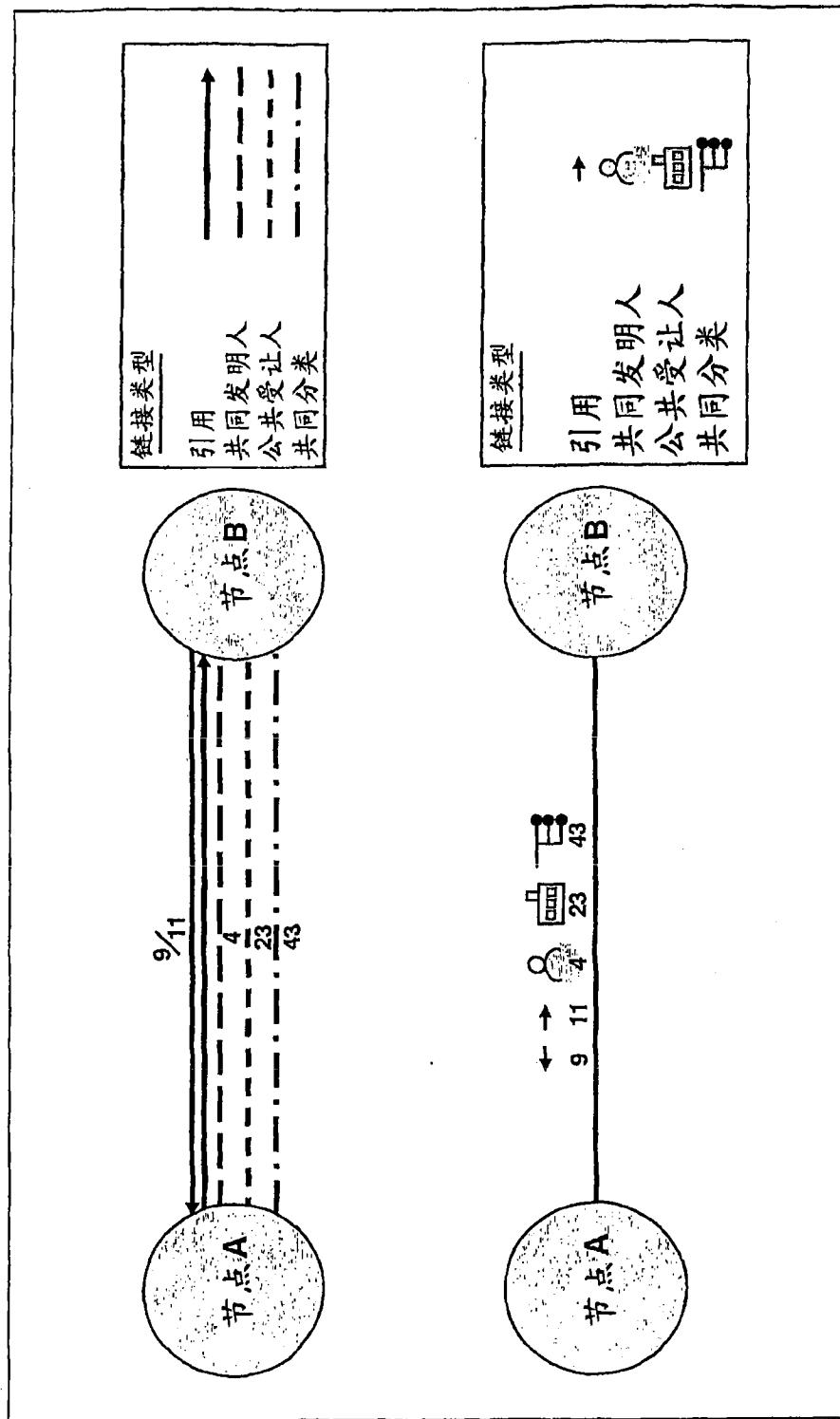


图 10

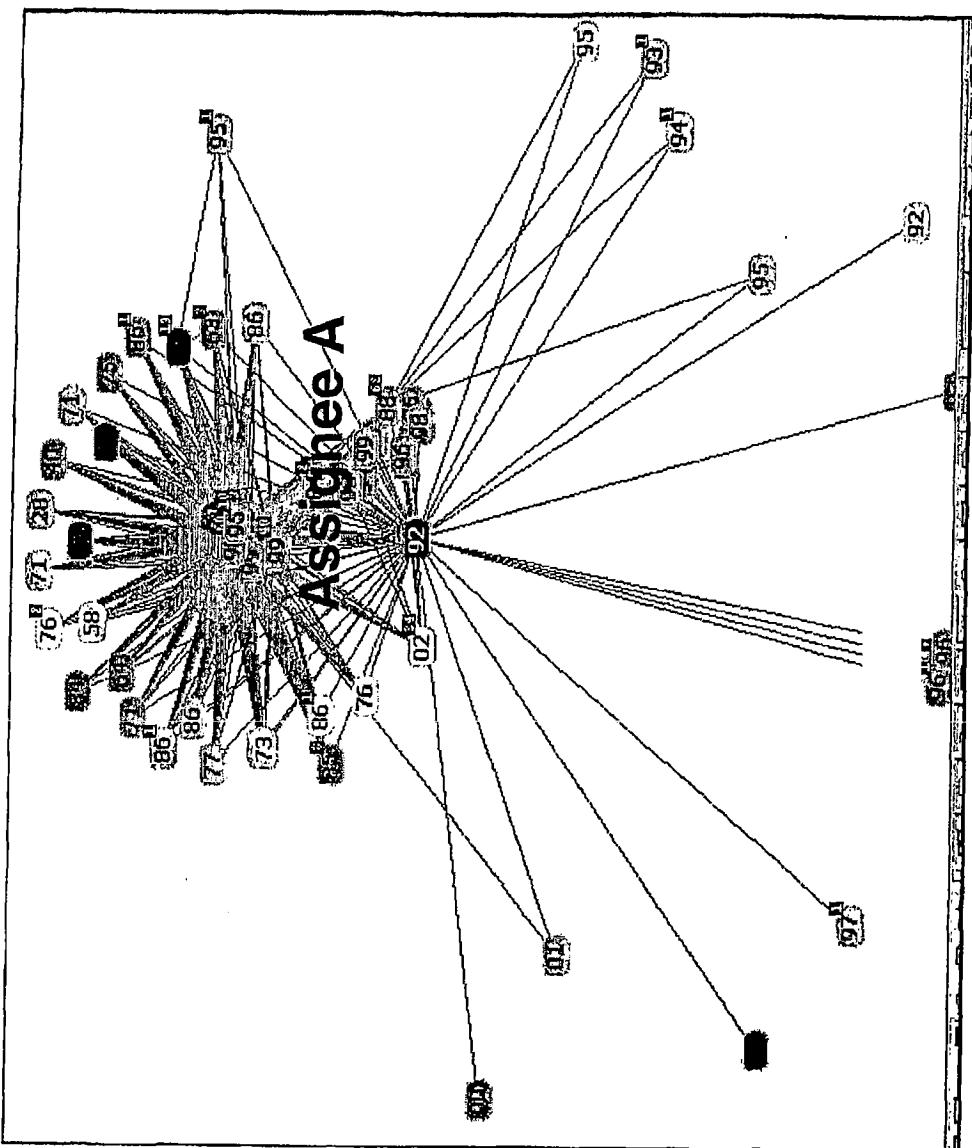
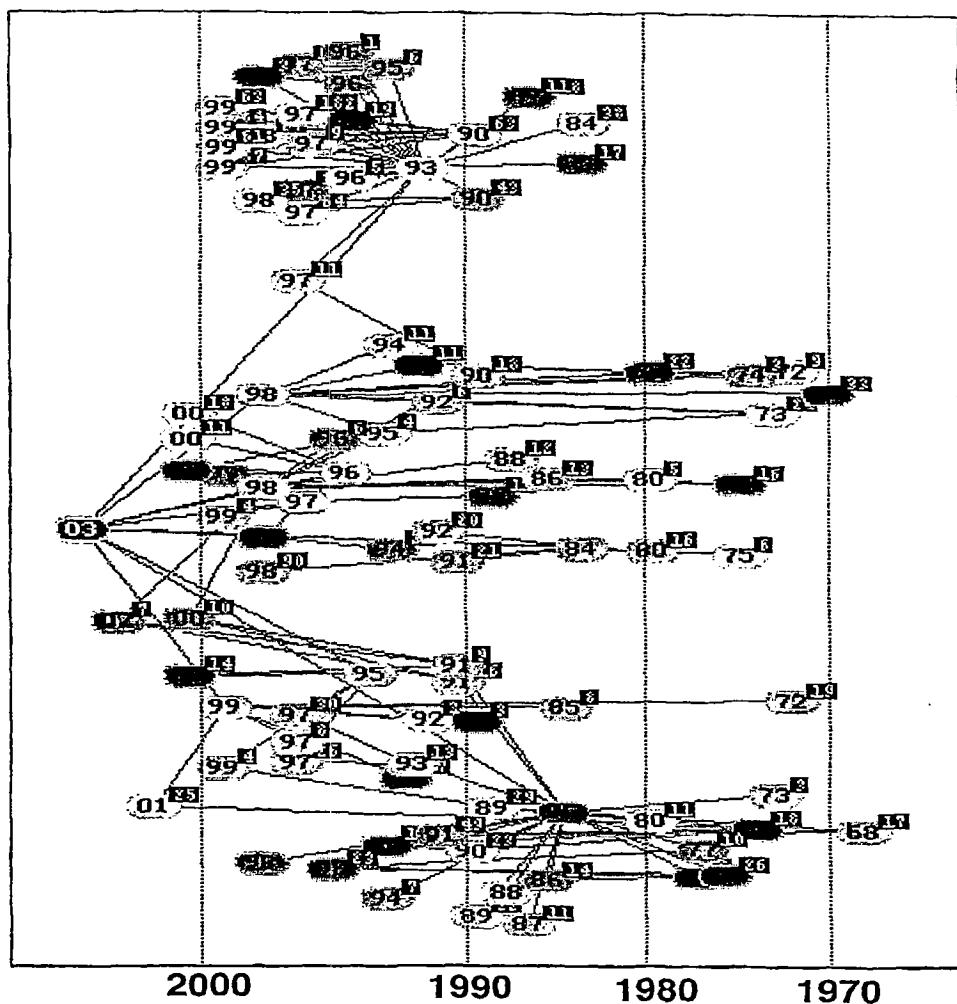


图 11



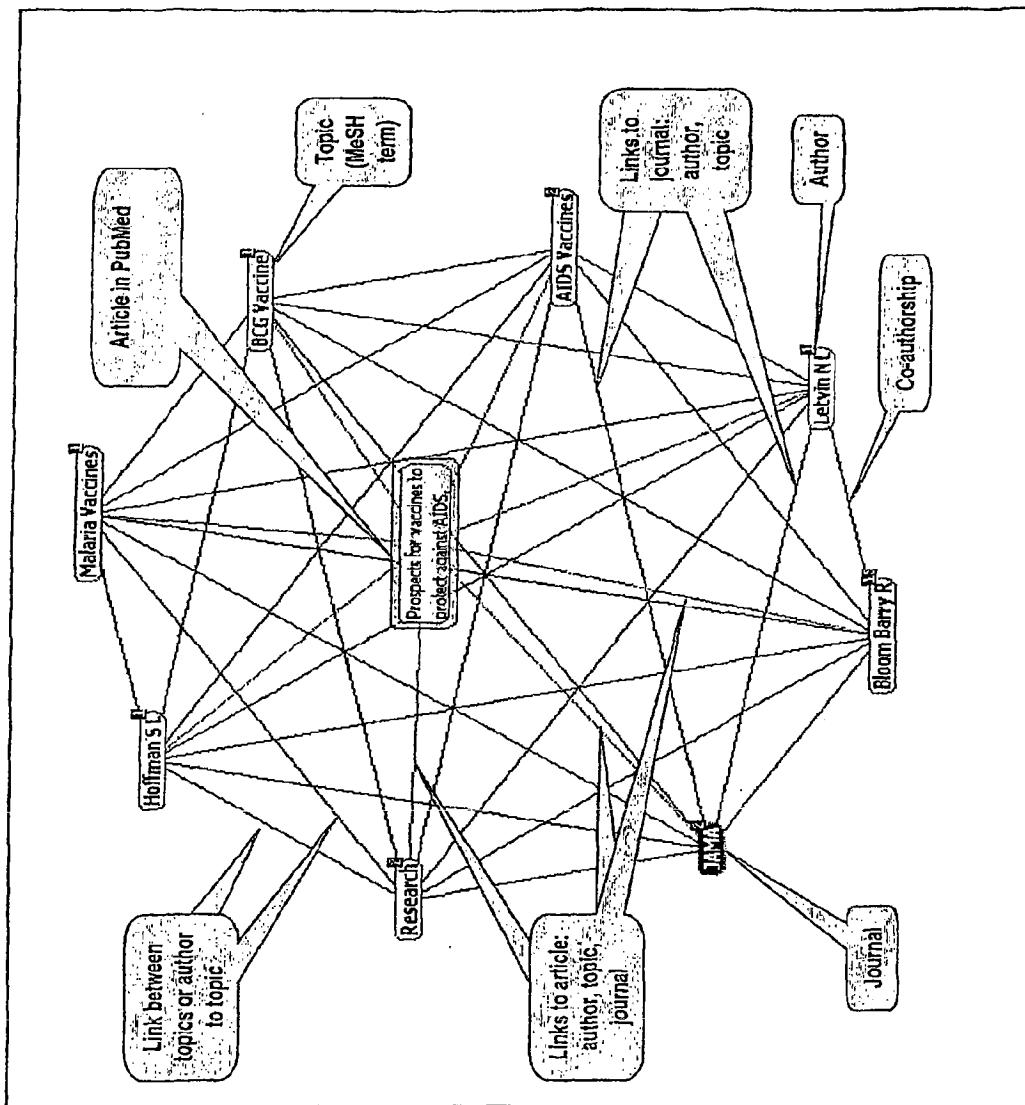


图 12

图 13

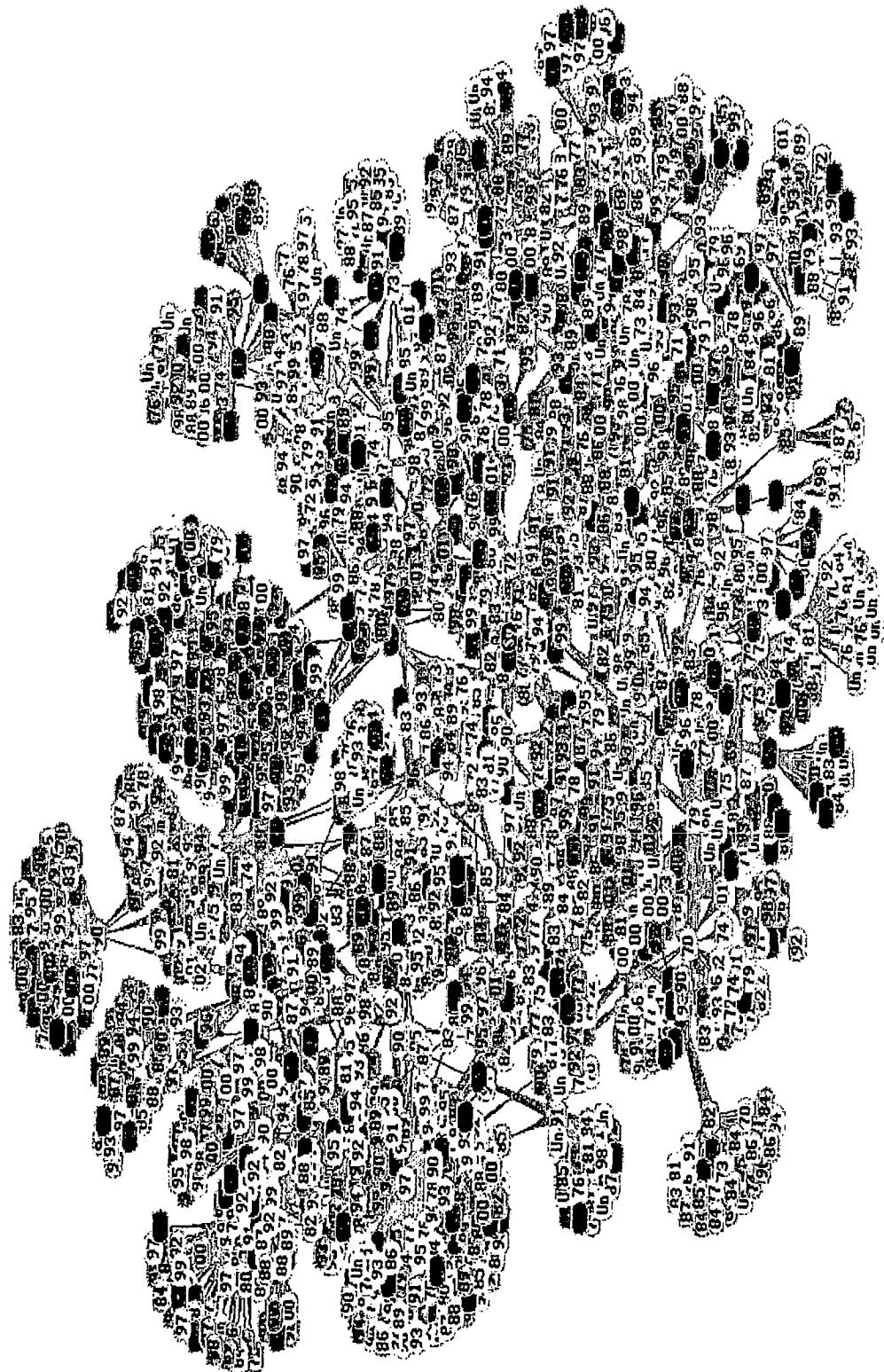


图 14

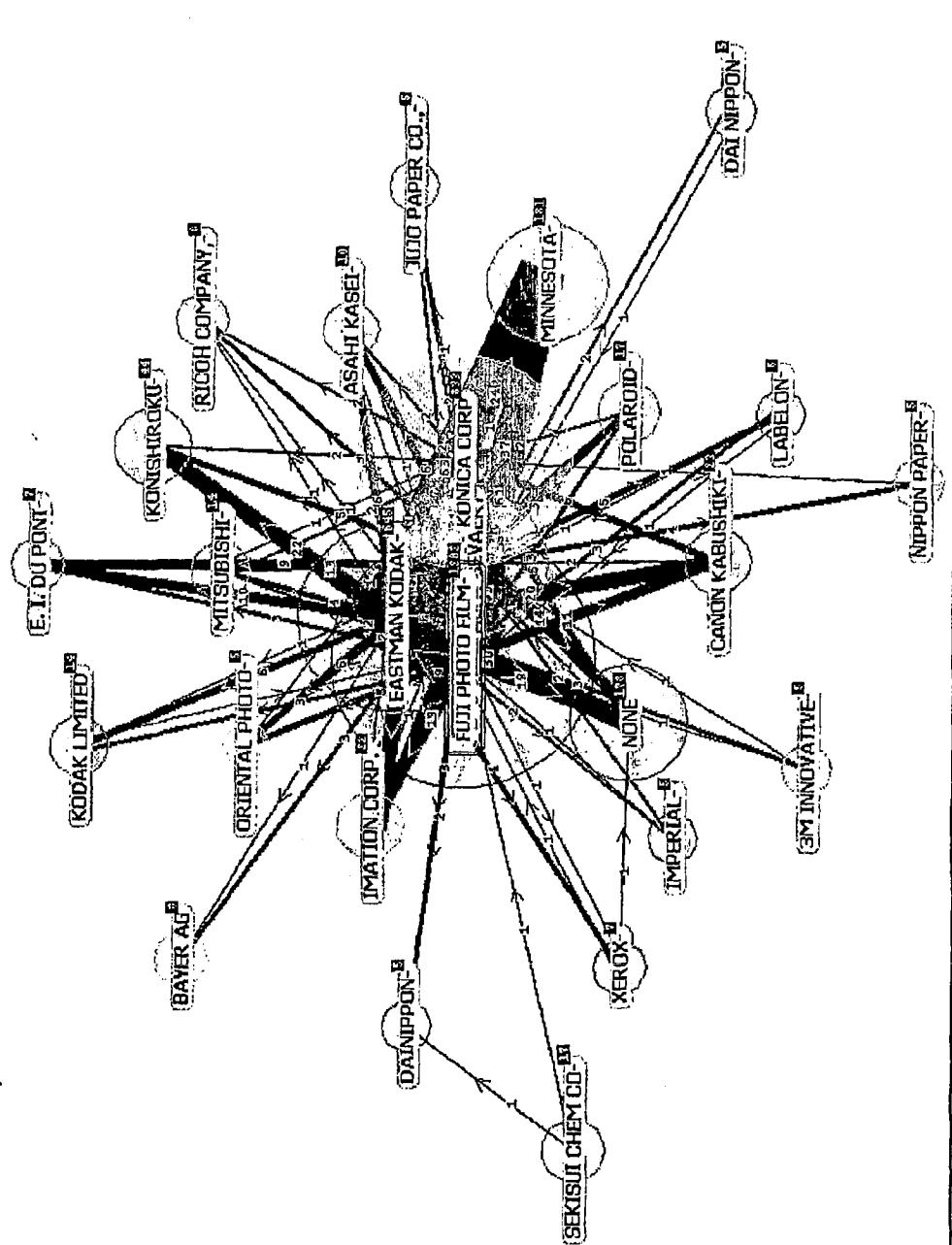


图 15

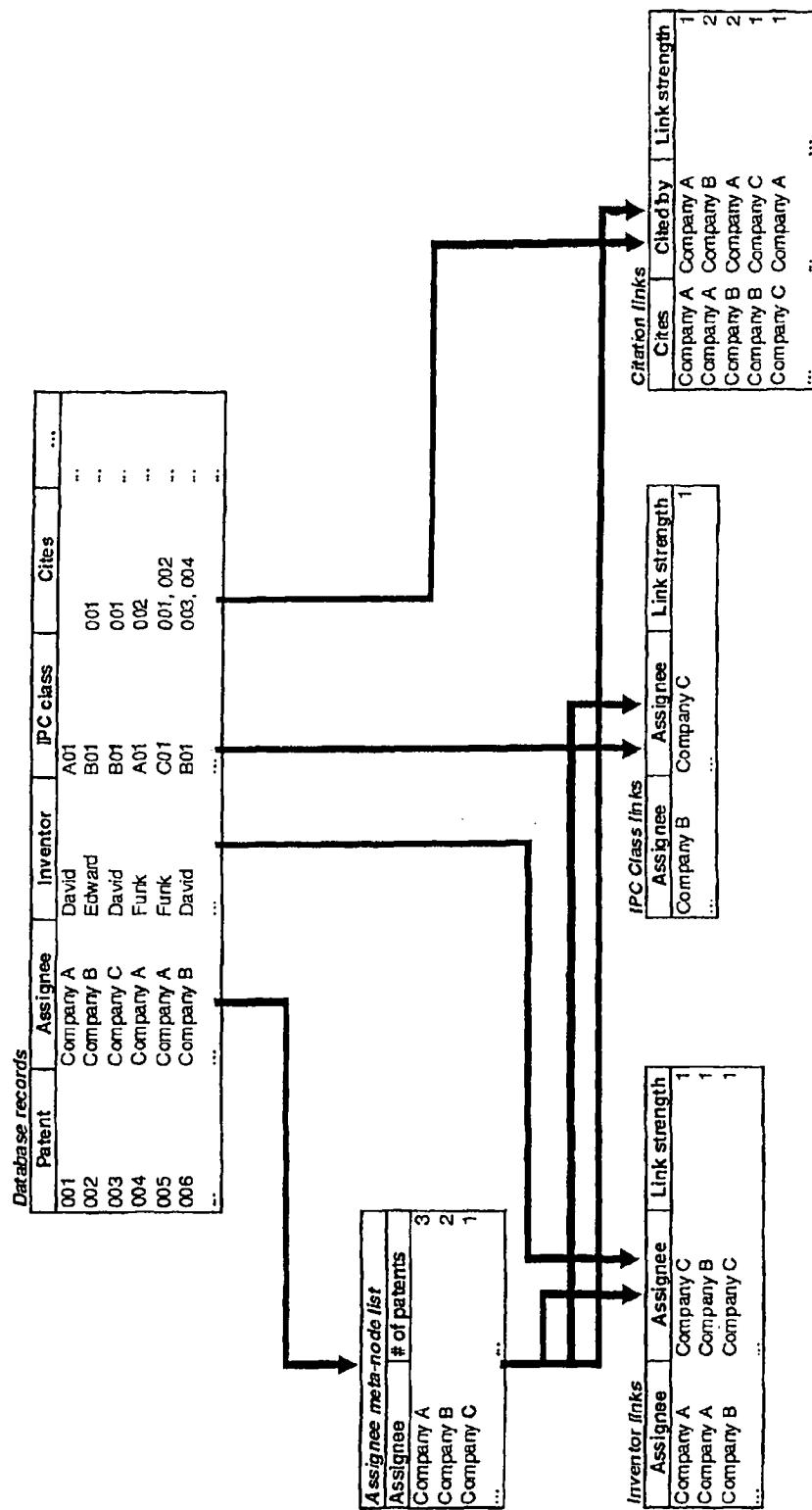


图 16

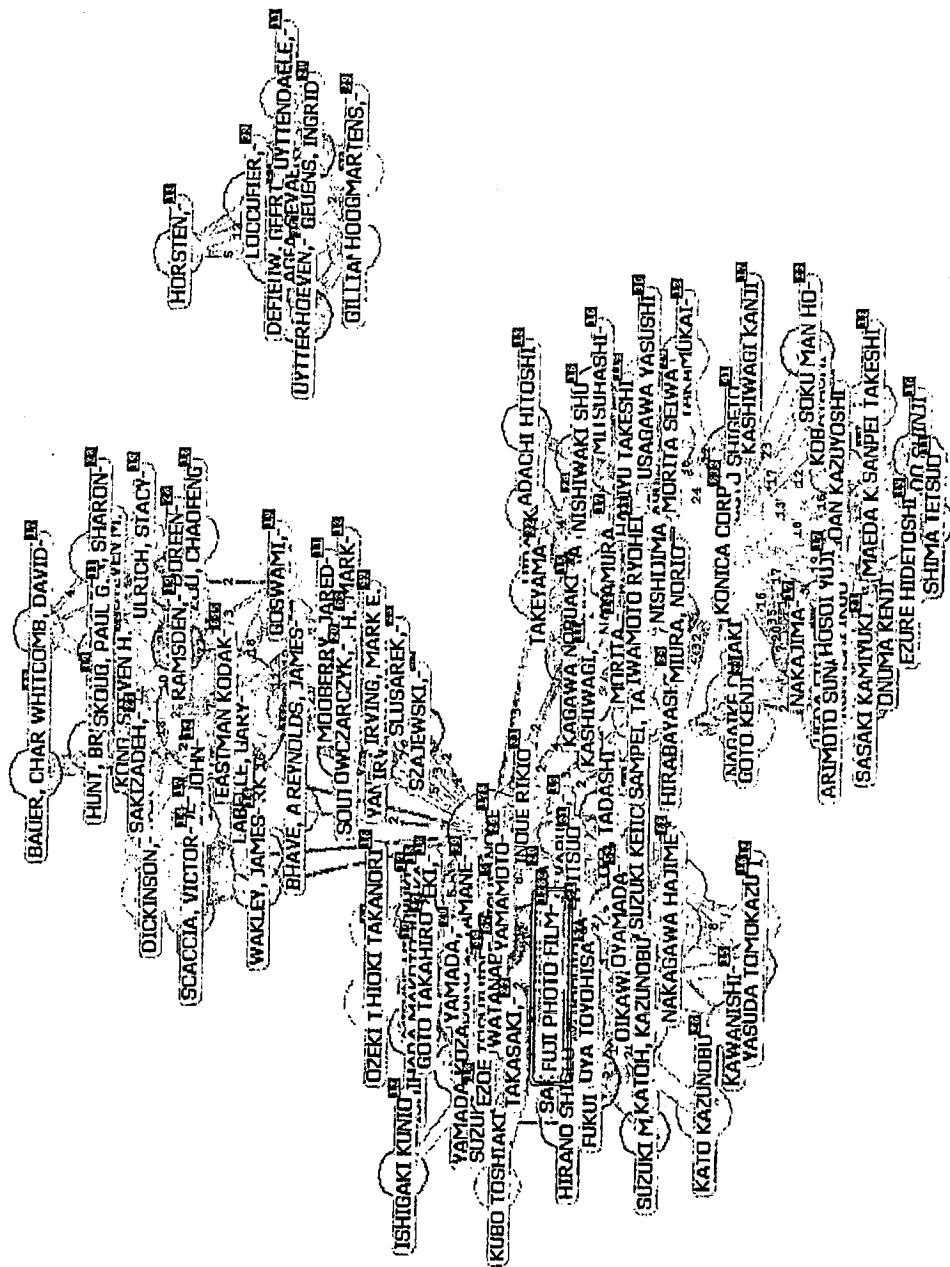


图 17

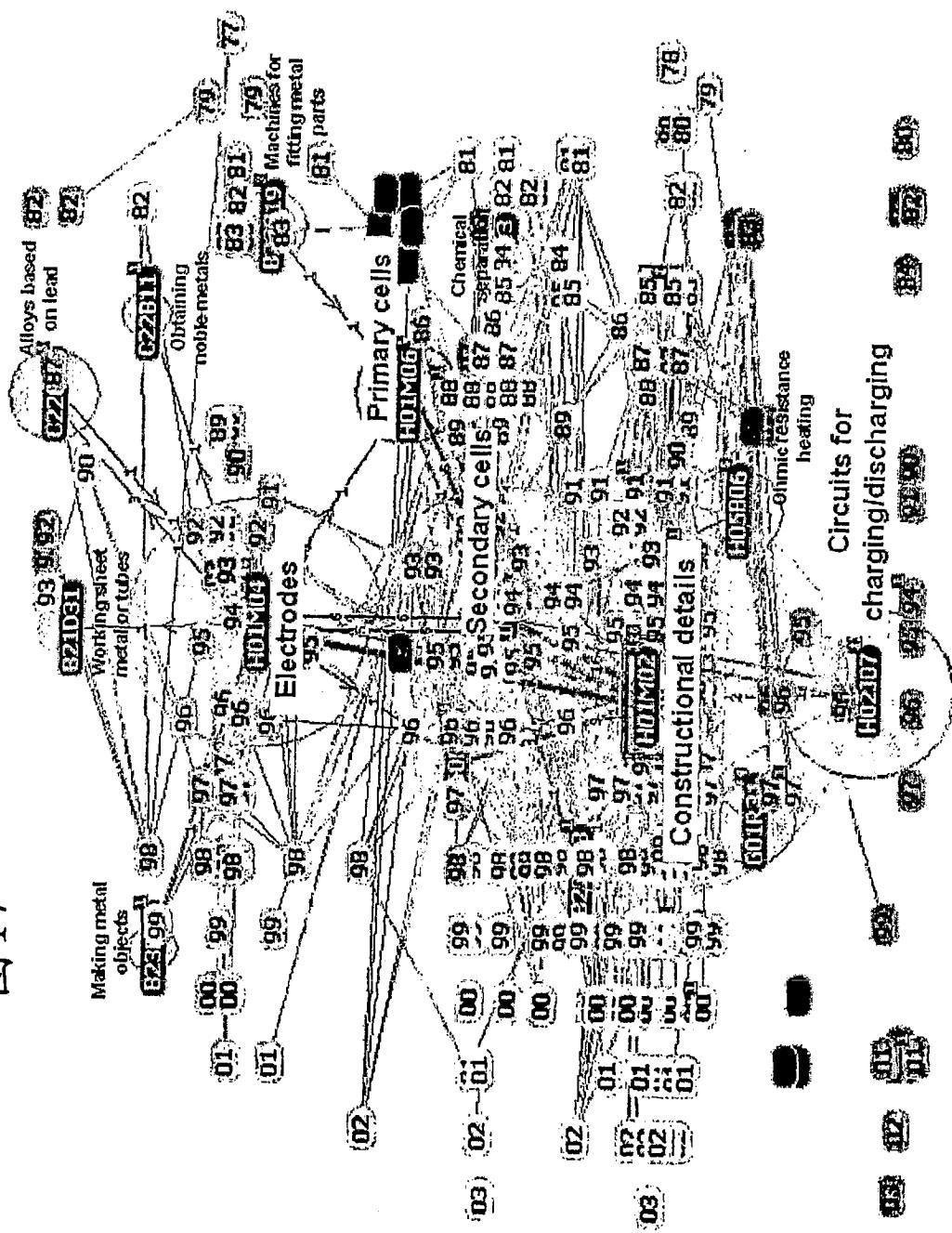


图 18

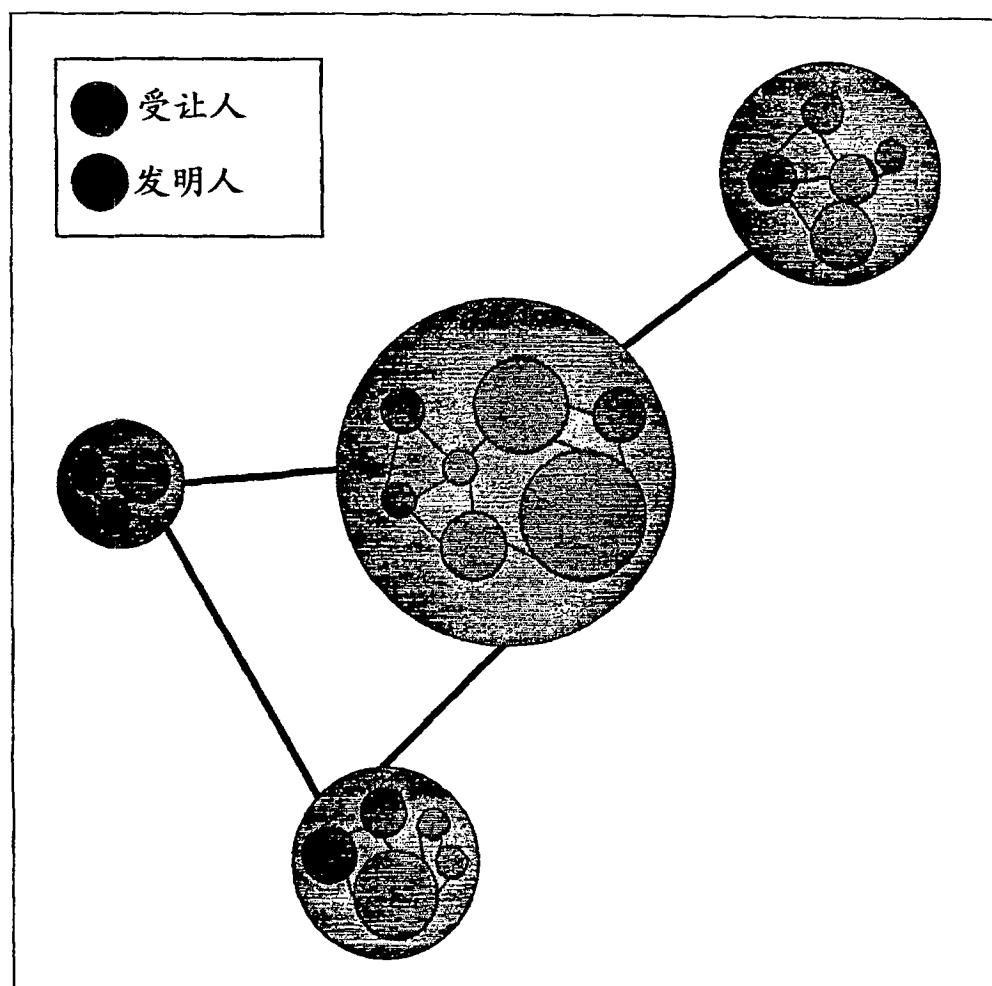


图 19

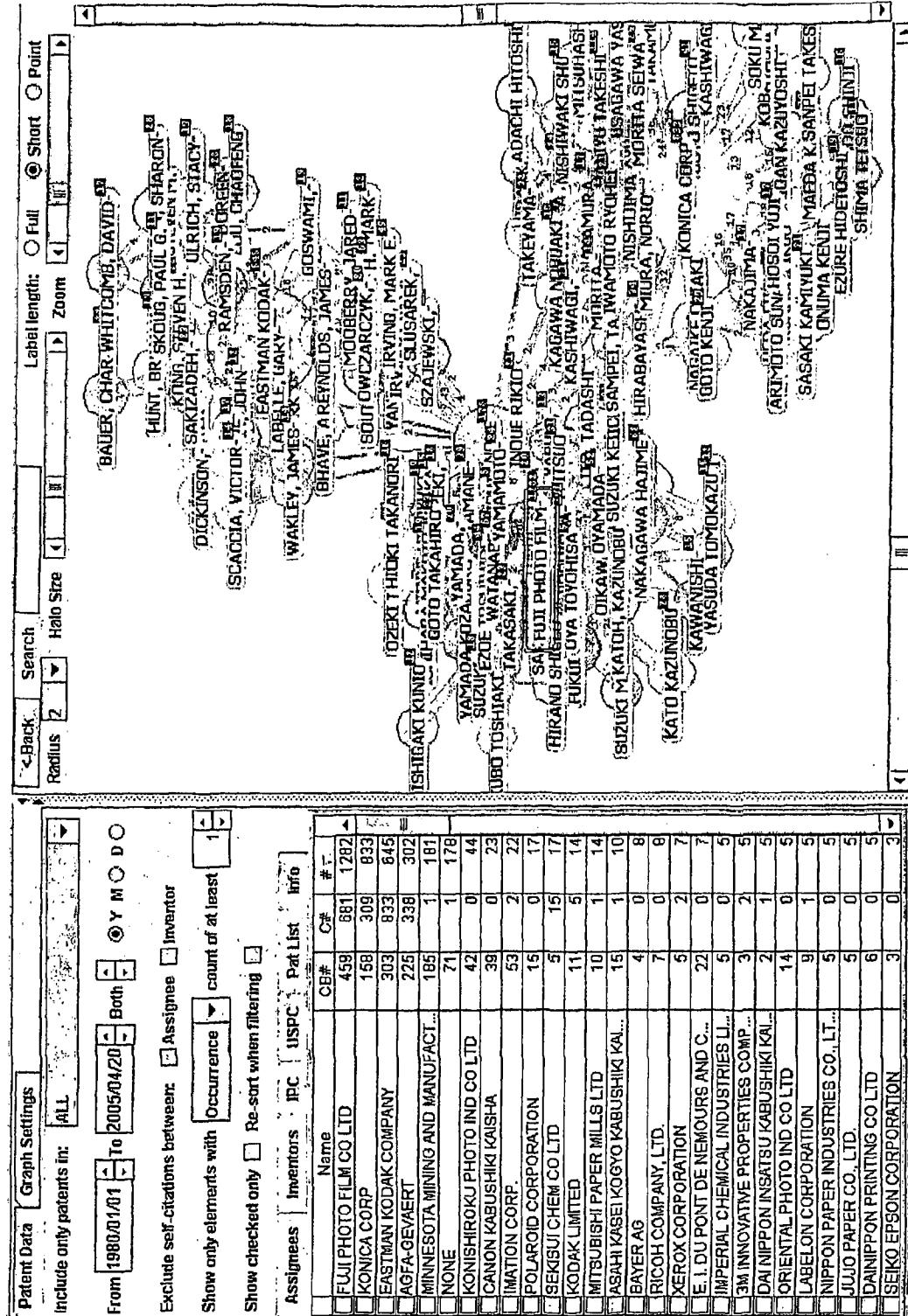


图 19A

Group	Name
② FUJI PHOTO FILM CO LTD	
	FUJI PHOTO FILM CO LTD
	FUJI PHOTO FILM CO., LTD.
	FUJI PHOTO FILM CO., LTD
	FUJI PHOTO FILM CO., LTD
	FUJI PHOTO FILM CO LTD(JP)
	FUJI PHOTO FILM CO.
	FUJI PHOTO FILM CO. LTD
	FUJI PHOTO FILM PHOTO, CO., LTD.
	FUJI PHOTO FILM, CO. LTD.
	FUJI PHOTO FILM, CO., LTD.
	FUJI XEROX CO., LTD.
③ KONICA CORP	
	KONICA CORP
	KONICA CORPORATION
	KONICA CORP.
	KONICA MINOLTA MEDICAL & GRAPHIC INC
	KONICA MINOLTA MEDICAL & GRAPHIC INC.
	KONICA MINOLTA HOLDINGS, INC.
	KONICA MINOLTA HOLDINGS INC
④ EASTMAN KODAK COMPANY	

图 20

专利数据的属性		学术数据的类似属性生	
专利	论文（文章）		
专利号	PMID号（PubMed ID号）		
申请的国家	期刊		
标题	标题		
摘要	摘要		
发明人	作者	作者	
• 地址（城市，州，国家）		• 地址（城市，州，国家）	
受让人	机构	机构	
• 地址（城市，州，国家）系）		• 地址（城市，州，国家）	
分类	分类	医学主题标目（MeSH）	
IPC/USPC/Derwent		• 医学主题标目（MeSH）	
引用	引用		
日期	日期	日期	
• 优先权日		• 无类似的	
• 申请日		• 提交日	
• 公布日		• 公开日	
• 授权日		• 受理日	

图 21

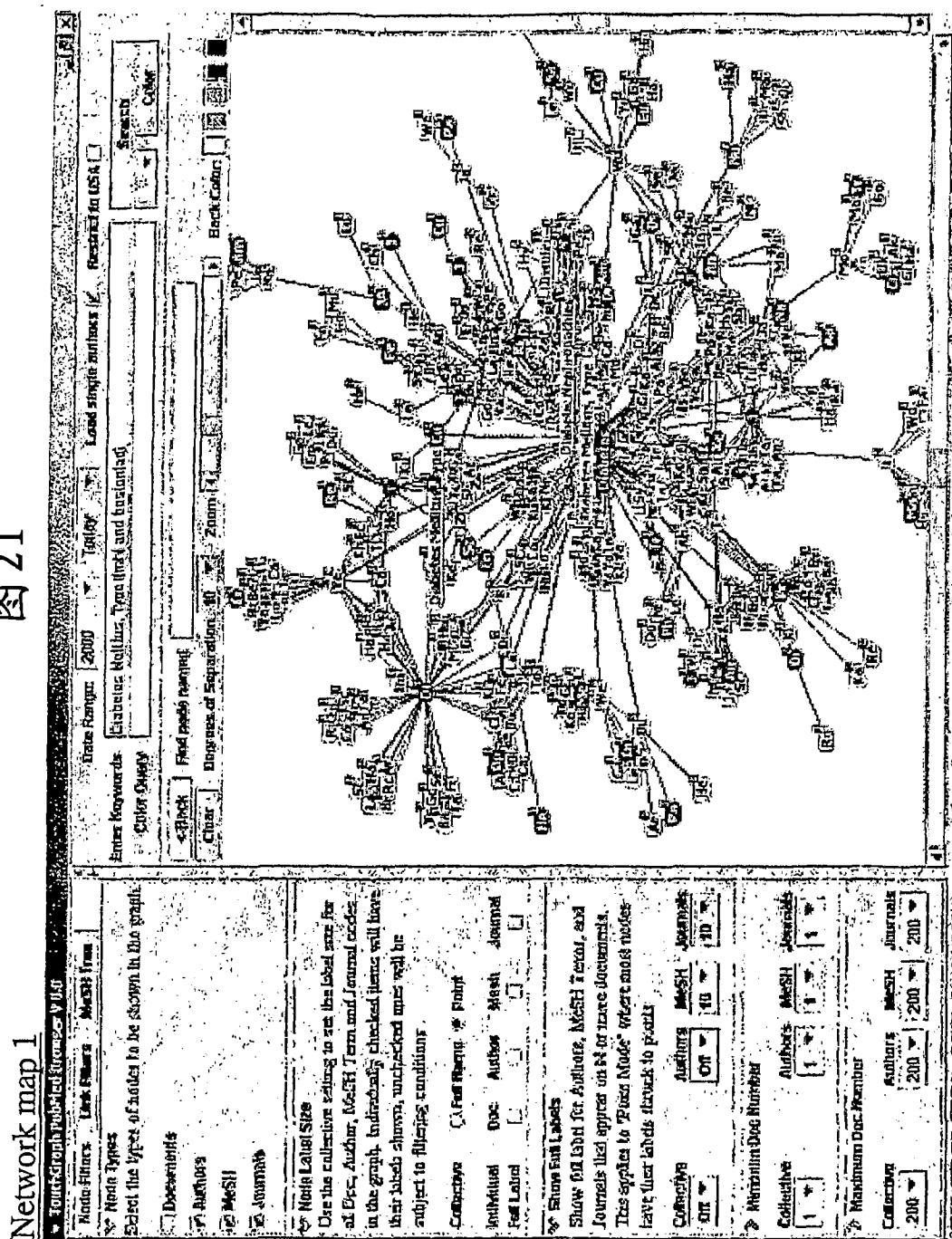


图 22

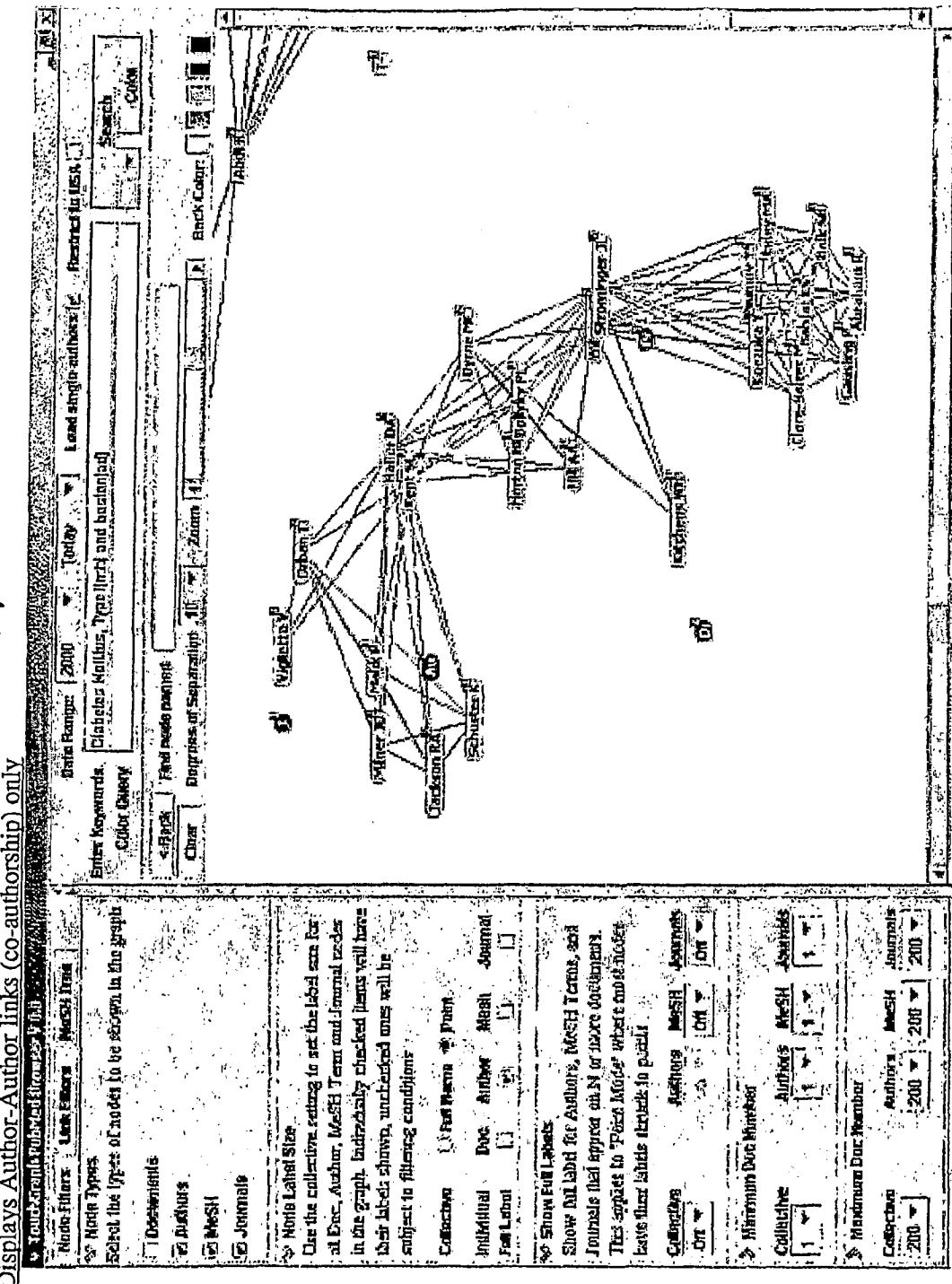


图 23

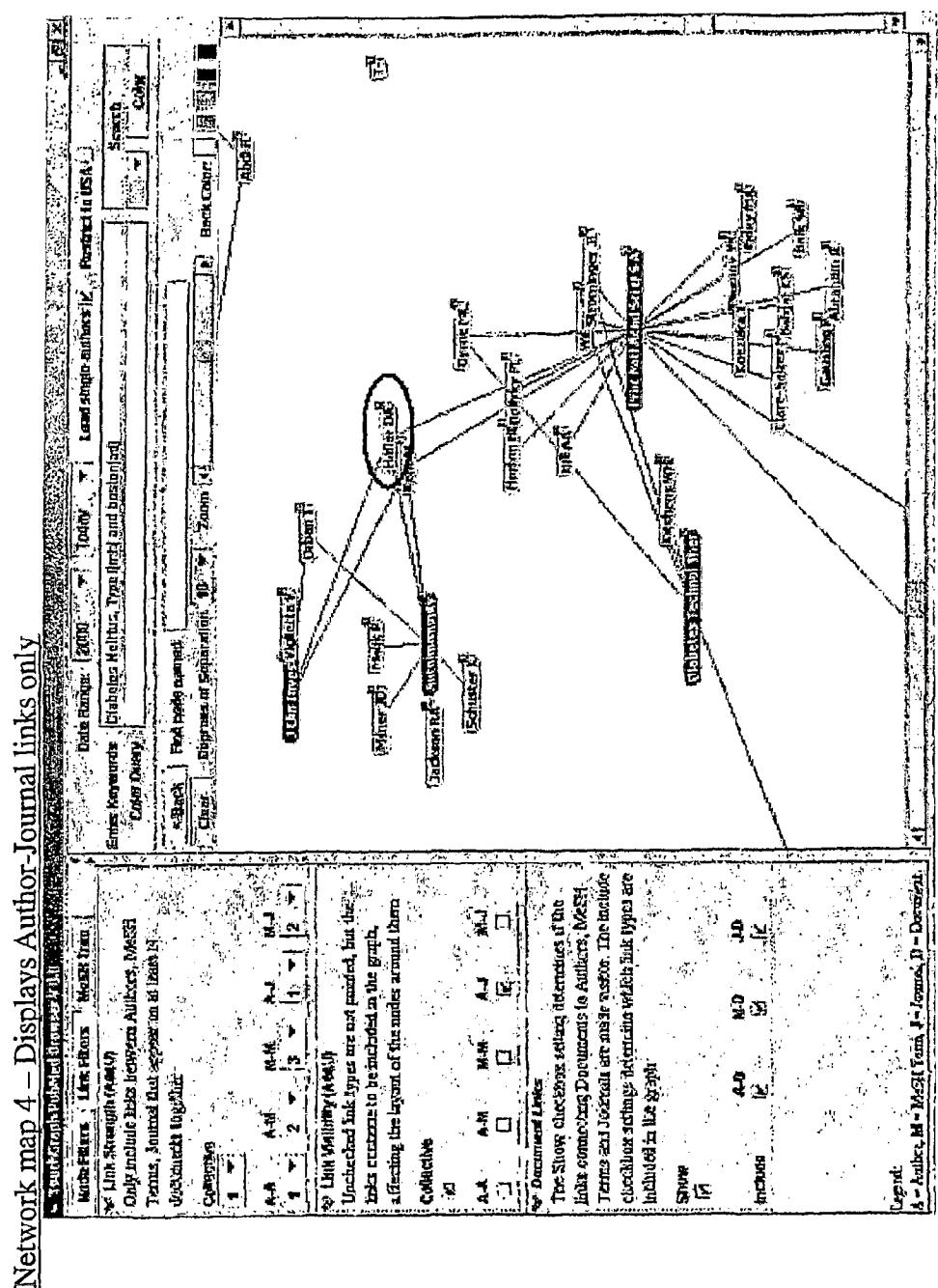


图 24

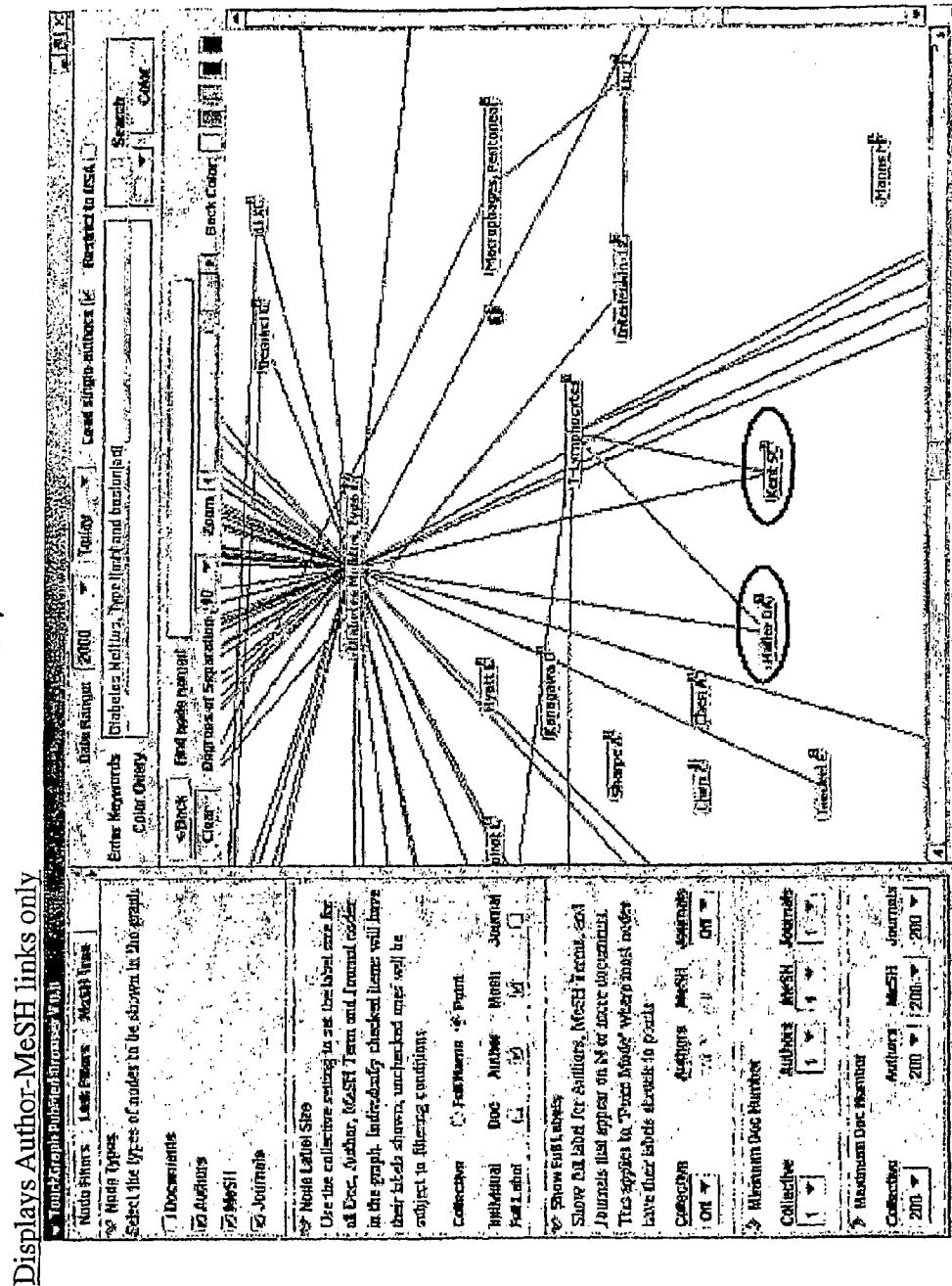


图 25

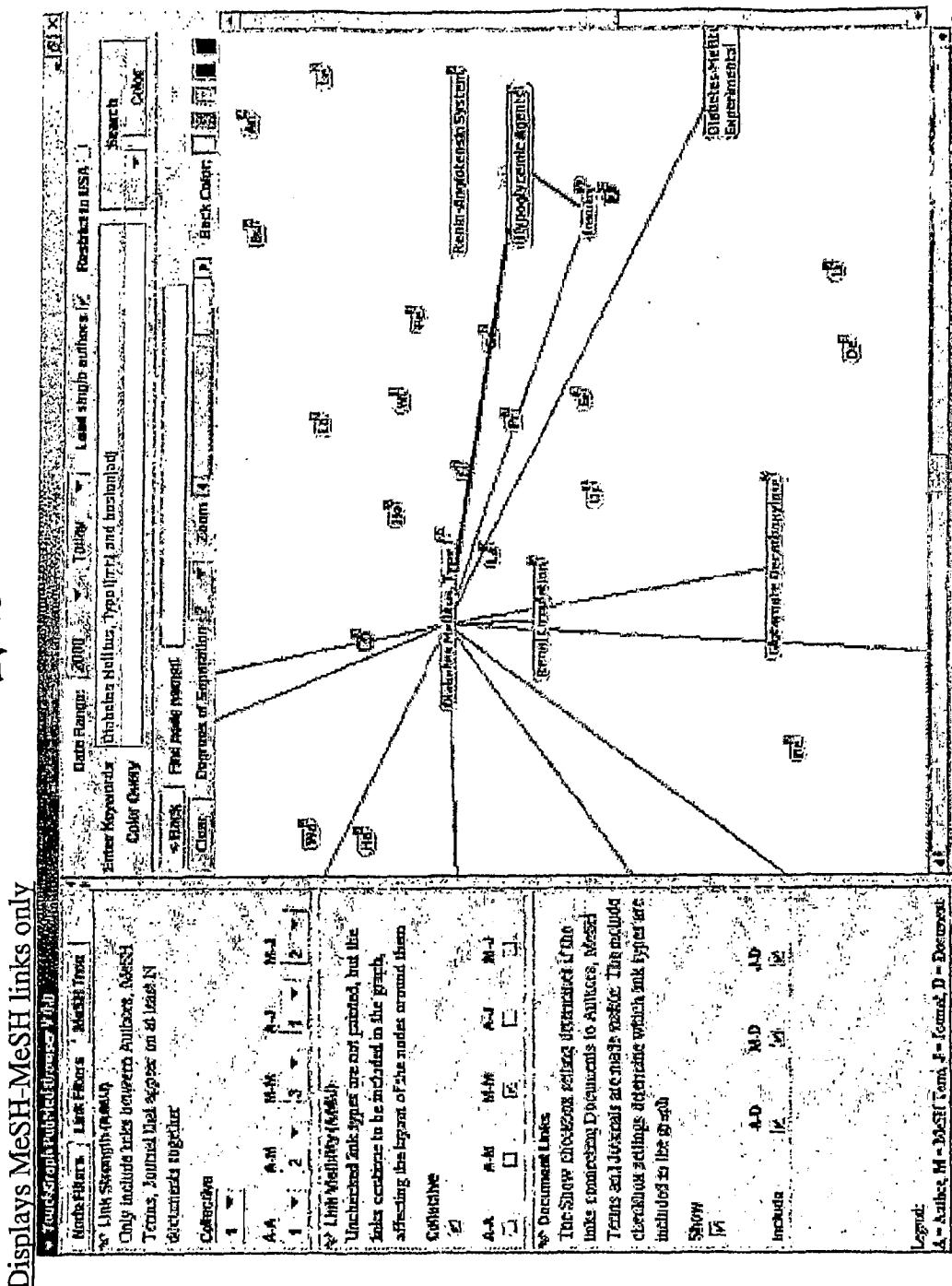


图 26

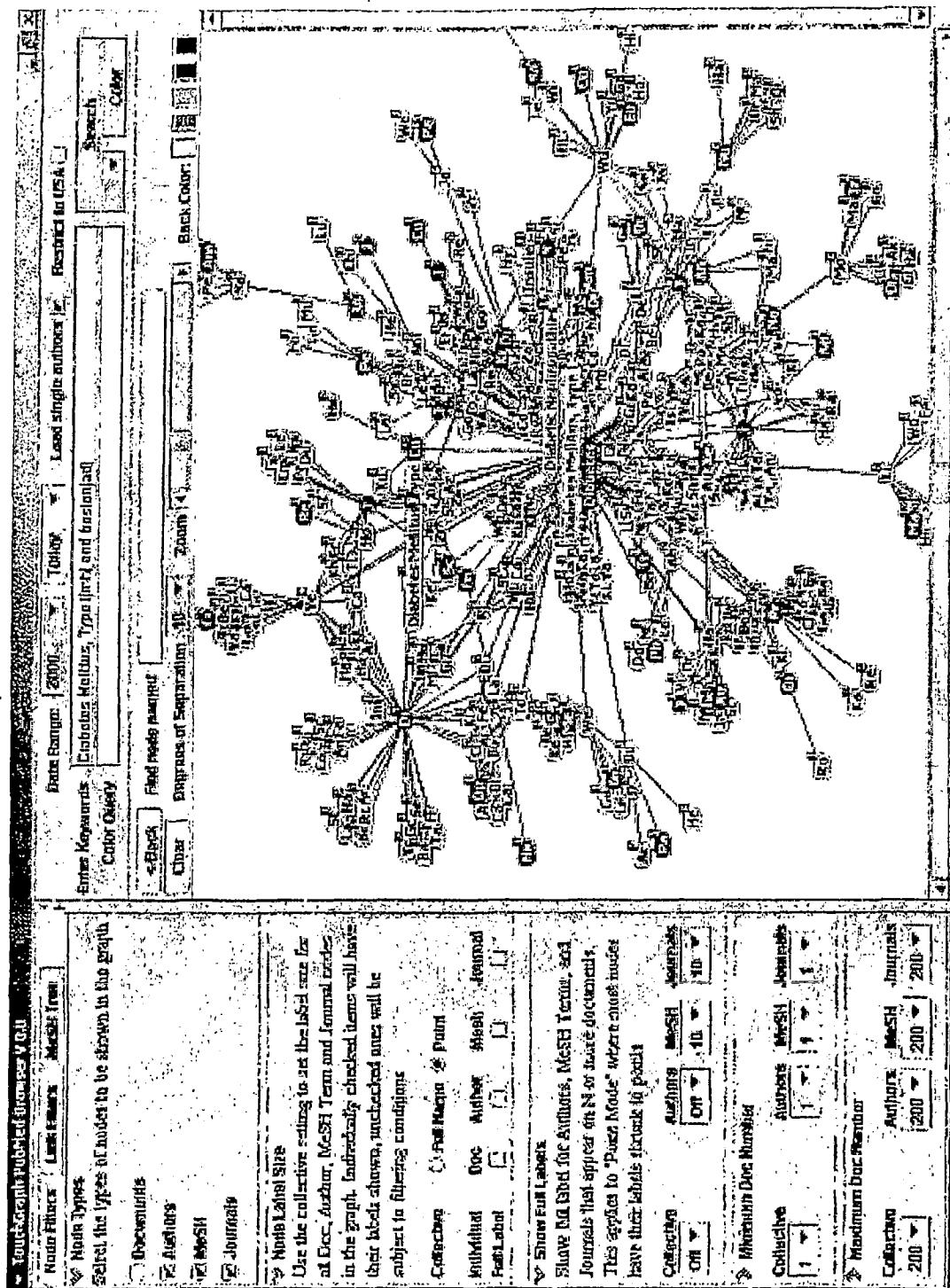


图 27

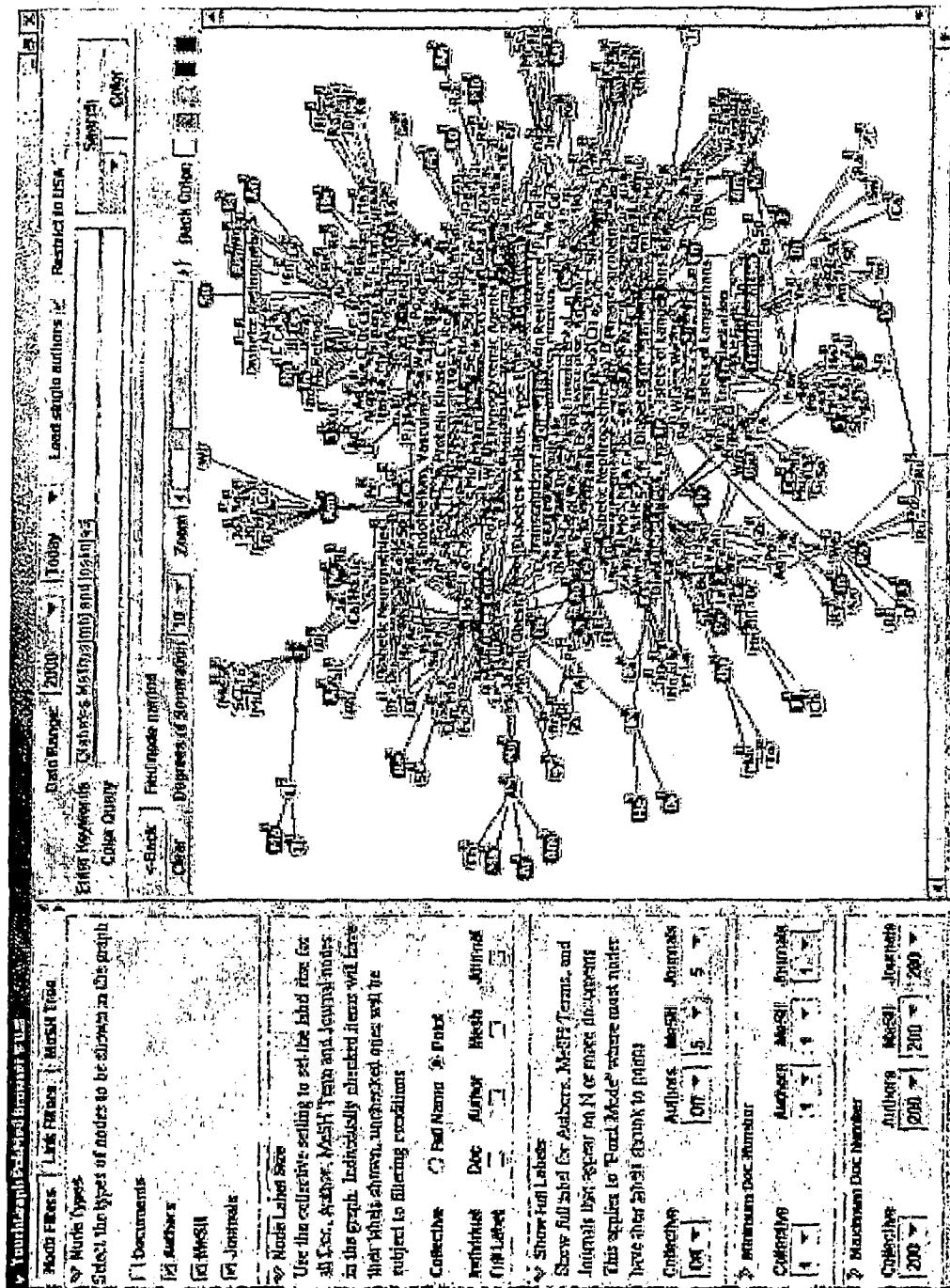


图 28

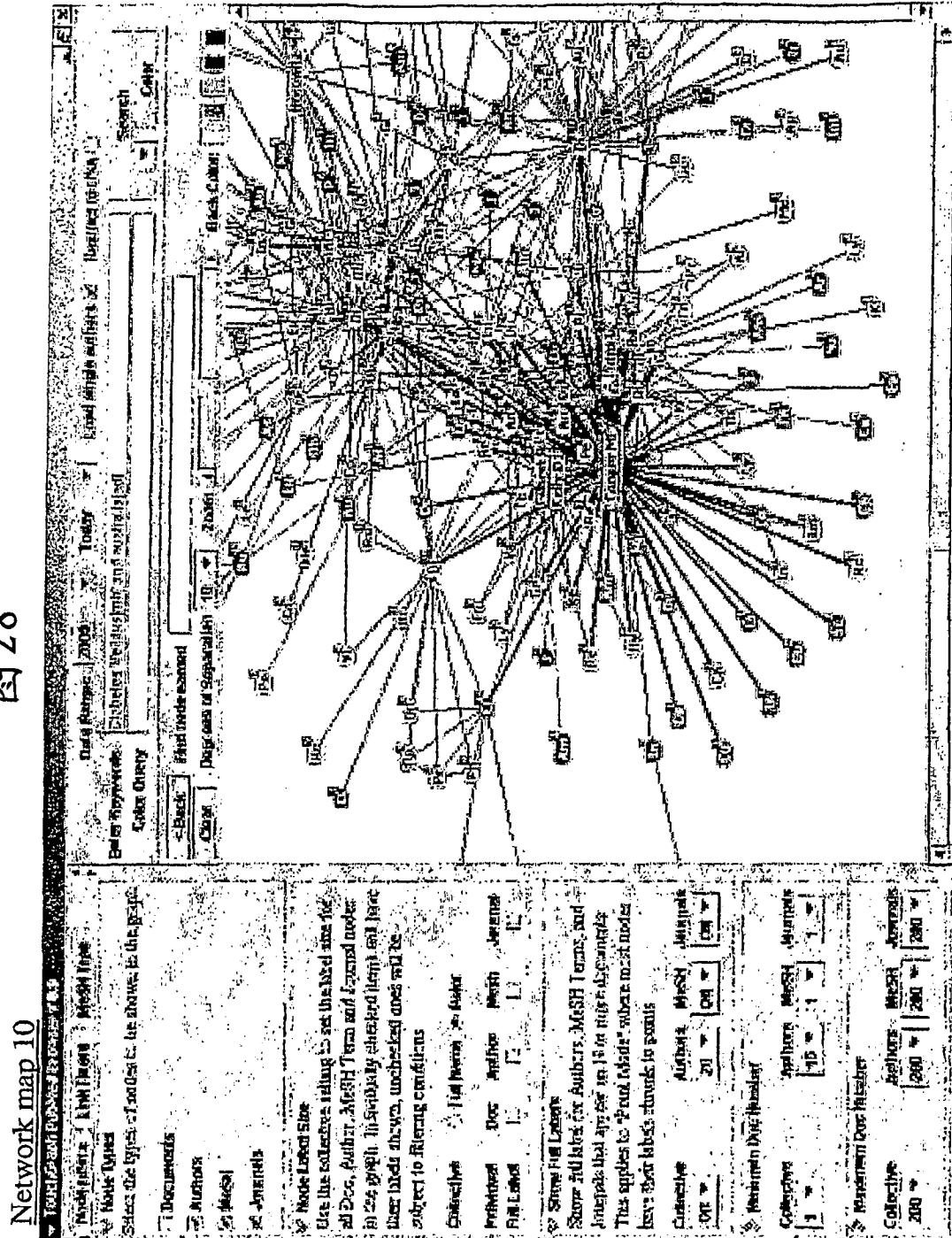


图 29

Network map 11

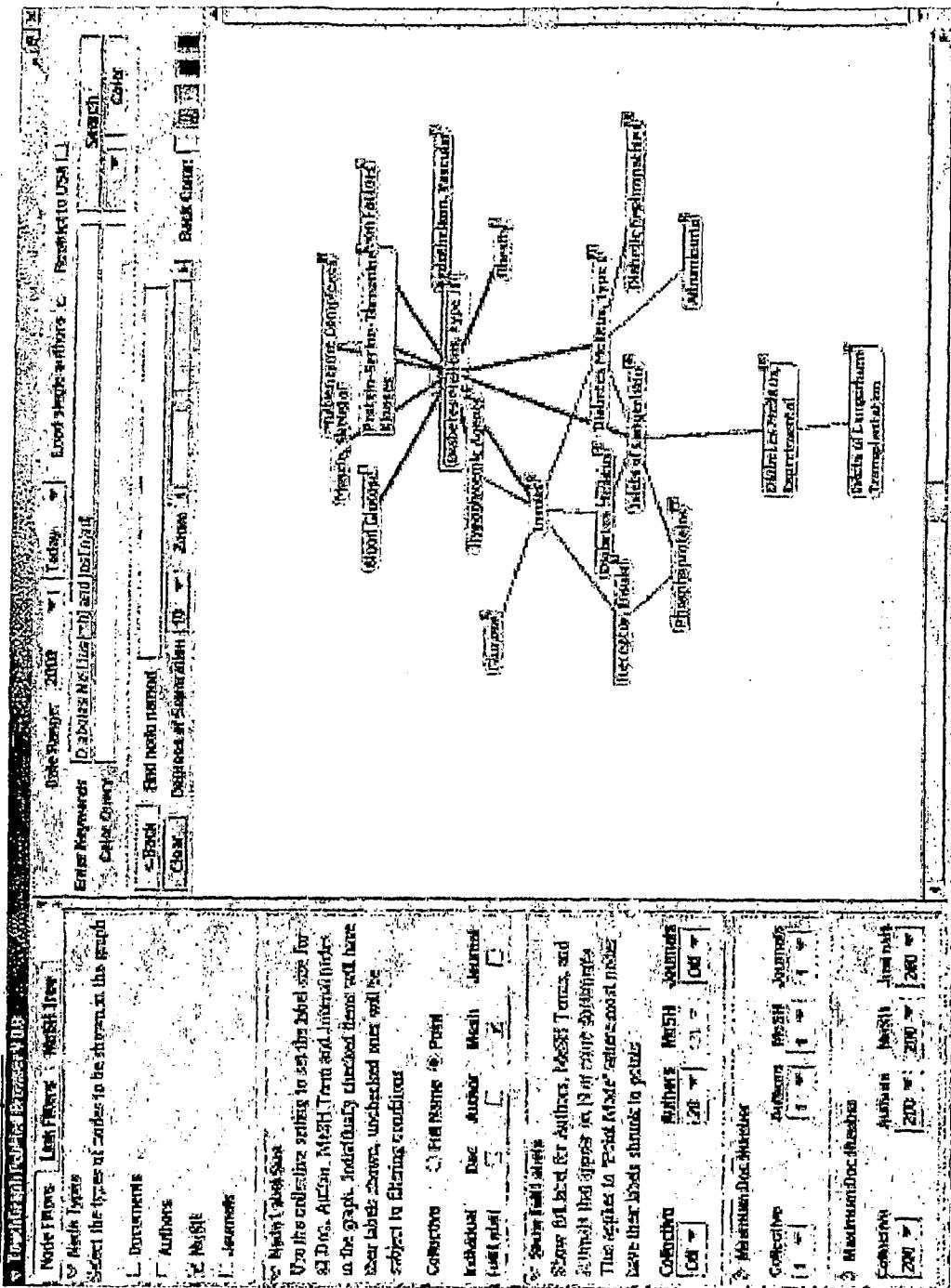


图 30

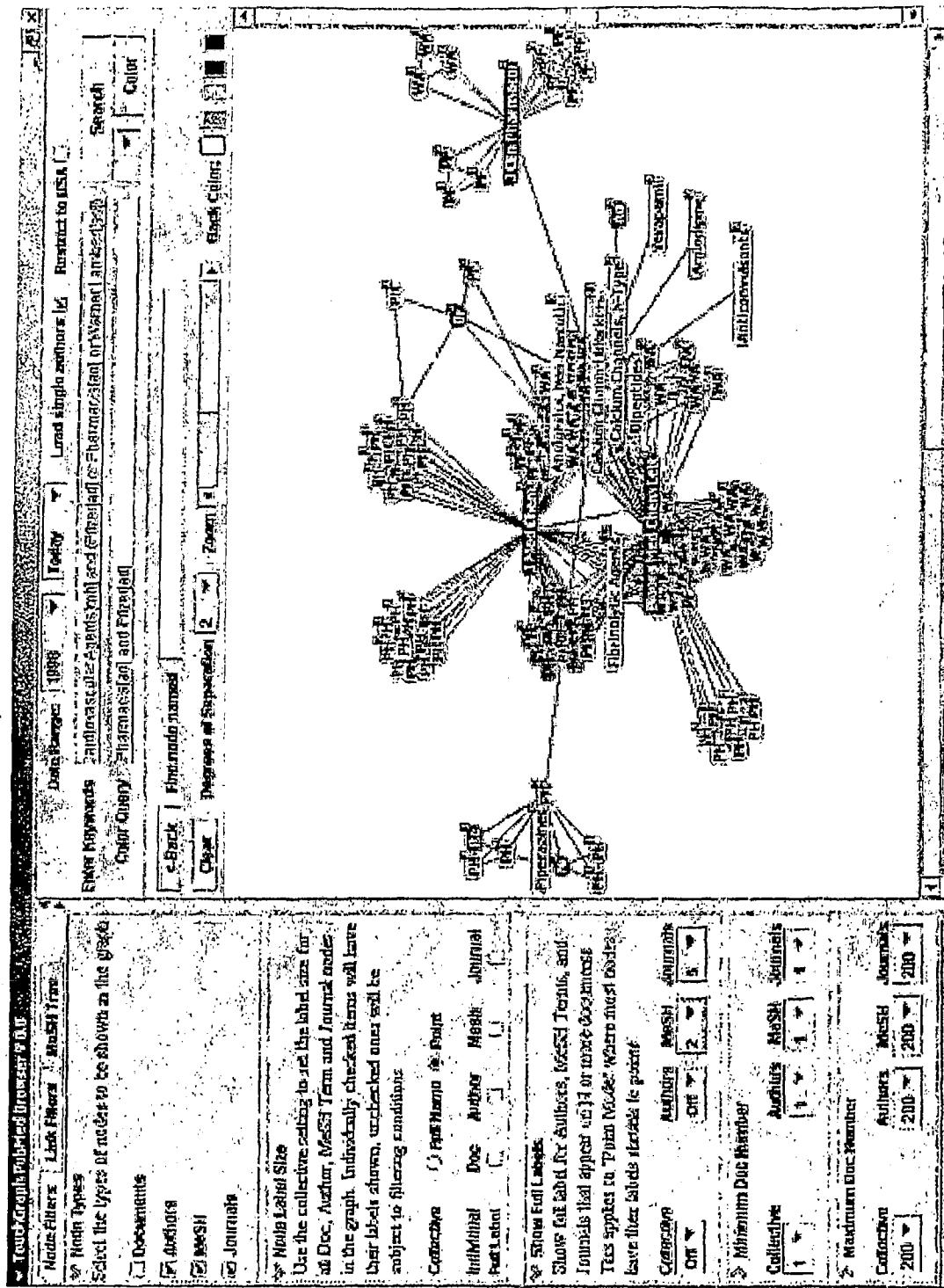


图 31

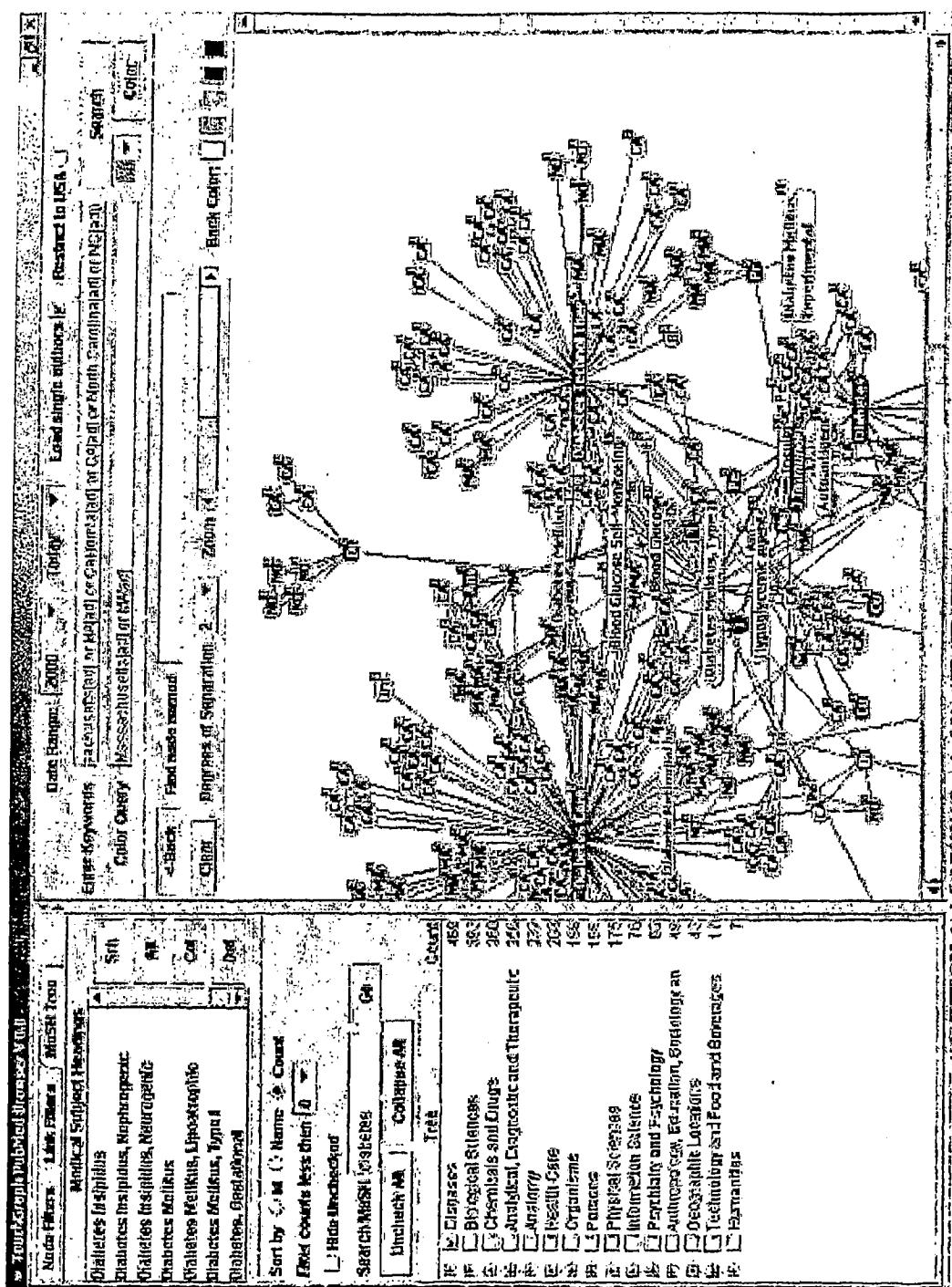


图 32

