问答系统综述与评测

报告人:王树西

2004-5-17

提　纲

1、问答系统综述

2、问答系统评测

3、引用的文章

问答系统综述

1、问答系统历史
2、聊天机器人
3、基于知识库的问答系统
4、问答式检索系统
5、基于自由文本的问答系统

问答系统历史

1、1950年，A.M.Turing提出“图灵测试”[1]；

2、John Searle提出“中国人房子问题（Chinese Room）”，质疑图灵测试对测定计算机智能的意义有多大[2]；

3、Searle否定不了图灵测试。陆汝钤院士认为，图灵测试永远不可能在图灵定义的层面上真正的实现[3][68]。

4、Jaseph Weizenbaum在1966年实现的“Eliza”，是第一个问答系统；

5、1969年，ROBERT F.SIMMONS就指出：数据集的规模不同，开发问答系统所用的方法也是不同的。更加难能可贵的是，ROBERT F.SIMMONS提出模式推理的思想[5]。

6、1991年，Hugh Loebner设立 “Loebner Prize”[39]；

7、1999年，TREC QA Track；

问答系统综述

1、问答系统历史
2、聊天机器人
3、基于知识库的问答系统
4、问答式检索系统
5、基于自由文本的问答系统

聊天机器人

1、典型的聊天机器人
2、聊天机器人相关网站、公司
3、Loebner Prize
4、聊天机器人小结

典型的聊天机器人

1、**ELIZA**[7]

用的是模式及关键字匹配和置换的方法，没有发展成一套系统的技术。例如，假设有关键字me和句型模式：(*you**me)，而且还有一个与该模式对应的置换规则：（what makes you think I ** you），那么，当输入句为“Yesterday you hurt me.”时，输出为“What makes you think I hurt you?”。ELIZA另外一个明显的弱点是：一旦发现用户输入中的一个关键字（词），便马上寻求答案，而不继续寻找用户句子中其它的关键字（词），这样使得一些优先级别比较低的模式几乎永远不会被匹配到 。

ELIZA下载地址之一:http://www.spaceports.com/~sjlaven/eliza.zip;

2、**FRED系列**[7][8]

多个版本：Agent Max Headcold、Milton Beasley、Barry DeFacto、FRED23、JFRED ；

于1998年和1999年，两次获得“Loebner Prize”比赛的冠军;

FRED可以从他人的会话中进行学习，然后把学来的经验用于自身的会话过程中。

FRED下载地址：http://www.spaceports.com/~sjlaven/fredtrek.zip

http://www.spaceports.com/~sjlaven/FRED23.exe

http://atlas.kennesaw.edu/~rgarner/robot/

典型的聊天机器人（续）

3、**ALICE**[7][41]

由宾夕法尼亚州Lehigh大学的Richard S.Wallac开发。获得2000年度、2001年度以及2002年度的“Loebner Prize”比赛冠军。它遵循GNU通用公共许可协议的开放源代码，有300多人对其发展做出了自己的贡献。

ALICE背后并没有复杂的算法，事实上，ALICE有40,000 多个模板，也是采用了模式匹配的方法来检索最合适的回答。但ALICE采用了一种很好的扩充机制，AIML文件可以进行内联，许多包含特殊领域知识的AIML文件可以方便的合并成一个更大的知识库。并且，ALICE通过对聊天记录进行分析，可以得到尚且没有明确回答的问题，并给出建议的模式。

Alice所在网址：http://www.alicebot.org/

4、**Eugene**[7][16]

Eugene是一个非常优秀的聊天机器人，它响应速度快、答案较为确切、对用户问题的理解较为准确、答案与用户问题比较相关、答案语句流畅且简短扼要，更加难得的是，它不但可以进行常识问答，而且可以进行知识问答，这说明它有一个知识库。

Eugene所在网址：http://www.mangoost.com/bot/bot.html

5、**Jabberwock** [7][17]

Jabberwock获得2003年“Loebner Prize”冠军，用户可以通过英语或者德语，与Jabberwock进行交谈。Jabberwock懂得20,000个单词，并且可以讲笑话和谜语。

Jabberwock所在网址：http://www.abenteuermedien.de/jabberwock/index.php

典型的聊天机器人（续）

6、**Talk-Bot** [7][19]

最初作为一个Internet在线聊天系统(Internet Relay Chatting，IRC)，Talk-Bot 是Chris Cowart 于1998年用javascript 和 PERL语言编写完成的，并于2001年和2002年两次获得“Chatterbox Challenge”比赛的冠军。Talk-Bot还有一个名为Kylie - TTS BOT 的“孪生妹妹”，可以用语音回答用户的问题。

Talk-Bot所在网址:http://www.frontiernet.net/~wcowart/chatbot.html

Kylie - TTS BOT所在网址：http://www.funonmars.com/kylie/kylie.html

7、**Ultra Hal** [7][28]

Ultra Hal是一个聊天机器人系列，有多个不同版本，包括：Original Web Hal、Speech-Enabled Web Hal、Teachable Web Hal、Dubya Bot、Hampy the Hamster Expert、Uncensored Teachable Web Hal、Create Your Own Chatter Bot。Ultra Hal是由从事自然语言处理（Natural-Language Processing, NLP）的公司Zabaware Inc开发的。

测试结果表明，Ultra Hal 响应速度快，言词贴切，是一个相当不错的聊天机器人。

Ultra Hal所在网址：http://www.zabaware.com/webhal/index.html

Ultra Hal下载地址：http://www.zabaware.com/assistant/download.html

典型的聊天机器人（续）

8、**Niall** [7]

Niall(The Non-Intelligent Acquired Language Learner)的作者是Mat Peck。系统运行之前，Niall的知识库是空的，但随着人机交谈的进行，Niall不断的从用户那里学习知识并存储起来，所以Niall的知识越来越多。例如，用户输入一条知识“Dog is an animal.”，然后紧接着提问“What’s dog?”，系统回答“Dog is an animal.”。虽然Niall还比较粗糙，但它代表了一个重要的研究方向：从会话中学习知识。

Niall 下载地址：http://www.spaceports.com/~sjlaven/niall.zip

9、**Cyber Ivar** [7][33]

Cyber Ivar是Jaczone的产品，Cyber Ivar向人们提供软件工程方面的知识，例如UML（Unified Modeling Language）。

测试结果表明，Cyber Ivar响应速度快，能正确理解用户问题，并且给出的答案准确、全面。在回答UML、WayPointer和 Jaczone之类的问题时，Cyber Ivar给出的答案相当准确、全面；对于常识性问题，它给出的答案也比较贴切。令人惊异的是，对于用户的提问“who is Maozedong?”，Cyber Ivar竟然回答：“he was a Chinese communist who reigned from 1949 until his death in 1983.”，在Cyber Ivar的知识库中，竟然有这么一条关于“Maozedong（毛泽东）”的知识（虽然这是一条错误的知识），这实在让人惊异，这说明Cyber Ivar的知识库还是挺大的。应该说，Cyber Ivar不但可以进行常识问答，而且可以进行知识问答。Cyber Ivar是一个相当优秀的聊天机器人。

Cyber Ivar 所在网址：http://www.jaczone.com/CyberIvar/cyberivar2.html

典型的聊天机器人（续）

10、**产品宣传机器人**

这类机器人的特点就是为公司的产品进行宣传。把这种机器人放到了网上，就好象商店门口摆了机器人招揽生意一样。它们有些能和网页相互作用，在一个frame里聊天，聊到公司负责人或者产品，就在另外的frame里调出相应的网页；有些把语音合成的功能结合了进来，和你谈的所有话都是“说”出来的。 如：

Kate[7][38]

Kate是一个为Ford Motor Company公司作宣传、进行客户服务的聊天机器人，它可以回答诸如：“How do I change the oil in my 2002 Explorer?”、“Where is my nearest Ford dealer?”等问题。用户可以向Kate提问有关Ford、 Lincoln-Mercury等产品及其相关服务的问题。Kate回答用户的问题，往往采用链接的方式，让用户通过链接自己寻找答案。如果用户输入的问题形式上比较规范（例如下列形式：“What kind of vehicles do you make?”），那么Kate的回答会较为准确。

Kate所在网址：http://www.customersaskford.com/kate/crc_frame.asp

11、**其它聊天机器人（ChatBot）**

Chatbox 、Hex 、Cybelle 、Parry 等，有几十个之多。

聊天机器人

1、典型的聊天机器人

2、聊天机器人相关网站、公司

3、Loebner Prize

4、聊天机器人小结

聊天机器人相关网站

聊天机器人很多，上面提到的只是其中的一些，篇幅所限,难免有所疏漏。还有很多介绍聊天机器人的网站，例如：

http://www.i-une.com/search.php?q=Chatterbots

http://www.botspot.com/search/s-chat.htm

http://www-personal.umich.edu/~abney/ling492/systems.html

提供聊天机器人的公司

有许多公司提供聊天机器人的相关产品与服务，例如：

◎ **Artificial Life**(http://www.artificial-life.com/)

◎ **Extempo**(http://www.extempo.com/)

◎ **Kiwilogic**(http://www.kiwilogic.com/)

◎ **Maybot**(http://www.maybot.com/)

◎ **NativeMinds**(http://www.nativeminds.com/)

◎ **Robitron Software Research**(http://www.robitron.com/)

◎ **Petamem**(http://www.petamem.com/)

◎ **Versality.com**(http://www.versality.com/)

◎ **Virtual Personalities**(http://www.vperson.com/)

◎ **Zabaware**(http://www.zabaware.com/)

聊天机器人

1、典型的聊天机器人
2、聊天机器人相关网站、公司
3、Loebner Prize
4、聊天机器人小结

Loebner Prize

1991年，Hugh Loebner决定对“图灵测试”进行实际操作。他设立了“Loebner Prize”，悬赏$100,000，奖励首次通过图灵测试的人。因为“图灵测试”的通过有待时日，所以，对每年一度“Loebner Prize”比赛的冠军，奖励$2,000。迄今为止，没有任何一个程序通过“图灵测试”。

“Loebner Prize”设立以来，许多程序参加了比赛，也产生了许多著名的聊天机器人程序。下面是历届“Loebner Prize”的赛况。

Loebner Prize（续）

时间	参赛系统中的“冠军”	作者
1991 年	PC Therapist	Joseph Weintraub
1992 年	PC Therapist	Joseph Weintraub
1993 年	PC Therapist	Joseph Weintraub
1994 年	TIPS	Thomas Whalen
1995 年	PC Therapist	Joseph Weintraub
1996 年	Hex	Jason Hutchens
1997 年	Converse	David Levy
1998 年	Albert	Robby Garner
1999 年	Albert	Robby Garner
2000 年	ALICE	Richard Wallace
2001 年	ALICE	Richard Wallace
2002 年	Ella	Kevin Copple
2003 年	Jabberwock	Juergen Pirner

聊天机器人

1、典型的聊天机器人
2、聊天机器人相关网站、公司
3、Loebner Prize
4、聊天机器人小结

聊天机器人小结

它们的背后几乎没有经过严格证明的定理、晦涩的数学公式，甚至没有复杂的算法，事实上它们几乎全部采用模式匹配的方法，来寻找问题最合适的答案。它们有一个共同的特点，那就是在与用户的交谈过程中，都是基于谈话技巧和程序技巧，而不是根据常识。在它们的对话库中，可以存放多个句型、模板，但几乎没有常识库，而要成功的进行图灵测试，除了要研究计算机对话过程中的语义和语用分析以外，一个很重要的手段，就是要使计算机拥有足够多的常识知识，并具有联想能力[3]。

对于知道答案的问题，聊天机器人往往给出人性化的回答；对于不知道答案的问题，有三种回答方法：①猜一个答案；②老实说不知道；③用转移话题的办法回避[3]。目前的聊天机器人，因为其知识库规模有限、甚至没有知识库，所以面对用户提出的许多专业性问题，用的就是第三种方法，也就是用转移话题的办法回避。无庸讳言，正是因为知识的缺乏，聊天机器人目前还没有太多的实际用途，除了和用户“聊天”，它并不能解决太多的实际问题，许多情况下，它更象是玩具而非工具。但是聊天机器人所体现出来的一些技术特色，却值得我们注意，例如，正确理解用户的询问，从用户的输入中进行学习，使得答案更加人性化、更合乎逻辑。

问答系统综述

1、问答系统历史
2、聊天机器人
3、基于知识库的问答系统
4、问答式检索系统
5、基于自由文本的问答系统

基于知识库的问答系统

1、定义；
2、知识库；
3、各类基于知识库的问答系统；
4、小结

定义

拥有一个或多个知识库，并利用检索、推理等技术，来理解与求解用户问题的问答系统，称为基于知识库的问答系统。一般来说，知识的数量与质量是一个基于知识库的问答系统性能是否优越的决定性因素，因此，基于知识库的问答系统的主要特征是有一个或者多个知识库，其中存储一个或者多个领域的知识。

基于知识库的问答系统

1、定义；
2、知识库；
3、各类基于知识库的问答系统；
4、小结

知识库

1、知识与信息不一样，知识是信息经过加工整理、解释、挑选和改造而形成的[48]。

2、一阶逻辑公式、内涵逻辑公式、语义网、框架、脚本、概念图、OO乃至知网、HNC等等，都是知识表示的不同方案。比较知识表示方案之间优劣的准则有三个：可表示性、表达能力、相应推理的可行性 [86]。

3、Feigenbaum：强调知识的大量积累可以使计算机的智能发生质变 [3]。

4、在Feigenbaum思想的影响下，Doug Lenat于1984年在“美国微电子和计算机技术公司（MCC）”发起了CYC研究项目，于1995年结束。CYC耗费了200人年的工作量，建立起一个拥有50万断言的知识库，并在此基础之上研究了自然语言理解、学习、问题求解等人类智能活动的机理。从1995年1月起，CYC课题组变成了一个国际闻名的Cycorp公司，成为许多重要企业投资的对象。

5、中国科学院从20世纪80年代起，即组织有关科学知识库的研究，并拨专项经费支持。这一项目自１９８６年起开始建设，迄今中国科学院已有４５个研究所参加科学数据库的建设和服务，截至去年１０月底，科学数据库总数据量达到８２００亿个字节，其中上网数据量４３００亿个字节[75]。

知识库（续）

6、现有的知识库还存在一些问题，如：知识获取的瓶颈问题、知识库维护的困难等。此外就是知识库的完备性和不确定性问题。

7、**知识库建设中的本体（Ontology）**

当前，建立本体大部分还是采用手工方式，建立本体还远远没有成为一种工程性的活动，每个本体开发组都有自己的原则、设计标准和定义的开发阶段，也正是由于缺少大家都认同和遵循的开发方法，目前，对本体的共享、重用和互操作还难以实现[3]。

知识工程领域对本体论的研究目前还刚刚起步，很不深入，也很不成熟；另一方面从本体论研究的前景来看，在未来的几年内，它将会吸引更多的研究者，并将会成为一个异常活跃的领域[3]。

基于知识库的问答系统

1、定义；
2、知识库；
3、各类基于知识库的问答系统；
4、小结

基于知识库的问答系统

现有基于知识库的问答系统，大致包括：基于本体的问答系统、自然语言界面的专家系统、基于受限语言的的数据库查询系统，等等。所有这些问答系统都有共同的一个特点：它们都是基于一个或者多个知识库（数据库），通过自然语言的形式与用户进行交流。和聊天机器人不同的是，这类系统擅长于知识问答，对于不能回答的问题，就老实回答说“不知道”，而非故意转移话题。

1、基于本体的问答系统；
2、自然语言界面的专家系统 ；
3、基于受限语言的的数据库查询系统 ；
4、基于FAQ的问答系统 ；

基于本体的问答系统

1、陆汝钤院士主持开发的“Pangu”

对话者:“动物园有一头黑熊死了。” // 计算机:“黑熊是怎么死的？”

对话者:“据说黑熊是吃塑料袋死的。” // 计算机:“准是哪个不文明的游客投 进去的。”

对话者:“是啊, 太可气了。” // 计算机:“饲养员怎么不阻止呢？”

2、由曹存根研究员率领的课题组，正在建立NKI（国家知识基础设施）海量知识库，并在此基础上搭建NKI问答系统；

3、在宋柔教授的率领下，北京语言文化大学计算机系语言信息处理研究所正在进行百科词典的知识提取工作，处理对象限定为行文较规范的《中国大百科全书》（光盘版），并且目前只提取比较易于形式化的信息。

自然语言界面的专家系统

现在已知的专家系统，有1968年Feigenbaum等人于斯坦福大学建成的DENDRAL，细菌感染诊断专家系统MYCIN、探矿专家系统Prospector，MIT大学开发的数学符号运算专家系统MACSYMA、CMU大学开发的语音识别专家系统HEARSAY、HEARSAY-II、HEARSAY-III，匹兹堡大学的H.E.Pople等人开发的内科病诊断咨询系统INTERNIST，斯坦福大学开发的AM专家系统、肺功能测试专家系统PUFF，DEC公司与CMU大学开发的XCON(R1)、青光眼诊断与治疗专家系统CASNET等。Feigenbaum本人在1988年作了一次调查，根据他所掌握的情况，当时投入运行的专家系统，约有2000个，分布在欧美和日本，这里没有包括发展中国家的数字[44][46]。

现在已知的专家系统开发工具，有各种专家系统语言：PROLOG语言、SMALLTALK语言、LISP语言等；专家系统外壳（又称骨架系统）：CLIPS、EMYCIN、KAS、EXPERT等；通用型专家系统工具：OPS5、ART等[47][48][49]。

基于受限语言的的数据库查询系统

基于受限语言的数据库查询系统，是指使用受限的自然语言，对数据库进行查询的系统，它的关键步骤是要将中文查询句转换为数据库的SQL语句。信息以关系数据库的方式存储，自然语言界面，用户用自然语言进行查询，查询结果以自然语言、表格、图形等方式返回。

近年来，国内研制出很多相关系统，如RCHIQL、NCHIQL、NLCQI等[40]。他们所用的是类似于语法和模板的技术，由于查询的对象是数据库，所以大部分系统都充分利用了ER模型[82]。

基于FAQ的问答系统

1、与产生式、语义网络和框架等传统知识表示相比，FAQ库中的知识，虽然也是经过人工处理的，但是处理方法却并非基于上述符号处理机制，而是采用自问自答的方式，知识描述的颗粒（粒度）很大、很粗，属于半结构化文本；

2、基于FAQ的问答系统，关键在于计算用户查询和FAQ知识库中问题的相似度，从而找到FAQ知识库中与用户查询最为相似的问题，然后把此问题对应的相关答案直接提交给用户，这是一个句子相似度计算的过程；

3、计算句子相似度的方法有很多，大致说来，可分为两种。第一种方法，不考虑两个句子的语义信息，而是直接利用模式匹配技术、关键字（词）匹配技术、基于向量空间模型的TF/IDF方法等，来计算两个句子的相似度；

第二种方法，考虑两个句子的语义信息，利用WordNet、HowNet、同义词词林等语义知识资源，计算用户查询与FAQ知识库中所有问题的语义相似度，如果语义相似度的值均小于阈值M，那么就可以认为FAQ知识库中没有用户所问的问题，否则将FAQ知识库中与用户查询语义相似度最高的问题所对应的答案，提交给用户。

基于知识库的问答系统

1、定义；
2、知识库；
3、各类基于知识库的问答系统；
4、小结

小结

基于知识库的问答系统，其优点是显而易见的。上面提到的基于知识库的问答系统，性能优良，对于用户提出的许多问题，回答准确，甚至可以进行一定程度的推理计算。并且由于是基于知识库的，所以系统具有良好的可扩展性。

但是其局限性同样明显。如果用户的问题落入系统的知识库范围之内，系统可以轻松的解决问题；一旦超出这个范围，系统性能很快下降为零。总之，目前这类系统的性能象一个窄的尖峰，适用范围非常狭窄。从知识库的角度分析其弱点的来源,可以发现系统的知识库规模不足、知识获取困难，如前所述，存在知识库的瓶颈问题。例如，依靠无限加大知识库的CYC计划只停靠在探索阶段[3]。

问答系统综述

1、问答系统历史
2、聊天机器人
3、基于知识库的问答系统
4、问答式检索系统
5、基于自由文本的问答系统

问答式检索系统

根据以自然语言方式提交的用户查询，从系统文档集合或WWW中，检索出相关文本或网页，并将其返还给用户，这种系统称为问答式检索系统，也称问答式搜索引擎、智能搜索引擎。

作为一种界面友好的信息检索系统，问答式检索系统需要做的，就是正确理解自然语言形式的用户查询，充分领会用户的查询意图，并检索出与用户需求最相关的文本或者网页。相应的，它所需要的技术主要包括两个：用户查询处理技术、信息检索技术。

问答式检索系统（续）

基于WWW的问答式检索系统的典型流程如下：

① 从WWW上采集大量的网页并加以索引，存入数据库。由于系统不断从WWW采集新的网页以及对原有网页进行更新，所以系统的数据库是不断更新的。这是一个预处理的过程。

② 接受用户用自然语言提交的查询；

③ 对用户查询进行分析，将其从自然语言形式转换为系统的内部表示；

④ 利用内部表示的用户查询，对系统的数据库进行检索；

⑤ 如果检索结果为空，那么向用户说明，转②；

⑥ 如果检索结果只是一个网页，那么直接将它提交给用户，转②；

⑦ 如果检索结果是多个网页，那么按照一定规则将它们进行排序并提交给用户，转②。

问答式检索系统（续）

为了开发一个上述通用的、面向所有领域的问答式检索系统，至少需要通过网页采集器(Crawler)建立一个较大的网页数据库，并且这个网页数据库必须不断的更新，这是工作量很大并且代价昂贵的一项任务。为了减少工作量，快速开发，许多问答式检索系统是基于现有搜索引擎开发的，具体流程如下：

① 接受用户用自然语言提交的查询；

② 对用户查询进行分析，理解用户的意图；

③ 抽取出用户查询中的关键词，并利用语义词典(WordNet、HowNet等)，进行关键词扩展；

④ 将上述关键词提交给搜索引擎(如Google等)，进行检索；

⑤ 将搜索引擎返还的前N个网页提交给用户；

⑥ 转①。

这种问答式检索系统，所做的工作主要是对用户查询进行分析，抽取出其中的关键词，并进行关键词扩展。

现有典型的问答式检索系统

（1）Start [7][18]

Start是世界上第一个基于Web的问答系统，自从1993年12月开始，它持续在线运行至今。Start的作者，是MIT人工智能实验室InfoLab Group的Boris Katz及其同事。

不同于信息检索系统（例如搜索引擎），Start旨在提供给用户“准确的信息”，而不是仅提供一系列简单的链接。现在，Start能够回答数百万的多类英语问题，包括“place”类（城市、国家、湖泊、天气、地图、人口统计学、政治和经济等）、电影类（片名、演员和导演等）、人物类（出生日期、传记等）、词典定义类等。

Start是一个基于知识库的问答式搜索引擎，系统包含两个知识库（“START KB”、“Internet Public Library”）以及一个搜索引擎。如果通过这两个知识库就能回答用户的问题，那么系统立刻给出准确的答案；否则，首先解析用户输入，得到其中的关键词，然后利用这些关键词，通过系统自身的搜索引擎进行检索，最后将得到所有相关文本，以链接的形式提交给用户，供用户点击并在打开的网页中自行寻找答案。

测试结果表明，Start是一个优秀的问答系统。

Start所在网址：http://www.ai.mit.edu/projects/infolab/start.html

现有典型的问答式检索系统（续）

（2）AnswerBus

对于每一次用户查询，AnswerBus返还5个网页链接，并给出XML和TXT格式的“Possible answers”。

AnswerBus所在网址：http://www.answerbus.com/index.shtml

（3）IONAUT

为了更加正确理解用户的查询意图，IONAUT建议用户查询以“who”、“where”、“when”等关键词开头，从而标明需求的类别。

对于每一次用户查询，IONAUT返还10个网页链接（Top Ten Documents），并且对每个网页的内容进行扼要介绍。

IONAUT所在网址：http://www.ionaut.com:8400/

（4）LAMP

为了更加正确理解用户的查询意图，LAMP列出person、organization、location、date、time、money、percent等7种查询类别，让用户进行选择。与上述各系统不同，LAMP并不是返还网页链接，而是直接返还答案。

LAMP所在网址：http://hal.comp.nus.edu.sg/cgi-bin/smadellz/lamp_query.pl

现有典型的问答式检索系统（续）

(5)QuASM

QuASM(Question Answering using Semi-Structured Metadata)不是用来回答推理类、过程类问题的，并且答案可能是过时的。这是因为，它的数据库是固定的，其中的数据来源于网站www.fedstats.gov及其相关链接，并且这些数据是经过人工处理的。

QuASM所在网址:http://ciir.cs.umass.edu/~reu2/

(6)AskJeeves

对于自然语言形式的用户查询，AskJeeves返还文本形式的答案，并给出答案的来源--包含答案的网页链接。例如，对于问题“Who is President Clinton?”，系统准确的回答:“Bill Clinton spent the 1970s as a law professor and then… …”，并且附上Clinton总统的照片，以及答案的来源:包含答案的网页链接。

(7)Encarta

Encarta是微软的产品。Encarta回答问题较快，答案也比较准确。

Encarta所在网址:http://encarta.msn.com/

(8)“小灵通”[56]

“小灵通”是中科院计算所开发的一个中文问答式检索系统，系统的关键模块包括：模板库，用户查询分析器，检索-排序器，管理程序等。其中的核心是模板库。“小灵通”虽然基于Web，但并非面向所有领域，而是面向旅游领域。

问答式检索系统技术分析

一般来说，问答式检索系统主要需要两种技术：用户查询处理技术、信息检索(IR)技术。由于信息检索(IR)技术目前已经比较成熟，所以不再赘述，这里主要讨论问答式检索系统所面临的第一个问题：如何正确理解用户用自然语言提出的查询。

如果存在一个优秀的自然语言理解系统，可以准确无误的理解用户查询的语义，问题自然解决了。问题是，如此优秀的自然语言理解系统，到现在为止还没有出现，所以我们不得不寻求其它的替代解决方案。

第一种方法，对用户查询进行浅层分析，识别出其中的关键词，然后利用查询扩展技术，借助HowNet、WordNet等语义词典，将关键词的同义词、近义词等高度相关的词，一并提交给后继的检索系统。这种方法，虽然允许用户使用自然语言查询，但并没有充分利用用户查询的信息，特别是语义信息，其能力等同于一般的词表法。许多号称自然语言查询的系统其实都是这么做的。

第二种方法，就是使用问句模板。AskJeeves就是这样的系统，系统拥有一个问句模板库，并且为这些问句模板构造有页面作为答案；“小灵通”拥有237个模板，覆盖了大部分旅游常见的问题类型。如果系统面向的领域非常狭窄，如“小灵通”，那么这种方法的好处是显而易见的，数量很少的问句模板就可以覆盖绝大多数的用户提问方式。但如果系统面向的领域非常广阔，甚至是开放领域，那么仍然利用这种方法的话，模板库和模板答案的数量就非常多，由于模板库和模板答案一般是人工产生和维护的，所以工作量非常大，实际上，AskJeeves雇佣了数百专职人员来完成该任务。

问答式检索系统小结

目前信息检索(IR)技术日臻成熟，现有搜索引擎(如Google)的检索效果(准确率、召回率)也越来越令人满意，一般来说，只要输入几个关键词，Google就能够检索出用户所需要的网页，准确率相当高。应该说，在信息检索方面，Google比现有绝大多数问答式检索系统都要高明的多。

问答式检索系统可以接受用户以自然语言方式提交的查询请求，实际上，Google基本上也可以做到这一点。

从上述技术分析可以看出，在处理用户查询方面，问答式检索系统主要是抽取用户查询中的关键词，以及利用语义词典对关键词进行扩展，从而得到一组描述用户需求的关键词。从用户的角度出发，这一点往往是没有必要的，因为最清楚用户需求的是用户自己而非系统，所以在描述用户需求方面，用户自己给出的关键词自然要比系统分析得出的关键词准确的多。

综上所述，与现有搜索引擎相比，问答式检索系统的优势并不明显。一系列测试结果表明，无论是速度还是准确性，Google都要比现有大多数问答式检索系统高出许多。并且，问答式检索系统返还给用户的，只是和用户查询相关的文本或者网页，而不是用户问题的具体答案，所以严格说来，问答式检索系统不算是一个真正的问答系统，而是一个信息检索系统。目前实用的问答式检索系统很少，并且这方面的文章也不多。应该说，这个研究方向并不被广泛的看好。

问答系统综述

1、问答系统历史
2、聊天机器人
3、基于知识库的问答系统
4、问答式检索系统
5、基于自由文本的问答系统

基于自由文本的问答系统

所谓自由文本，又称原始文本、非结构化文本，是指未经人工处理的文档、网页等。基于自由文本的问答系统，是指这样的问答系统，它接受用户以自然语言提交的问题，然后利用信息检索(Information Retrieval , IR)等技术，从系统的自由文本库中检索出相关的文档、网页，最后利用答案抽取等技术，从这些检索出来的自由文本中抽取出问题的答案并提交给用户。

基于自由文本的问答系统，基本上分为三个步骤，首先，处理用户查询;其次，检索相关的自由文本;最后，从自由文本中抽取答案。它所涉及到的技术，包括信息检索(IR)技术、信息抽取(IE)技术、文本挖掘、Ontology、文本聚类、文本摘要、个性化信息需求建模、模式推理和几乎所有的基础NLP技术(词法分析、句法分析、语义分析、指代解析、消岐等)，是这些技术的集大成者。

与上述各类问答系统相比，基于自由文本的问答系统，不需要建立大规模知识库，而是基于自由文本进行知识问答，节省了大量的人力物力；并且系统返还给用户的，是用户问题的具体答案而不只是和用户查询相关的文本或者网页。应该说，基于自由文本的问答系统，代表着问答系统的发展方向，现有的绝大多数问答系统，都属于这一类，特别是基于WWW的开放领域问答系统，更是研究的热点。

基于WWW的问答系统

基于WWW的问答系统，又称为基于Web的问答系统、Web问答系统，是指基于WWW开发的问答系统。WWW[55]是World Wide Web的简称，是由CERN[54]的Tim Berners-Lee于1991年开发的。WWW只有十余年的历史，却已经得到很大的发展，现在的WWW规模庞大，已经拥有数十亿网页和数千万网址，并且这个数字正以惊人的速度增长。WWW已经成为一个全球性的信息基础设施，对于满足人们的信息需求来说，这是一个理想的资源。毫无疑问，WWW是一个开放领域的知识库，那么，基于WWW的问答系统，就是开放领域的问答系统。

传统的信息检索系统，也就是基于web的搜索引擎，诸如“GOOGLE”和"NorthernLight",返回的是和用户查询相关的文本，而不是用户问题的答案，和传统的信息检索系统不同，Web问答系统能够精确回答用户用自然语言提出的问题，以信息点而不是包含答案信息的文档作为返回结果。作为比一般信息检索更进一步的研究，基于Web的问答系统同样面临海量的问题，但更强调精确性[57]。面向英文的问答技术正在迅速走向实用，但是，目前的问答系统大多局限在某个特定领域或者特定范围之内，能够回答的问题类型也比较简单，真正的面向Web开放域的问答系统的正确率和精确性都不高，还不能提供良好的商业服务[57]。

现有各种Web问答系统

(1) 跨语言问答系统Marsha[84]

汉语问答系统Marsha，采用了类似于TREC问答系统的技术，Marsha包括三个主要模块：查询处理模块；汉语搜索引擎；答案抽取模块。Marsha也包括汉语处理的一些特定技术，诸如分词、序数处理，并采用TREC QA Track的评测机制，对本系统的性能进行评价。测试结果表明，Marsha和TREC－8 QA Track中的英语问答系统具有很大的可比性。Marsha的英语版本进一步表明：Marsha中所用到的启发式方法对于英语问答系统具有可适用性。

(2) **NSIR**[60]

NSIR是一个较为典型的Web问答系统。NSIR的运行过程包括5个步骤：查询调整、文本检索、段落抽取、短语抽取，答案排序。本系统采用了Probabilistic Phrase Reranking (PPR)算法，利用近似性和问题类型特征，在TREC8文本集合中，得到20％的交互文本。

(3)Textract[61]

Textract是Cymfony公司的产IE品，Textract参加TREC-8 QA Track比赛，获得较好的成绩，回答问题的正确率达到66%。Textract的主要目的之一，是检测在诸如QA的应用中，信息抽取(IE)是如何有助于信息检索(IR)的。

现有各种Web问答系统

(4)MULDER[89]

MULDER是一个优秀的基于Web的开放领域问答系统。MULDER的体系结构，包括：（1）问题解析（Question Parsing）；（2）问题分类（Question Classifier）；（3）查询公式化（Query Formulation）；（4）将用户查询提交给搜索引擎Google进行检索，得到相关文本；（5）使用自然语言分析器对Google的返回结果进行语法和语义分析，抽取候选答案（Answer Extraction）；（6）采用启发式策略进行答案选择（Answer Selection），将可信度最高的答案返回给用户。

(5)Webclopedia[90]

Webclopedia 致力于对各种语言提出的问题，从多种语言的文本集合或者Web中，抽取出答案。Webclopedia的接口包括:CONTEX问题分析器;查询构成模块;MG信息检索器;三个文本分离器;BBN的命名实体识别器IdentiFinder;

(6) Aranea[103]

Aranea利用数据标注和数据挖掘技术，从WEB上抽取答案。数据标注，利用半结构化的数据库技术，对于回答一般的问题十分有效。Aranea的测试结果表明，数据挖掘技术用于问答系统，可以减轻WEB上面数据量过大的问题，从而解决许多自然语言处理的问题。

Web问答系统(续)

(7)**AQUAINT**

在ARDA 的AQUAINT (Advanced QUestion Answering for INTelligence)资助下，Columbia 大学和Colorado大学正在合作研究的问答系统。目标是建立一个能够回答复杂问题的系统，所谓复杂的问题，需要和用户交互，以及记忆问题的上下文，答案或许存在于文本的非同类数据库，需要结合、综述不同来源的信息。

(8)IBMPQ

IBM用统计的方法来做的问答系统。参加for TREC-11的统计问答系统 [113] 。

三个系统:“IBMPQ”;“IBMPQSQA” ;“IBMPQSQACYC”[114]。

(9)DIOGENE

DIOGENE:多语言问答系统。标准的体系结构:问题处理、搜索、答案抽取;MULTIWORDNET是WORDNET的扩展版本[117]。

(10)问答系统QALC：利用Web和WordNet [121]。

(11)QUANTUM

QUANTUM的体系结构:利用模板、浅层解析、正则表达式等来分析问题，并选择最合适的抽取函数。通过Okapi来检索出来段落，并抽取出来候选答案。[126]。

Web问答系统(续)

(12)Tequesta

Tequesta: Amsterdam大学的文本问答系统。这是参加TREC-10 Question Answering track的一个系统，名字Tequesta。就是用了模式匹配的方法。通过一些例如模式匹配、查词典或两者结合的简单方法，来完成识别工作。为了识别问题的目的，把问题分为18个类型，用模式匹配的方法来识别问题类型，有67个模式可供匹配。如果多于1个模式同时匹配上问题，那么，这个问题有多个目标。这些模式被排序，分出先后。并且，答案选择模块遵守这样的原则：对于更加详细的目标，问题被分类来选择答案。如果上述方法不灵，就把问题类型设定为未知[127]。

(13)SHAPAQA

SHAPAQA:World Wide Web上问答系统的浅层解析

我们介绍SHAPAQA，它是对World Wide Web上在线的、开放领域的问答系统进行浅层解析的方法。给定一个基于形式的自然语言(受限预言？)输入的问题，系统利用一个基于记忆(memory-based)的浅层解析器，来分析web页面，其中，这些web页面是利用正常的关键词在搜索引擎上检索出来的。本系统的两个版本在一个200个问题的测试集合上被评测。[134]

Web问答系统(续)

(14)QALC

在这篇论文中描述了QALC系统(the Question-Answering program of the Language and Cognition group at LIMSI-CNRS).这个系统参加了TREC-8 的QA-track.QALC的基本结构包括5个平行的模块。2个模块进行问题处理，三个模块进行文本处理。

(15)**其它:问答系统AskMSR[100]、开放领域问答系统JAVELIN[108]**。参加TREC2002 QA track的系统：DLT。

推理在问答系统中的应用

问答系统缺乏推理能力，推理系统缺乏自然语言理解能力，这是一个老问题了。正是这个问题困扰着大型知识库系统的建设，也使花费巨大的人力物力建立起来的知识库系统难以面向公众开展达到一定质量的知识服务。

Montague认为，自然语言和逻辑人工语言没有实质的区别，自然语言和逻辑人工语言本质上是相通的，作为符号系统，它们都遵循共同的结构规律，这就是所谓通用语法的思想。

把用户问题进行解析，解析成为Prolog能够推理的地步，然后，利用prolog作为工具进行推理，从用户问题中抽取出需要的信息。也就是说，通过谓词逻辑推理，来获取用户的提问意图。这种研究方法的局限性在于：语序变化的多样性；谓词构造的不容易；谓词构造，主要靠关键词，关键词语序是变化的[104]。

推理规则的获取

在基于文本的问答系统中，一个较大的挑战在于:问题和候选答案文本经常用不同的词汇表述。为了解决这个问题，人们考虑引进词汇级别的推理技术，以及结构级别的推理技术，例如，下面就是一个这样的推理规则:"X writes Y" →"X is the author of Y"，在回答用户问题的时候，这样的推理规则是非常有用的，但是这样的推理规则很难大规模的建立、构造[69]。并且，这样的推理规则并非万能的，只在一定情况下应用，请看下述的无效推论:

All socialists are vegetarians.

Nobody is a socialist.

=>Therefor,nobody is a vegetarian.

推理规则的归纳通常是一个人工整理的过程，依靠的是语言学家的语言知识，所以比较适合处理有规律的语言现象，而缺乏对一些特例的处理能力。但这存在以下两个方面的问题：一是从规则的应用范围上看，靠人工的方法只可能产生一些共性规则，不可能产生数量较多的针对个别情况的个性规则，而个性规则尽管应用范围小，但也是保证正确率的重要手段；二是人工方法产生规则的准确率有待验证。因此，在统计方法正确率不易再提高的前提下，能否自动高效地获取规则是实现汉语词性标注中的关键问题[84]。

问答系统中的推理

自然语言语义的形式化问题很困难，原因是目前数学和逻辑学都没能为之提供一个令人信服的工具。首先，自然语言的句法与语义的界定是一件不可能的事：与人工语言不同，自然语言的句法和语义纠缠在一起，几乎在所有的层面上，二者都是不可分割的。其次，为描述自然语言而构造的句法和语义无歧义的形式语言的描述能力值得怀疑[66]。

规则必须人工制定，所以仍然需要大量人力物力。 **所以考虑自动获取。**

Web问答系统工作流程

现有的Web问答系统，处理顺序一般是用户查询处理、从Web上检索相关网页、从得到的Web网页中抽取相关文本并形成答案，最后将答案提交给用户。它所需要的技术，不仅仅包括IR技术、IE技术，而且还包括问题类型分类、命名实体识别、指代解析、查询扩展。时态变换、浅层解析、语法、语义、语用、语义消歧等。例如，以“Who”开头的问题一般需要“PERSON”类型的命名实体作为答案；以“When”开头的问题需要“DATE”类型的命名实体作为答案。以“Where”开头的问题需要“LOCATION”类型的命名实体作为答案。

利用匹配的方法发现可能的答案 [116]。

问答系统的查询转换。查询转换的机器学习方法，可以提高检索答案的准确率。运行期间，问题被转换成一系列的查询[120]。

问答系统中基于Web的模式挖掘和匹配方法。对每种类型的问题，利用TREC QA track的数据作为训练例子，都可以从web上自动学习很多文本模式。这些文本模式的评估借用数据挖掘的方法。给定一个新的未知的问题，这些文本模式可被用来对web上面可能的答案进行抽取和排序[123]。

利用SMART系统进行查询扩展[137]。

问答系统评测

问答系统的语义和复杂性：自然语言工程的Moore定律。一个问答系统任务的复杂性，依赖于有可能抽取的答案的复杂性。并且，这些答案中的每一个，其复杂性依赖于那个答案出现的期望水平[135]。

- Timeliness.要实时。
- Accuracy准确性;因为错误的答案甚至比没有答案更加恶劣。
- Usability有用性;
- Completeness. 完备性
- Relevance.相关性

Qaviar是一个实验性的问答系统自动评测系统。研究的目的，是为了发现一个在问答系统中，和人类评测机制类似的自动计算方法。Qaviar通过人为给出答案中的关键词，来进行判断。如果人给出的答案的关键词的个数超过一个给定的召回率。如果召回率超过一个阈值，就判断答案正确。Qaviar的评测结果，和人工评测结果相同率，达到93%～95%。41个问答系统被Qaviar和人工评测者同时进行评测，同样的数据下，人工评测准确率达到0.956，Qaviar正确率达到0.920。本文中，我们报告了Qaviar，这是一个实验性质的对问答系统应用进行自动评测的系统，我们的研究目标，是发现一种对于问答系统方面，和人类评估标准一致的自动评测方法。Qaviar通过计算系统返回的答案中，人类给出的答案中关键词的召回情况。如果召回的答案中的关键词超过一定数目，就判定这个答案是正确的。Qaviar和人类判断的一致性达到93% 到95%。Qaviar和人工评测者评价了41个问答系统，[99]。

问答系统评测（续）

答案的评测标准，从以下方面入手：

- 答案的相关性；
- 事实、数字和例子的准确性；
- 逻辑的一致性和清晰性；
- 答案和主题的一致性；
- 分析的全面性；
- 语句的通顺性以及是否有感情修辞；

问题不等于查询;答案不等于文本;[125]

问答系统评测（续）

应该说，目前的QA测评标准，还没有成熟的。就连TREC QA Track的评测标准，也有相当的排脑袋的成分在里面。

目前TREC QA Track的评测标准，是这样的：评测委员会人工给出标准答案。对于每个问题，参赛系统给出5个答案（系统运行结果），然后与标准答案进行比较。

如果第一个答案就是正确的，那么系统得1分；

如果第一个答案错误，而第二个答案正确，那么系统得1/2分；

如果前两个答案都是错误的，而第三个答案正确，那么系统得1/3分；

如果前三个答案都是错误的，而第四个答案正确，那么系统得1/4分；

如果前四个答案都是错误的，而第五个答案正确，那么系统得1/5分；

如果所有答案都是错误的，那么系统得0分。

总的来说，这个评测方法是手工完成，而不是一个自动的过程。

问答系统评测（续）

目前TREC Web TRACK检索用的较多的，是查全率（Recall）和查准率（Precision），以及两者的调和平均数F。具体来说，就是：

Recall = 查询出来的准确答案个数 / 数据库中所有的准确答案个数

Precision = 查询出来的准确答案个数 / 查询出来的所有答案个数

F = 2*P*R/(P+R)

这个评测方法目前用的很多，所以我在想，它能否用于问答系统的评测过程？

具体说来，对于用户提出的一个问题，如果系统给出的答案是完全正确的、和标准答案完全一致，那么系统得分是1；如果系统给出的答案是完全错误的、和标准答案根本不沾边，那么系统得分是0；如果系统给出的答案位于完全正确与完全错误之间，那么系统得分就是0和1之间的一个值，或者说，系统给出的答案越是接近标准答案，那么系统得分就越是接近1，否则就越是接近0。这个评测标准，就叫做正确性评测标准吧。

对于用户提出的一个问题，知识库中可能存在多个正确的答案。作为一个问答系统，最理想的状况，就是把所有这些正确的答案全部提交给用户，这种情况下，系统得分是1；最糟糕的状况，就是无法向用户提交任何一个正确答案，这种情况下，系统得分是0；如果系统向用户提交的答案个数，位于这两种极端的情况之间，那么，系统得分就是0和1之间的一个值，或者说，系统给出的正确答案个数越多，那么系统得分就越是接近1，否则就越是接近0。这个评测标准，就叫做完备性评测标准吧。

问答系统评测（续）

根据我的直观感觉，一个优秀的问答系统，应该满足如下标准：

首先，其响应时间必须能够让人接受，响应时间越小越好，最好是实时响应。这是一个可以用时间进行度量的问题，或者说，是可以量化的、是客观的。

其次，系统给出的答案应该是准确的。一个不能够给出准确答案的问答系统，很难说它有什么实际用途。那么，标准答案应该是什么？我个人认为，标准答案应该是人工给出的，至于系统运行结果是否与标准答案相符，也是人工评测的。

再次，系统给出的答案应该是全面的。对于同一个问题，可能有多个答案，或者答案包括多种情况。一个完整的答案，应该能够处理上述情况。如何评测呢？人工评测。

再次，系统给出的答案应该是语句流畅、简短扼要，而非生硬拗口、长篇大论。如何评测呢？人工评测。

最后，还有一个提问问题的难易问题。对于每个问题，不妨标出其难度系数，如果系统对难度系数高的问题给出正确答案，那么对于本题来说，系统的得分就高。

一些相关问题

子目标控制：用深度优先遍历搜索。

（1）所谓正确性，是一个语义上的概念。也就是说，你首先要指定一种给每一个句子赋予真假值的机制。然后，如果你不管选择这种机制下的任何一种赋值方案，只要在这种方案下前提（事实库和规则库）都是真的，那么你的推理机制得出来的结论也一定是真的。

（2）在模式推理中，模式和模式之间的形式上的推论关系，本质上只有两种：变量的实例化关系（模式可推出则模式的实例化也可推出）和推论的传递关系（凡是可推出的前提推出来的东西都可以当作前提）。这两种都是和一阶逻辑的推理规则相符的。因此，只要事实库和规则库的东西都是真的，我们的模式推理机制得出的结论必然是真的。也就是说，模式推理机制具有正确性。当然你要去证明。

（3）即使个别事实和个别规则不是真的或者不知道是不是真的，模式推理也不会比一阶逻辑更加扩大矛盾的传播，因为它只用了一阶逻辑的推理规则和逻辑公理的一个子集。

提　纲

1、论文的总体结构
2、已经完成的部分
3、尚未完成的部分
4、时间规划
5、课程完成情况
6、科研工作完成情况
7、论文发表情况

尚未完成的部分

1、多元多次模式之间的合一算法。
2、模式合一、模式推理在问答系统中的应用。
3、一个问答系统（准备用多引擎的方式实现）。

提　纲

1、论文的总体结构
2、已经完成的部分
3、尚未完成的部分
4、时间规划
5、课程完成情况
6、科研工作完成情况
7、论文发表情况

时间规划

1、~2004.4：完成论文的“引论”、“综述”模块；

2、~2004.7：完成论文的“多元多次模式之间的合一算法”、“模式推理”模块；

3、~2004.10：完成一个问答系统；

4、~2004.12：改进问答系统并部分的解决“指代消解”问题。

5、提交论文，准备答辩。

提　纲

1、论文的总体结构
2、已经完成的部分
3、尚未完成的部分
4、时间规划
5、课程完成情况
6、科研工作完成情况
7、论文发表情况

课程完成情况

1、完成了包括《博士生英语》在内的**19门课程**的学习任务，共获得**61个学分**。

2、博士生英语**84分**，其余课程，**85分以上共有8门**。

提　纲

1、论文的总体结构
2、已经完成的部分
3、尚未完成的部分
4、时间规划
5、课程完成情况
6、科研工作完成情况
7、论文发表情况

科研工作完成情况

1、2000.04~2000.7：参与973项目“天罗搜索引擎”项目，完成了部分模块的代码改进工作（组长余志华）。

2、2000.08~2000.12：参与“北京图书馆数据采集推送系统”项目，完成了“关键词标识”模块（组长余志华）。

3、2001.01~2001.07：参与“中科力腾企业信息平台(EIP)系统”项目，完成了 “Office与Lotus格式转换”模块（组长廖华明） 。

4、2001.08~2001.12：参与“中科院计算所所务信息平台系统”项目，完成了“数据的图形化转换”模块（组长虎嵩林）。

5、2002.1 ~2002.12：参与“新词发现”、“www.nlp.org.cn数据整理”等项目（组长刘群老师）。

6、2003.1~2003.10： 参与“安全中心敏感信息发现（TDT）”项目，参与项目的早期讨论并完成了“采集数据单元”模块（组长骆卫华）。

7、2001.11~ ：在老师指导下，做“浅层结构模式推理”研究项目。

提　纲

1、论文的总体结构
2、已经完成的部分
3、尚未完成的部分
4、时间规划
5、课程完成情况
6、科研工作完成情况
7、论文发表情况

论文发表情况

以第一作者，在各种期刊、会议上，共发表论文7篇。

1 、王树西、白硕、姜吉发等；模式合一的“斩首”算法；中国人工智能学会第10届全国学术年会论文集（上）； P528～P532；北京邮电大学出版社；2003年。

2、王树西 、刘群、白硕等；基于动态知识库的问答系统研究；语言计算与基于内容的文本处理；P587～P592；清华大学出版社；2003年。

3、王树西 、刘群、白硕；A Survey on Question Answering System；Advances in Computation of Oriental Languages 2003； P498～ P506；清华大学出版社；2003年。

论文发表情况(续)

4、王树西、刘群、白硕；自然语言界面的专家系统的研究；计算机工程与应用；第39卷第17期；P35 ～ P37；2003年。

5、王树西、刘群、白硕；一个人物关系问答的专家系统；第七届中国人工智能联合学术会议论文集； P31～P36；广西师范大学出版社；2003。

6、王树西、白硕、姜吉发；模式合一的“斩首”算法及其应用；计算机工程（已经录用）。

7、王树西、刘群、白硕；红楼梦人物关系问答系统；第一届学生计算语言学研讨会（SWCL2002）论文集；P168 ～ P174；2002年。

参考资料

1、A M.Turing. Computing Machinery and Intelligence. MIND,1950,59(236):433—460.

2、白硕. 计算语言学讲义（电子版）. 2001年.

3、陆汝钤. 世纪之交的知识工程与知识科学. 清华大学出版社. 2001年.

4、陆汝钤. 人工智能. 科学出版社，2000年.

5、Rohini Srihari and Wei Li. Information Extraction Supported Question Answering. http://trec.nist.gov/pubs/trec8/papers/cymfony.pdf. 1999 .

6、Jinxi Xu, Ana Licuanan, Jonathan May, Scott Miller and Ralph Weischedel. TREC2002 QA at BBN:Answer Selection and Confidence Estimation.
http://trec.nist.gov/pubs/trec11/papers/bbn.xu.qa.pdf.

7、Ulf Hermjakob, Abdessamad Echihabi, Daniel Marcu. Natural Language Based Reformulation Resource and Web Exploitation for Question Answering.
http://www.isi.edu/~ulf/papers/trec2002.pdf.

8、Jun-Tae Kim,Dan I. Moldovan.Acquisition of Linguistic Patterns for Knowledge-Based Information Extraction. http://www.computer.org/tkde/tk1995/k0713abs.htm

9、Patrick Pantel and Dekang Lin.A Statistical Corpus-Based Term Extractor.
http://www.cs.ualberta.ca/~lindek/papers/ai01.pdf.

10、白硕 程学旗 郭 莉. 大规模内容计算. 语言计算与基于内容的文本处理. 清华大学出版社. 2003年.

11、许洪波.大规模信息过滤技术研究及其在Web 问答系统中的应用（博士论文）. 2003年.

… …

谢谢