Artificial Intelligence and Robotics Research 人工智能与机器人研究, 2021, 10(1), 9-28
Published Online February 2021 in Hans. http://www.hanspub.org/journal/airr
https://doi.org/10.12677/airr.2021.101002

Hans汉斯

人工智能时代知识工程的初步探索

杨福义[1]，叶其松[2]

[1]鞍山师范学院，辽宁 鞍山
[2]黑龙江大学俄罗斯语言文学与文化研究中心，黑龙江 哈尔滨
Email: yangfuyi@sina.com

收稿日期：2021年1月7日；录用日期：2021年1月21日；发布日期：2021年2月7日

摘　要

本文分析了人工智能时代知识工程的理论与实践的关键问题。进行了系统理论、工程技术与应用实践结合问题的一系列探索，分析出大数据人工智能时代知识工程的核心问题，研讨了知识工程的基本单元与复杂网络技术在知识工程中的应用思路。最后给出国家实施人工智能时代知识工程的一些讨论与建议。

关键词

人工智能，知识工程，粒计算，复杂网络，构词矩阵

Knowledge Engineering Exploration in the Era of Artificial Intelligence

Fuyi Yang[1], Qisong Ye[2]

[1]Anshan Normal University, Anshan Liaoning
[2]Center for Russian Language Literature and Culture Studies of Heilongjiang University, Harbin Heilongjiang
Email: yangfuyi@sina.com

Received: Jan. 7th, 2021; accepted: Jan. 21st, 2021; published: Feb. 7th, 2021

Abstract

This paper analyzes the key issues of the theory and practice of knowledge engineering in the era of artificial intelligence. A series of explorations on the combination of system theory, engineering technology and application practice are carried out, the core issues of knowledge engineering in the era of big data artificial intelligence are analyzed, and the basic units of knowledge engineering and the application ideas of complex network technology in knowledge engineering are discussed. Finally, some discussions and suggestions on the implementation of knowledge engineering in the era of artificial intelligence are given.

文章引用: 杨福义, 叶其松. 人工智能时代知识工程的初步探索[J]. 人工智能与机器人研究, 2021, 10(1): 9-28.
DOI: 10.12677/airr.2021.101002

Keywords

Artificial Intelligence, Knowledge Engineering, Granular Computing, Complex Network, Chinese Character Word Formation Matrix

1. 引言

随着科学技术和网络应用的发展，以知识为基础的经济活动正在全世界范围内兴起。知识组织、知识管理和知识工程迅速成为实业界和学术界关注的焦点。大数据技术与人工智能(artificial intelligence，简称为 AI)以及知识工程的应用迫切需要理论与工程实践的密切结合，而当前人工智能领域面临着一系列的困惑。

徐匡迪院士之问揭开当下中国人工智能虚伪的面纱。2019 年在上海召开的院士沙龙活动中“徐匡迪之问”引发共鸣：“中国有多少数学家投入到人工智能的基础算法研究中？”核心算法缺位，人工智能发展面临“卡脖子”窘境。中国制造正从“硬件组装厂”向“软件组装厂”蔓延，政产学研浮躁如故、积习难改[1]。”

张钹院士在《后深度学习时代的人工智能》中指出：“现代的电子计算机还需要在很长一段时间里依赖冯·诺依曼结构。”

“当前，连大脑的运行机制都没有研究清楚，怎么可能开展完全的类脑计算呢？类脑计算研究的开展，需要学科的交叉，我特别推荐数学、认知科学、心理学、神经科学和语言学等领域的学者积极开展交叉学科研究，从而推动人工智能理论的进一步发展和创新。”

谭铁牛院士特别提醒说，人工智能领域的误解和炒作普遍存在，包括人工智能就是机器学习(深度学习)、人工智能与人类智能是零和博弈、人工智能已经达到 5 岁小孩的水平、人工智能系统的智能水平即将超越人类水平、30 年内机器人将统治世界、人类将成为人工智能的奴隶等等，这些错误认识会给人工智能发展带来不利影响[2]。

机器人所显示出的人工智能，归根到底，是人的智能的表现形式，没有人的智慧和能力，哪里会有什么人工智能。我们现在对这一点是很清楚的：机器人的那些高超技能，是由人对其所设计并输入的程序而显示出来的，“人工智能”这个提法已经很清楚了；机器人的“智能”归根到底是人的智能，而不是什么机器的潜在智能。

人类历史上的任何重大科技发明和发现。都为人类社会的生产带来巨大的改变。院士之问与论述，需要引起我们对人工智能的深入哲学思考，以至于清醒地确定研究方向，解决人工智能的实际应用问题。

专家系统是人工智能的重要应用。“对于人工智能的未来发展，要不忘初心、继续探索，回归人工智能的研究本源；特别是对于人脑智能机理的挖掘，孕育着信息科技的重大变革。[2]”本文从各学科交叉研究和近年专家系统开发应用的实践，探讨知识理论技术的工程应用与知识工程中最核心的两项技术：认知单元的确定与处理和网络体系结构的设计问题，进行了初步的探索，现汇报给同仁前贤，予以思考、评议和讨论。

2. 人工智能与知识工程

2.1. 重要术语的定义

科学技术语言中最重要的是基本术语的定义。没有准确的术语定义，定理、推理和结论一系列的证明无法展开。《证明与反驳》这本书译者的话指出([3]，序 2)：“冲突可以促进数学生长，总有点费解。作者给你细讲了几条线索中的一条：证明与反驳互为触发剂，协同作用于数学知识的革新。”“新概念(新方法)的形成是数学史上的里程碑。”。

在定义释义中出现的术语依然需要定义，这是无穷的递归，以至于某个环节出现循环定义而成为两个定义的循环映像。为了解决术语概念的定义递归的终结，先哲和前贤作了以下论述：

亚里士多德是世界古代史上伟大的哲学家、科学家和教育家。

亚里士多德指出：“定义是一个有关本质不可证明的陈述。([3], p. 129)”。

帕斯卡(Pacal, B.)是伟大的哲学家、数学家、物理学家。为了纪念它对人类的伟大贡献，物理学以他的名字命名了国际单位制中表示压强的基本单位帕斯卡，简称帕，符号 Pa。计算机科学命名了语言语法严谨，层次分明，程序易写，可读性强，第一个结构化编程语言为 Pascal 语言。

帕斯卡的定义规则([1659], 第 596~597 页])：“凡是一目了然的现成术语。不要下定义。凡是有一丝一毫模糊性或多义性的术语，不准不下定义。在术语的定义里。只使用一目了然或做过解释的词。([3], p. 128)”。

ISO 标准是指由国际标准化组织(International Standard Organization, ISO)制定的标准。关于人工智能与专家系统，中国国家标准予以等价采用。主要术语定义如下[4]。

人工智能[artificial intelligence]一门交叉学科，通常视为计算机科学的分支，研究表现出与人类智能(如推理和学习)相关的各种功能的模型和系统。表现出与人类智能(如推理和学习)相关的各种功能的功能单元的能力。

知识工程[knowledge engineering]一门学科，研究从领域专家及其他知识源获取知识，并将这些知识组织成知识库。

知识(用于人工智能) [knowledge(in artificial intelligence)]事实、事件、信念以及规则的汇集，以便有系统地使用。

对象(用于人工智能) [object(in artificial intelligence)]具有一种或多种属性的物理或概念实体。

以下概念定义由国家标准和国际标准规定。

概念[concept]通过对特征的独特组合而形成的知识单元。(GB/T 15237.1-2000 术语工作词汇 eqv ISO 1087-1:2000)。

符号[symbol]：用字母、数字、图符以及它们的任意组合来指称概念的形式(GB/T 16786-2007)。

客体[object(entity，item)]可感知或可想象到的任何事物(GB/T 15237.1-2000 术语工作词汇 eqv ISO 1087-1:2000)。

奥地利著名术语学家赫尔穆特·费尔伯在《术语学、知识论和知识技术》中论述了知识理论、术语与知识工程。其重要术语定义如下[5]。

符号世界：符号构成物的总体。

符号句子：符号复合体，它对应于一个逻辑句子。

符号句子链：符号句子复合体，这些符号句子对应于一个由逻辑句子形成的链条。

本文在以上定义与知识理论指导下探索研究人工智能与知识工程问题。

2.2. 第三代人工智能的特点

第一代人工智能是依靠符号，又称为符号 AI。符号 AI 的基本思路：「人类思维的很大一部分是按照推理和猜想规则对‘词’(words)进行操作所组成的」。根据这一思路，他们提出了基于知识与经验的推理模型，因此又把符号 AI 称为知识驱动方法。符号 AI 的开创者最初把注意力放在研究推理(搜索)的通用方法上，与解决复杂现实问题相差很远。寻求通用 AI 的努力遭到了失败，符号 AI 于 20 世纪 70 年代初跌入低谷。以自然语言形式表示(离散符号)的人类知识，计算机难以处理，必须寻找计算机易于处理的表示形式，这就是知识表示问题。

第二代人工智能依靠连接，又称为连接 AI，第二代 AI 的学习理论有坚实的数学基础，引用了概率假设，即大容量假设：样本(Xi，Yi)数量巨大(n→∞)。取得了巨大的成功。深度学习的成功来自于以下 3 个要素：一是数据，二是算法，三是算力，运行 AlphaGo 的机器是由 1920 个 CPU 和 280 个 GPU 组成的分布系统。因此第二代 AI 又称数据驱动方法。

2014 年，深度学习的诸多缺陷不断地被发现，预示着这条道路遇到了瓶颈。深度学习的「黑箱」性质是造成深度学习推广能力差的另一个原因。

第一代知识驱动的 AI，利用知识、算法和算力 3 个要素构造 AI，第二代数据驱动的 AI，利用数据、算法与算力 3 个要素构造 AI。由于第一、二代 AI 只是从一个侧面模拟人类的智能行为，因此存在各自的局限性。为了建立一个全面反映人类智能的 AI，需要建立鲁棒性好与可解释的 AI 理论与方法，发展安全、可信、可靠与可扩展的 AI 技术，即第三代 AI。其发展的思路是，把第一代的知识驱动和第二代的数据驱动结合起来，通过同时利用知识、数据、算法和算力等 4 个要素，构造更强大的 AI。

第三代人工智能是把知识放在第一位，数据放在第二，算法放在第三位，算力放在最后，这个排序是经过仔细琢磨，不是随便排的。主要是强调“知识”在发展人工智能中的重要性[6]。

综上所述，可以分析出第三代人工智能要解决的关键问题是，1) 认知计算的基本粒子问题。2) 基本粒子所构成的立体复杂网络结构体系的研究与分析问题。即一栋建筑大楼的基本材料(元素)与整个大楼的关系。也相似与一颗人造卫星(工程最终产品)与各层次基本元素构成的材料、零件、部件、组件、设备、子系统直至整个装置的多层次关系。

2.3. 知识组织系统

美国未来学家 J • 奈斯比特在《大趋势》一书中指出：“我们淹没在信息中，但是却渴求知识”[7]，就是因为“失去控制和无组织的信息在信息社会里并不构成资源，相反，它成为信息工作者的敌人”。因此，有效组织信息，提供全面、快速、准确的信息检索服务已经成为网络时代亟待解决的问题。英国著名分类法专家布利斯 1929 年提出：“知识组织是图书馆学情报学的分类法和叙词表的基础上发展起来的，其核心就是知识序列化，在这一过程中，一方面传统的情报检索语言继续发展，另一方面，不断有新的知识工具不断涌现。”“新的术语‘知识组织系统’应运而生([8], p. 241)。”

关于信息这一术语。诺依曼是这样叙述的：“一定的周期性或近似周期性的脉冲序列,传送着消息,亦即信息[9]。

诺依曼提出人脑的语言不是数学的语言，并提出第一语言和第二语言的观点。诺依曼预见的人类大脑记忆元件和人脑信息处理系统本质的论述，到目前依然是科学研究的尖端课题，值得我们在科学研究中参考。

图像、声音、电影、动画、文献、知识都是一种符号语言，作为信息是在网络上传送的二进制数字信号序列。

现在不仅仅是知识组织的问题。而是大型海量知识的加工整理存储传播的过程-即人工智能时代的知识工程问题。

2.4. 知识工程

知识工程也是对知识进行组织、存储、加工，分类后，按用户需要输出产品(知识库文化产品与机器人设备)的工业生产过程。

知识挖掘是更高阶段的知识工程。我国著名知识工程专家杨炳儒在知识发现(数据挖掘)方面构建了多层次、递阶的基于内在认知机理的知识发现理论体系并通过成功的应用，验证其科学性与有效性，体现了其科学创新价值与应用创新价值。

当前时代的特点是世界已经变成地球村。世界各国以从来未有的速度以互联网进行知识交流。而知识交流的核心，是统一全世界的知识交流语言。知识交流的语言通过会议论坛和各种文本著作图书在全世界流通。所有知识的表述，离不开语言文本的传播，离不开信息技术。语言工程是人工智能时代的重要基础。世界各种语言文本的描述，都是采用符号。具有人工智能的计算机系统，对应的就是符号的输入输出。而星际通信的基础也只能是光信息所表示的符号二进制信息代码。在以云计算为基础的互联网所有的文本、声音和图像信息，都是转变成为最基础的二进制代码而传输的。因此，信息技术是涉及到人类科学技术的一个基础。与各学科有密不可分的联系。

2020 年 10 月国家自然科学基金委员会网站更新了机构设置信息，在 8 大科学部(数学物理科学部、化学科学部、生命科学部、地球科学部、工程与材料科学部、信息科学部、管理科学部和医学科学部)的基础上增加了“交叉科学部”，意味着“交叉科学部”成为了基金委第 9 个科学部。

10 月 21 日，科技部部长王志刚表示，下一步，将加大对冷门学科、基础学科和交叉学科的长期稳定支持。交叉学科是指不同学科之间相互交叉、融合、渗透而出现的新兴学科。近代科学发展特别是科学上的重大发现，国计民生中的重大社会问题的解决等，常常涉及到不同学科之间的相互交叉和相互渗透[10]。

交叉科学其实质是科学学的研究。涉及各门类学科技术的协调和统一。而科学技术的交流随着经济发展交流的需要，科学技术文献各类标准规范，各种经济活动，都需要技术语言的支持。其中最重要的一条就是各行各业的科技术语。科技术语是知识的结晶，是人类活动知识信息交流的基本单元。知识组织系统的互操作。离不开全世界术语的交流和统一。因此现代语言学研究中的一个重点方向，就是科学技术语言的研究。需要有全世界统一的科学技术术语。

2.5. 知识与知识工程的分类

2.5.1. 知识的分类

分类是以事物的本质属性为根据，把一个属概念划分为若干个种概念的过程。知识分类，是一项极其复杂的科学认识活动。不同的知识论者各有自己的分类理论与分类方法，因此在知识分类史上，就出现了形形色色的知识分类方式。最具有代表性的有 10 种：1) 按照知识的效用分类。2) 按照研究对象分类。3) 按照知识属性分类。4) 按照知识形态分类。5) 按事物的运动形式分类。6) 按照思维特征分类。7) 按照自然现象和社会现象分类。8) 按照知识研究方法分类。9) 按照知识的内在联系分类。10) 按照学科发展趋势分类[11]。

知识的集中存储就是世界各地图书馆存储的大量图书与科技文献，而图书文献的分类实质也是知识的分类。

类是指一组具有某一共同属性的事物对象集合。类又称为**类目**([8], p. 33)。目前所用的文献分类，都是首先把知识领域划分为学科，然后再进一步依靠其他特性进行区分和组织的。所谓学科，是关于客观

世界中特定事物对象的本质特征和规律的知识体系。**学科分类体系**是**知识分类体系**的一种类型，它与科学分类是密切相关的([8], p. 39)。科学分类依据特定的原则，确定知识门类区分和组织的总体框架。另一种检索方法是**主题法**。所谓**主题词**是指论述对象，包括事物、问题、现象、规律等的语词。主题法具有直接、规范、组配、相关、通用和动态的特性([8], pp. 102-104)。在互联网的大量信息中，可以由用户自己进行标注。方便了海量信息的知识抽取与加工。

知识的分类方法有多种，海量文本的自动分类也是人工智能程序下的知识工程。图书文献记载了浩瀚的人类认知世界的知识。文献分类，图书分类，产品标准分类都是可参考的方法。知识分类最直接的方法是参考图书文献的分类方法

目前我国使用的中国图书馆图书分类法(简称**中图法**)是几十年经过全国科技工作者与情报信息工作者实践过的有效方法，目前使用的是第五版。

另一种方法是 **UDC 分类法**(Universal Decimal Classification)，又称为通用十进制分类法。是世界上规模最大、用户最多、影响最广泛的一部文献资料分类法。近百年来，UDC 已被世界上几十个国家的 10 多万个图书馆和情报机构采用。UDC 目前已成为名符其实的国际通用文献分类法。每一项国家标准的封面都有 UDC 分类标识号。

ICS 是由国际标准化组织编制的**标准文献分类法**。它主要用于国际标准、区域标准和国家标准以及相关标准化文献的分类、编目、订购与建库，从而促进国际标准、区域标准、国家标准以及其他标准化文献在世界范围的传播。ICS 是一个等级分类法，包含三个级别。

《说文解字》是首部按部首编排的汉语字典。原书作于汉和帝永元十二年(公元 100 年)到安帝建光元年(公元 121 年)在这本字典中，对篆书进行了以部首为语义分类的标注。

许慎指出：方以类聚，物以群分。同条牵属，共理相贯。杂而不越，据形系联。引而伸之，以究万源[12]。

中国的清朝雍正皇帝在另一本编写的大型辞书《骈字类编》的序言[13]中指出：《易》曰：‘方以类聚’，则类之始也。今字则从骈。义虽不同而不妨于并列，编则从类，事虽互见而不至于混淆，于是乎乾坤之蕴，民物之繁，大备于一书之中。而蔚然萃群书之秀矣。形而上谓之道，形而下谓之器，引而伸之，触类而长之，无器非道也。

辞书、字典和词典都是知识工程的核心资源,是知识工程的根基。

2.5.2. 知识工程的分类

1) 按知识原料的历史阶段分类

按历史阶段可以分为古代、近代和现代知识工程。

知识工程是对人类已有知识。加工存储整理和转换。早在计算机诞生以前。在中华民族汉字 5000 年发展的历史上，很早就有了甲骨文。甲骨文是汉字的祖先，是图画文字的高度抽象而形成的文字符号。为中华民族记载了大量历史事件。记载了中华民族生产生活各个方面的知识。这就是中华民族传统的古籍。对传统古籍的整理、出版与翻译对外交流都是对知识的加工。构成了中华民族传统的以人力实现的知识工程。

现代知识工程诞生于 1977 年，是在电子计算机与人工智能发展下对信息，数据进行处理产生知识产品的工程。不仅在语言工程、自然语言处理中，而且也在现代自动控制工程中发挥重要作用。

2) 按使用的方式分类

对知识工程产品的生产方式有人工、半自动化与自动化方式，在大数据时代，从海量信息中抽取知识，更多的是依靠自动化方式以减轻繁杂的人工劳动。

3) 按知识工程的产品分类

知识工程的产品。对国家具有重要的战略意义。各类标准和规范都是知识工程的产物。都是各行各业的专家对历史知识和最新知识进行探索研究整理而得出的。这些都需要有科技理论与技术基础的支撑，需要工程实践。

知识产品可以分为丛书、类书等，也有标准规范类、词典辞书类、技术文件类和科普教材类等。

4) 按产品应用的领域分类

可以分为科研产品、教育产品和文化产品(如科普读物)等。

5) 其他分类方法。

3. 知识发现与知识的哲学模型

3.1. 人类智慧与知识发现

深入探讨智慧(灵感)与现代科学发展逻辑的内在联系，应该成为创造学、科学学、心理学、思维学又一个综合研究的课题。([14], p. 216)。

爱因斯坦在著作《开普勒》中说：“知识不能但从经验中，而只能从理智的发明同观察到的事实两者的比较中得出。([14], p. 227)。”

智慧只有在和智慧的碰撞中才会发出动人的火花，创造者与创造者之间的切磋、探讨和争辩，是激扬创造智能，砸断常规思维程序的利器([14], p. 272)。

在《术语学、知识论和知识技术》一书中，介绍了中国道教的老子对自然的思考和智慧([5], p. 246)。

老子曰：“道生一，一生二，二生三，三生万物”(老子《道德经》第四十二章)，是老子的宇宙生成论。这里老子说到“一”、“二”、“三”，乃是指“道”创生万物的过程。主要讲述了一、二、三这几个数字，并不把一、二、三看作具体的事物和具体数量。它们只是表示“道”生万物从少到多，从简单到复杂的一个过程。

知识是人类意识的一个特性，人类在知识积累的基础上，通过智者头脑的智慧思考与科学研究，把感性知识上升为理性知识。发现事物之间的规律，并通过各种模型和数学公式、物理定律和化学反应方程式等各种逻辑严密的语言形式构成认识世界的科学积累，极大地丰富了人类知识库。

智慧是人类大脑最神奇的功能。是人类社会发展的动力。

3.2. 知识世界的哲学模型

在《术语学、知识技术》一书中，费尔伯提出了术语学和知识理论的哲学模型。四个世界与四个构成物的分层关系([5], p. 287)。

1) 现实世界：大自然对象客体构成物。实例：事(物的运动)；物(宇宙的一切)。

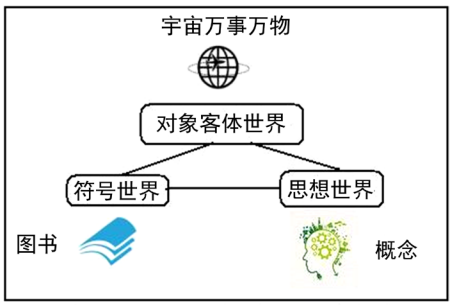

Figure 1. Structure diagram of the knowledge world
图 1. 知识世界结构图

2) 概念世界：人脑思维构成物。实例：概念(字、词与术语)和想象等。

3) 符号世界：符号(含图形符号)构成物。形式有：书籍与文献的句子、段落、篇章等。

4) 知识世界：知识构成物。实例有：书库、图库、实物展览库和影视库(包含所有构成物之间的关系、关联、分类、聚类和解析等的视觉信息)。

图 1 所示是知识世界结构图。

4. 知识工程的核心－认知计算的基本单元

4.1. 关于知识单元的论述

认知计算是认知科学的核心技术子领域之一，是人工智能的重要组成部分，是模拟人脑认知过程的计算机系统。近些年。知识理论中关于知识单元有许多论述：

中国科学院陈琳院士指出：发展新一代人工智能的核心基础科学问题是：“认知和计算的关系”。“研究任何一种过程，建立任何一种过程的科学理论，必须首先回答一个基本问题：这种过程操作的基本单元是什么？每门成熟的基础科学都有其特定的基本单元，如高能物理学的基本粒子、遗传学的基因、计算理论的符号以及信息论的比特等。对于认知科学而言，必须回答：什么是认知过程操作的基本单元？[15]”

中国科学技术信息研究所宋培彦(2014)等人研究了从术语定义句自动抽取知识单元的方法。并进行了术语知识单元的抽取实验。采用了哈尔滨工业大学的依存句法分析器作为工具，抽取 100 条术语定义的实例。指出：“依存树的叶节点与父节点构成了修饰关系；粗颗粒的知识单元表现为父节点，而细颗粒的知识单元则要深入到叶子节点[16]。

北京中医药大学齐熠、都立澜、李晓丽等人讨论了中医术语翻译中的翻译单位问题。指出：“翻译单位就是在译入术语和译出语中相互匹配的意义单位。是个动态的单位”“中医术语翻译的基本单位应是基于语义的最小词语搭配结构。最重要的是，翻译单位的判断不应该是基于经验的推断，而应该是基于实证的总结[17]”

张钹指出。人工智能。和认知科学研究者观察到人类智能的一个公认特点。那就是人们能够从极不相同的粒度上。观察和分析同一个问题。人们不仅能在不同的粒度的世界上。进行问题求解。而且。能够很快的从一个粒度世界跳到另一个粒度世界。往返自如，毫无困难。这种处理不同粒度世界的能力。正是人类问题求解能力的强有力的表现([18], p. 42)。

中国科学院计算所的研究专家史忠植指出。“人类不仅从不相同的粒度上观察和分析同一个问题的能力。更具有根据特定的背景。把相应的数据、知识和资源。建立成和问题求解的粒度空间的能力。也就是说，根据具体具体的任务背景。把论域中的对象抽象为不相同的粒度和构建出粒度之间的关系。进一步生成适合于该任务的粒度空间。通过在该粒度空间上的粒之间，粒集之间及粒度层次之间的往返跳跃，得到问题的解，这是人类智能的一个重要特点，也是模型的哲学基础([18], p. 42)。”

综上所述，可以看到，对象实体研究的基本单位是结点，而对象结点之间的关系则构成世界各种万万千千复杂的知识。

笔者认为：认知计算的基本单元是对象客体在人脑映像形成的概念。而这些概念对应的基本符号就是意音相关的符号－汉字。对概念基本符号汉字及其关系的各层解析，就构成结构分明，可以进行认知计算的知识组织系统完整的全貌。

4.2. 汉字是认知计算的基本单元

4.2.1. 人工智能与语言学的研究现状

2020 年 10 月 31 日，中国语言学会副会长北京大学教授陆剑明在第十九届中国语言学大会的报告中

指出[19]：

然而当今基于人工神经网络的“机器深度学习”所获得的人工智能，与语言学本体研究，不是结合得更加紧密，而是越来越疏远了。在人工智能发展中，语言学科逐渐被边缘化，汉语语言学人面对人工智能普遍感到茫然。如今 NLP 研究如火如荼，然而汉语本体的研究成果却没能在当下的人工智能研究中派上用场，其根本原因在于没有解决好中文信息处理与汉语本体研究的接口问题。这一方面导致汉语言学的“掉队”，另一方面也使得 NLP 中缺乏“语言知识”。因此他建议，汉语本体研究应当更多地关注中文信息处理的需求，从当下的注重“理论”思辨，转向深入句法语义等的研究。更要从中文信息处理的需求出发加强对语言事实的研究。

今后汉语本体研究似需加强以下几方面的研究：第一，树立“关联”意识，运用关联理论(Relevance Theory)，进一步深入句法语义研究。第二，加强词语和句法格式的句法、语义的特征研究。第三，进行汉语语言信息结构的探究。第四，加强现代汉语“边缘(periphery)结构”的研究。

而汉语的本体研究与中文信息信息处理的核心是和汉字的形音义。

4.2.2. 世界语言史·埃及语·英语

世界语言史就是语言帝国的侵略扩张史，多少语言在历史中湮灭了，多少语言顽强地维护自己的新生并承受外来语的侵袭。在《语言帝国：世界语言史》这本著名的的书中都可以找到答案。英国语言学家尼古拉斯·奥斯特勒从全世界数千年的语言动力学的宏观角度分析了语言的扩展与萎缩消亡的历史[20]。

索绪尔在普通语言学教程中指出[21]：

世界上只有两种文字体系。

一、表意体系。一个词只用一个符号表示。而这个符号不取决于词赖以构成的声音。这个符号和整个词发生关系。因此，也就间接地和它所表达的观念发生关系。这种体系的简单例子就是汉字。

二、表音体系。它的目的是要把词中一连串连续的声音模写出来。表音文字有时是音节的。有时是字母的。记忆言语中不能再缩减的要素为基础。

关于埃及语，这一具有古老文明特色的语言，在被侵略中衰落消失。

“侵略最终还是断送了埃及语在故土上的生命。如今的埃及毕竟是个穆斯林国家。基督徒占极少数，人人都说阿拉伯语。”“最终，埃及语不再是埃及的主要语言。因而它也无法继续维系自己的生命。”“3 个世纪之后，埃及受到了一次更大的震动，他被一个截然不同的外来帝国吞并了。这个帝国的中心位于埃及东边。这一打击对埃及而言大大超出了其承受能力。这是埃及第一次也是最后一次走向了衰落” ([20], pp. 150-153)。

关于英语，中国知名英语教育专家，雅思培训界大师，悉尼大学双硕士刘洪波指出[22]：“英语中80%是外来词汇。是一种典型的大杂烩语言。其中古希腊语和拉丁文。占了很大的比例。所以古希腊和古罗马人的言行风俗对英语词汇的影响相当深远。”

但由于大英帝国的征服世界，使世界四分之一版图纳入它的怀中，英语很快就灭亡了世界大量的语言。目前，世界语言消亡了二分之一，只余下近 6000 种。

“如果英语经济体变得僵硬而缺乏变化。那么它在语言上的领先优势。也很有可能将随之消失殆尽。作为一种通用语。英语依然面临各类困难。以往各种语言的命运。都可以为其佐证([20], p. 493)”。

4.2.3. 汉字、汉语与符号学

汉语由汉字记载着历史与文化科技。汉字所承载的是概念集合。是中文信息处理的核心。汉字是术语的最基本元素，是各种语句的基础，也是目前世界各国语言中唯一即具有视觉信息又具有听觉信息反映概念实质的基本描述单位。在认知计算与数据挖掘知识工程获得广泛的应用。

汉字是古老的自然文字中的表意文字。而且是 5000 多年发展。至今没有改变性质的表意文字。所以。它的发生和发展都植根于中华民族的生存环境[23]。

“汉字的最大优势就在于，它用一种巧妙的方式代表了不同方言所共有的、最为普遍的形式和意义。所有的现代方言，也包括文言，从本质上说。就是一系列表示意义的音节。这些音节也许在不同的方言里有着不同的发音和先后次序，但是他们一旦进入到图画文字，就能够一一辨认出来。通常汉语的一个音节就是一个汉字。所以不论说的是哪种方言，只要识字。就一定能看懂一篇汉语文稿的意思。([20], p. 143)”“此壮举无异于一个精妙、含混至极的奇迹，所以说传统的汉字还是幸存下来([20], p. 144)”

图纸，是工程界的语言。包含着大量抽象的符号表示各类对象客体。图形符号在地理学中的地图，电工电子学的原理图、布线图，交通工程的路线图与施工图都获得广泛的应用。图形符号承载着经济建设中的各类先进技术，是信息传递的有效工具，而汉字由图像符号变换而来，具有比类取像的意义。

例如：汉字甲骨文“立”的取像图形如图 2，前四画是站立的人形，点是人的头部，横是人的两臂，三四画是人的双腿，最后的横是脚下的大地。

Figure 2. The evolution of the font source of the Chinese character “立”
图 2. 汉字“立”的字形字源演变

汉字的一切由图形而诞生，表意是汉字的根本性质。汉字的科学性直接通过图形的高度抽象而产生，通过视觉形象直接在人脑形成了最基本的块单元。

4.2.4. 汉语与汉字的科学性受到了历史冲击与倒退

在元朝和清朝，统治者强行推广外族语言和文字，并定为官方必用之政令，但遭遇了失败。“元朝统治者忽必烈汗。甚至还命人为自己帝国内的所有认知语言。诸如蒙古语汉语突厥语。还有波斯语。设计一种字母式的书写系统。专门用于正式场合儿。人们称它为八思巴文。”“仅仅过了一个世纪就随着元朝一起消失了。[20]”

鲁迅错把中国的落后归结于汉字的难学，他临逝世(1936 年 10 月)“答救亡情报访员”时，更坚决地说：“汉字不灭，中国必亡。”理由是：“因为汉字的艰深，使全中国大多数的人民，永远和先进的文化隔离，中国的人民，决不会聪明起来，理解自身所遭受的压榨，理解整个民族的危机。[24]”

以合并汉字，篡改部首作为革新，汉语汉字的文化倒退冲击着中华民族的优秀传统。“我们在汉语的近代史中也发现了种种类似于导致埃及语消亡的文化倒退现象，世界上有 1/5 的人使用汉语，但它如今也出现了一些不祥的征兆([20], p. 157)”

北方民族大学在科研基金项目“汉字简化与汉字记号化研究”[25]中。指出。从《说文解字》到现代汉字，部首字的数量由 540 部减少到 201 部。发现在汉字简化的过程中。由于部首合并、混形、变异以及从属字消失等原因，导致大量的部首字消失，从而导致汉字进一步记号化。

记号是帮助记忆和识别而做的标志。单纯由记号。直接组成的字。叫做记号字。对《三千高频度字表》中的前 500 个汉字做了统计分析，发现记号字、半记号字。占到 73%。这种变化我们称之为汉字记号化。

汉字记号化，违背了汉字向科学化规范化发展的方向，是一种严重的的文化倒退。受到了新时代掌握科学的人民大众的顽强抵制。互联网流行的“亲不见，爱无心，产不生，厂空空。”这是对汉字不科学不规范的所谓“改革”的绝妙讽刺。当前由于部首改革乱象的破坏，使得汉字产生多开门现象，一个字归于多个部首。编辑随意处置。造成人力物力的浪费，也出现大量的辞书编辑的混乱现象。万能的记号“又”，出现了“邓对戏观艰难凤”字混同一起，杂乱无章的情况。

由于把鸡(鷄)归为“又”部，造成鸡鸭分家(指鸡进入又部，没设甲部，鸭仍归鸟部)。原归于鸟部的鹏字，进入了月部。不仅造成简体字和繁体字的分裂和分类归部不一。鹏程万里的鹏字。归为月部。加剧了月部字的混乱。“发生”这个词，到底是“髮生”还是“發生”词典只好加注。多个汉字合并为一字，冲击了汉字科学体系，为中文信息处理构成难以逾越的难关，无法按字对文本进行分类。所以割裂汉字形音义之间的关系。所谓的据形定部法，没有任何科学道理，是一种文化倒退现象。所谓“排检法”的部首，侵占破坏文字学意义部首，这种乱象应该尽快予以纠正。一个国家一个民族不能没有灵魂，应该回归“正本清源、守正创新”的正确道路上来。

《汉字部首解说([26], p8)》中指出：GB13000.1 字符集汉字部首归部规范确定了 20,902 字的部首。对于“归部规范”这一文件，学术界、教育界颇有争议。“一些形声字，按新的方法归部，违背造字原则，不便于识字教学。

目前。现代汉语还正在发展中。汉字汉语需要走向科学化规范化的发展道路。中华民族的共同语。应该称为华语。目前所出现的国际汉语。世界汉语。造成了中华民族共同语的分支乱象。应该引起国家语言文字管理部门的重视。

汉字的科学性与现代汉语的结构性决定它在世界语言交流中的重要地位，也决定它的未来。这是语言知识工程与信息处理中建立世界语言互译中应该处于中心地位而值得高度重视的工作。

4.3. 粒计算的基本单元与结构层次

概念：是人脑对客观事物本质特征的认识,是高级认知活动的基本单元。利用粒度空间结构在解决问题时，根据具体的情况，在不同的粒度层次之间进行往返跳跃。从而提供一种知识发现的描述的新方法。该模型可以在数据挖掘和机器学习等领域中得到广泛的应用([18], p. 43)。

粒计算。是人工智能领域新兴起的一个研究方向。是信息处理的一个新的概念与计算范式。主要用于处理不确定的、模糊的、不精确的、部分真的和海量的信息。其基本思想是利用不同粒度上的信息进行问题求解。目前，该方向已成为人工智能领域的研究热点之一。而且有望成为复杂系统中智能信息处理的一种有效的理论框架。粒计算的主要模型有粗糙集模型。模糊集模型。商空间模型。三者有一个共同之处，就是都考虑到了人类智能中有从不同粒度度思考解决问题的特点。他们思考问题的出发点与求解问题的任务却不尽相同，各有特色和优势。([18], p. 142)

粒计算把上面所叙述的各类单元都抽象化为知识粒，从而用粒计算与复杂网络理论解决知识工程的实际问题。

4.4. 粒计算的概念格理论与汉字

概念被理解为由外涵和内涵所组成的思维单元。它们按照某种规律组合成语言以表达人们的思想。基于概念的这一逻辑理解。是德国数学家 WIlle1982 年所提出的形式概念分析，又称为概念格理论。用于概念的发现排序和显示。**概念格**是一种基于概念和概念层次化的数学表达，在概念格理论中。形式背景表达为一个二维数据表的形式。基于形势背景的概念是最小的**知识单元**。称为**概念粒**。概念格就是通过概念粒之间的层次关系。给出形式背景上知识的结构化表示([18], p. 243)。

一个汉字 = 一个概念 + 一个形体 + 一个音节，是形音义的结合体，这是语言学家徐通锵的字本位理论。笔者认为，汉字是知识概念最小粒结构的映像与表现。

4.5. 知识、知识粒度和知识库的数学定义

在关于信息熵与粒计算的论文中，苗夺谦给出了基于粗糙集的知识、知识粒和知识库的基本定义([18], p. 143)。

• **知识**定义：设 U 是非空有限集，称为论域。任何子集 $X \subseteq U$，称为称为 U 中的一个概念。U 中的一簇概念，称为关于 U 的知识。

记 $A=\{X_1, X_2, \cdots, X_n\}$，若满足：

1) $X_i \subseteq U, X_i \neq \varnothing$。

2) $X_i \cap X_j \neq \varnothing, i \neq j, i, j=1,2,\cdots,n$ 。

3) $U_{i=1^n} X_i = U$ 。则称 A 为 U 的划分。([13], [粒计算 143 页])

• **知识库**的定义

知识库可以形式地定义为序对 $K=(U, R)$，其中 U 为论域，R 为 U 上的等价关系簇。域等价关系 $R \in \boldsymbol{R}$ 为知识，称 R 生成的等价类 $[u]_R$ 为基本知识颗粒。称商集 $U/R=\{[u]_R \mid u \in U\}$ 为论域 U 的 R-粒划分。([18], p. 243)

粒计算的数学理论严格地证明了**粒度**、**知识分辨度**和**知识熵**之间的关系。**熵值**。也是知识颗粒状的一种度量。因此。涉及到汉字的熵的计算。专业领域汉字集合的熵。从而进行知识工程结构分析的基本四个粒度层次：字、词和短语与句子的计算。

更高层的层次结构分析则涉及到段落、篇章。从粒计算的角度出发。不难解决这些句子、段落和篇章的计算以及它们之间的相似度。运用 Jakcad 的公式算法。可以计算出之间的相似度。这样就完成了海量文本基于粒计算理论指导下的知识抽取、分类和聚类的计算工程。因此陈霖院士所提出的认知科学的基本单元。在其中就承担了重要的角色，获得实践应用的验证。

杨炳儒指出：“图论通过点和线的构型来构成模拟各类系统的数学模型，并根据图的性质进行分析，提供研究各种系统的科学的、巧妙的方法。任何一个包含了某种二元关系的系统都可以用图论的方法分析，而且它往往具有形象直观的特点。” “图论的研究具有十分广阔的客观原型，并在许多领域中有着极其广泛的应用([27], p. 184)。

综上所述，以粒计算为特色的知识工程，离不开离散数学关系代数理论的支持。它是建立在离散数学的基础之上。维纳指出，离散数学的理论。要比连续数学的理论简单得多。在人脑的分析研究过程中。类脑结构一直是研究的重点。应该引入关于粒计算的最新研究成果。引入多学科交叉研究自然科学与社会科学具有共性的人工智能问题。把粒计算的理论与第三代人工智能的理论与技术，与离散数学图论的理论与算法结合起来。这也是是当前重点的研究方向。

知识的构成核心就是概念与概念的关系，粒计算理论是最好的知识工程工具。知识理论和知识技术研究都离不开代数语言学的关系计算。

不同层次的知识粒，是知识单元的基本表示。同层单元粒与粒的关系聚集成更高层次的粒。下层的粒是上层粒的分解单元。

4.6. 知识单元的粒结构层次：

结构层次：基本粒的组成分为六层。具有不同的粒度

1) 字基本单元：简称字元，是基本粒。

2) 词基本单元：简称词元，词粒

3) 语基本单元：简称语(块)元，语块粒

4) 句基本单元：简称句元，句子知识粒

5) 段基本单元：简称段元，段知识粒

6) 篇章基本单元：简称篇元，粒度最大－篇章知识粒。

图 3 展示了不同粒度知识单元各层的组网特性，以及由较小粒度聚合为高层次较大粒度复杂网络图形。每一层的知识粒，由下层粒组合而成。必要时赋予大概念标识符(层的粒标志)与各种属性。分层知识库各层知识单元复杂网络的集合，构成知识总库。

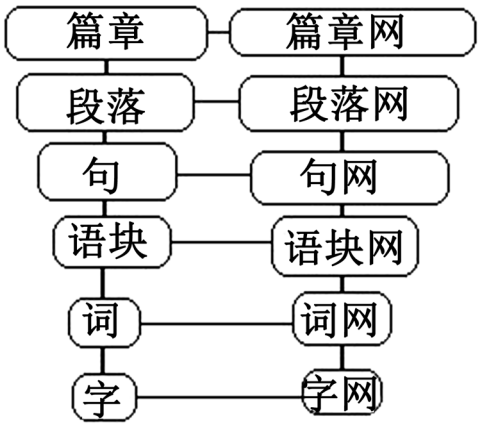

Figure 3. Multi-layer three-dimensional complex network composed of six different granular units
图 3. 由六种不同粒度单元构成的多层三维复杂网络

在同一层中，粒与粒作为结点。相互之间构成邻接矩阵，形成复杂网络。图 4 是字、词和语块三种粒度的只是相关连接图。清楚第描绘出由字构词，由词毗连成语的相互关系。

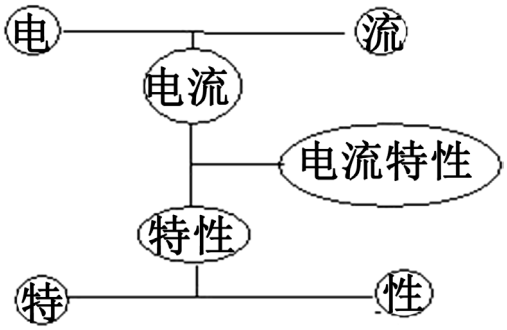

Figure 4. Example of the structure of a word block (term)
图 4. 字－词－语块(术语)的结构实例

图论中点与边的集合，就是知识基本表示单位。在复杂网络系统结构中，知识就是万万千千个结点之间的相互关系。如果把符号看做复杂网络中的一个点，那么一篇文章就是这万万千千点的集合，涉及把序列链结点到并行树结构的一系列关系计算。这个庞大数据的立体交叉网网络的建设、存储、检索和推理，构成了知识服务的核心。图 5 是现代汉语词典 100 个字与字构词关系复杂网络图。

图 6 是以一个汉字“羊”为核心的构词关系图，除了和其他汉字构成向心和离心词汇外，相关的汉字也有一系列的构词关系。所有这些关系都可以用邻接矩阵来描述。

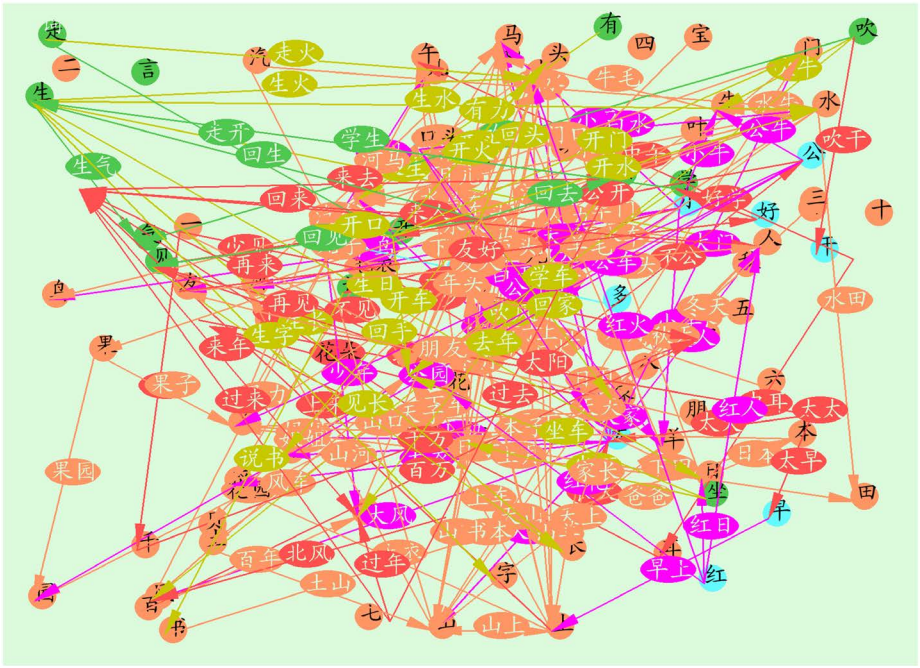

Figure 5. The network diagram of the word formation relationship between Chinese characters and characters
图 5. 汉字与汉字的构词关系网络图

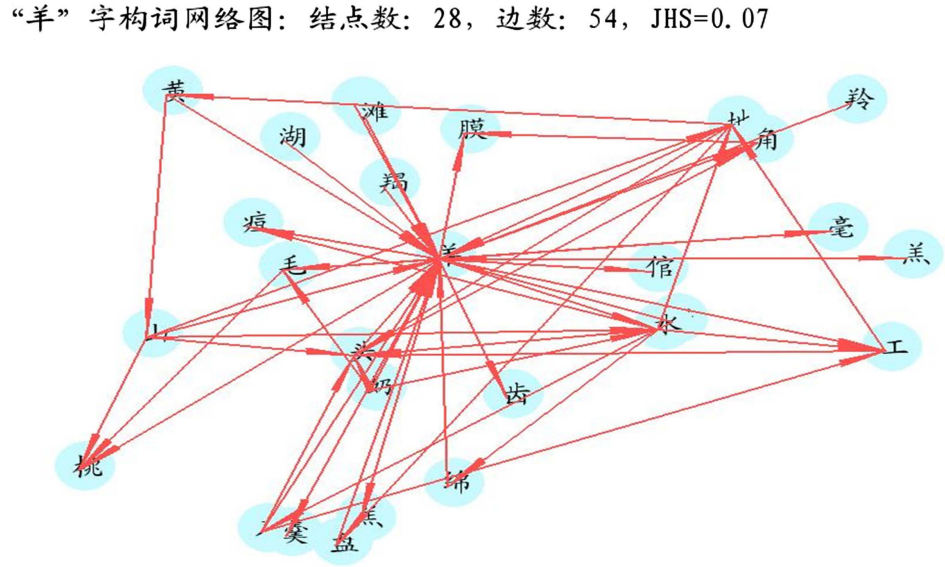

Figure 6. The relationship between the Chinese character “羊” and other characters
图 6. 汉字“羊”与其他字的一度关系图

在专家系统与人工智能问答机器人的系统设计中，知识库是系统应用的核心，而知识库的框架体系结构与知识单元的粒度合成与分解就成为知识工程的核心技术。数学家、控制工程专家都需要能够清晰系统结构的粒计算数据集。组织和各种基于粒度的算法设计也需要规范化的不同粒度的数据集。例如，中医方剂百余种药材的不同集合对应着万千种病症的集合。诊断治疗过程中疗效与症状变化的反馈决定治疗过程中方剂的改变或加减某种药材及其剂量。

笔者根据粒计算的理论与应用的核心技术，进行了大型知识库的粒计算结构设计，并在中医专家知

识库的建设中做了部分试验和人机对话信息精准解析的尝试。实验结果表明，粒计算方法理论扎实，具有实用性。并可以在医疗知识问答机器人的设计与生产中，获得广泛的工业应用。提供给各层次水平的服务对象。扩大医疗知识服务的领域，为建立世界范围内的中医系统化知识互操作做出贡献。

5. 知识工程与复杂网络

5.1. 知识工程与图论

知识工程的实践，需要理论研究提供严密的支持。大数据的生产涉及具有数十亿元素的大型矩阵，涉及这些元素关系的相关计算。

在几十种世界各国语言的科学文献文本之间进行相似度计算中找寻科研项目相关文献的集合，为科研提供基础材料，是知识工程对国家大性计算算力的殷切期望。当然科研工作者也可以长年累月地开动个人计算机，但大型立体关系矩阵的计算。需要国家动用大型巨型计算机来支持进行数据处理。这也是基层科研工作者长年累月地开动个人计算机所不能完成的。

图 7 是一个汉字与其他汉字的 SM 小世界网络理论的六度关系图。即围绕一个汉字，构成不同层次的关系群(类似微信朋友圈)。例如：图中与“蛇”字距离为 3 的字有 3605 个。从而进行汉字距离矩阵与词汇聚类形成词群的有关研究。

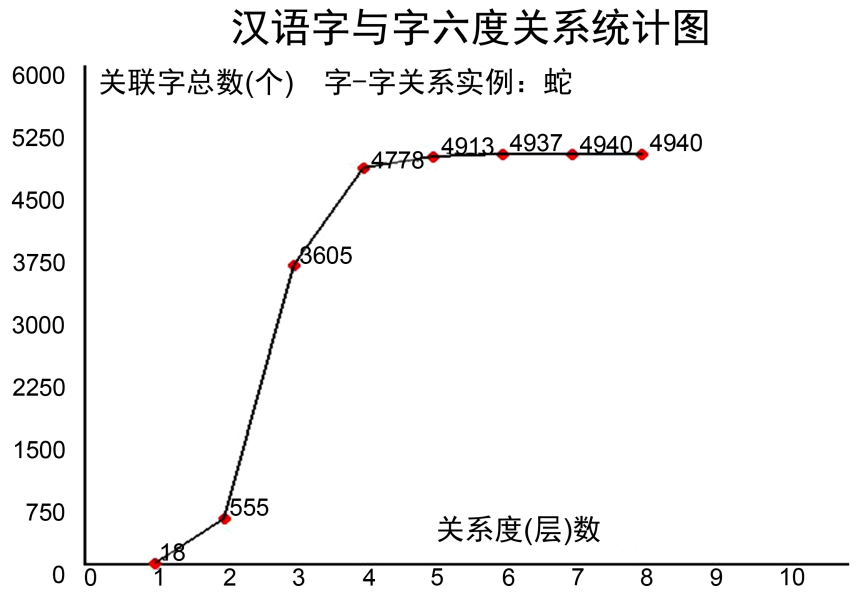

Figure 7. The distance relationship between the Chinese character “snake” and other Chinese characters (the total number of Chinese characters in the domain is 4941)
图 7. 汉字“蛇”与其他汉字的距离关系图(论域汉字总数 4941)

在数理语言学中，一个句子，就是字链接的一条有序节点集合。字与字的关系就是我们研究的重点。中华民族 5000 年来留下的历史文献当中。都是一种字的连续流。无论是古典文献还是现代电子文献。都是一种序列字的集合。而且汉字是和概念直接相对应的。在任何一篇文章当中,字与字的关系都是有序的。两个字之间。也就是两个概念之间的关系是明确的。这种关系可以分为多种类型。最简单的是两个字之间邻接关系，要么是构词关系。要么是非构词关系。字与字的关系。可以构成邻接矩阵。用图论矩阵来进行分类与聚类的使用研究，如文本聚类和分类。图 8 是现代汉语 100 个核心汉字的字与字构词关系图。

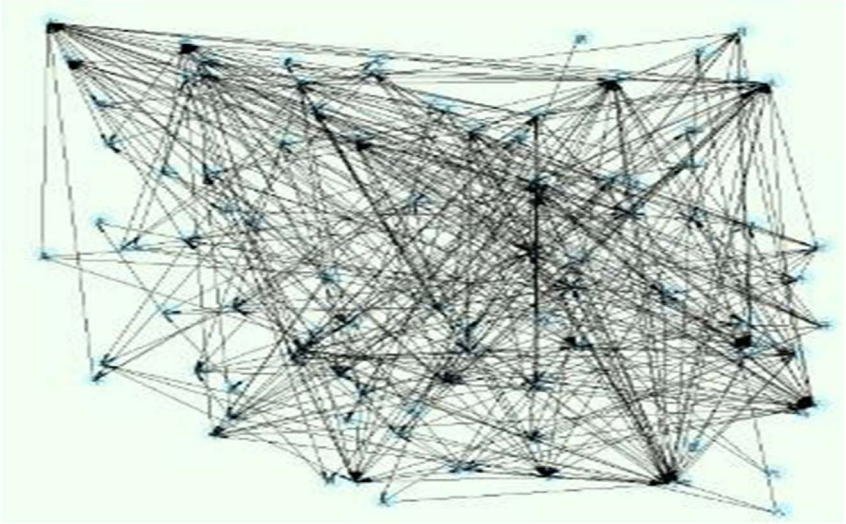

Figure 8. Word formation relationship among 100 Chinese characters in modern Chinese Dictionary
图 8. 现代汉语词典核心汉字 100 个之间的构词关系图

综上所述。任何篇章知识、段落知识和句子知识的构成，都是一种集合。不同语言，不同粒度的节点之间，则构成了层次分明的复杂立体网络。

节点与节点之间的连接构成路。一篇文章，就是链式符号的有序集合。离散数学的集合理论用于人工智能机器翻译。各种语言文本之间翻译。实质就是群集合之间互相映射的关系。

人工智能的本质，就是辅助人脑功能的延长。就是把人类社会的科学知识，用智能的方法依靠智能技术工具去解决国民经济生产与建设问题。解决人民生活医疗教育的实际需要。从这一点分析。知识工程，将对人类社会产生巨大的影响。科研是知识提高再生产发现与发明的工程，教育工程是知识普及工程，文化传播实质是知识的传递工程。

著名的科学家诺依曼指出：人脑的语言不是数学语言。诺伊曼在他人生最后一本著作《计算机与人脑》中。比较了天然自动机与人工自动机的区别。他指出。天然自动机的速度远没有计算机自动机的速度快。但是却完成了庞大复杂的逻辑运算。以人的视觉。听觉和触觉综合迅速的进行信息处理。这一点在第三代人工智能的研究当中。应该予以高度的重视。

张钹指出。第三代人工的智能的特点。是把符号主义与联结主义相结合。符号主义的核心。在于基本符号表达的概念世界。符号世界、人脑的概念世界和描述世界万千的事物对象客体世界。三者构成了知识世界。符号学理论及其应用的研究。在我国很少见到。数学的研究是一系列逻辑严格的推理，在最新的研究当中。在知识经济、知识工程、知识技术和知识服务中迫切需要多学科交叉和交叉科学的研究。

5.2. 知识理论与知识技术

在赫尔穆特·费尔泊的《术语·知识论和知识技术》中。论述了三个世界：符号世界、现实世界和人脑世界三者之间的关系。现实世界。包含的万事万物的运动。映像在人脑中。在人脑形成千丝万缕的联系。同时。现实世界的万事万物。所反应的概念。也映像在符号世界中。因此，类脑计算的研究就是对万事万物之间关系在人脑中映像的研究。其相互之间复杂关系的研究。这种研究也映像在表示概念的符号之间关系的研究当中，这就是**符号计算**。符号计算，将引入大规模的矩阵计算。离散数学、图论和复杂网络理论，在类脑计算中将发挥重要作用。

知识检索以及知识服务的管理、查找、匹配成为知识管理研究的重点。相似度计算是管理、查找、

匹配的核心。相似度的计算有多种方法，可以单独也可以结合使用，从而提高相似度计算的准确性与实用性[28]。例如：十万中医方剂的相似度计算与分类，方剂中药材集合与症状集合的关系研究问题。

各种模型，算法都是知识技术，数学模型，类脑模型，各种计算定理与公式，也是一种工具。都是知识生产与处理过程中的一台台自动机。在知识工场中运行，得出产品。

现代知识工程的工具是计算机，从计算机科学的角度分析，只有两种符号表示的数据，那就是程序于数据，程序就相当于人的大脑，对输入数据计算后输出数据。知识工程的各个环节都是自动机，一种处理各类输入产生输出的自动机。

基于互联网技术的云资源库中有着亿万本书籍和科技文献，经过加工转换标注等各种技术进入云知识库，经过云计算产生无数为人类生产与生活使用的数字化产品，也指挥万万千千工业、农业、交通运输业、商业的万千设备生产产品。从这点分析，回归理论与实践结合的人工智能的应用，是本质，也是初心。

“科技术语云工程，是采用云计算的技术方法，对海量知识资源进行处理的工程，建立语料库，加工语料库提取数据生成术语数据库。在语料库和术语数据库的基础上进一步使用人工智能的方法，构建服务于国民经济建设、科学研究和知识普及的知识库。这三大数据库群构成科技术语云工程的资源池，也是实施科技术语云工程的关键。只有在完成系统框架设计的基础上，充分研究知识组织系统的分层结构设计，才能通过网络接口平台，把互相关联的术语与知识送给亿万用户[29]。知识库涵盖了人类积累的全部数学、物理、化学和生物科学等各学科的知识。知识技术与知识工程的实践，为理论研究、计算研究和和类脑研究，提出了提供技术支撑，解决大量需要用理论与实践结合的一系列问题。当然也需要国家规模大型计算机构成的能完成海量数据立体矩阵群的运算群组与大型算法库群。在人工智能与机器人研究中。离不开哲学理论、数学理论、知识理论、计算机科学理论和现代控制工程理论的支持。这是一个重要的结论，交叉科学学指导着第三代人工智能的研究方向与重点。

6. 结语与讨论

6.1. 进一步重视汉字本体的应用研究

符号串，是语言学中形式化方法的基本概念。语言是词的集合上的自由幺半群([30], p. 67)。数理语言学所讲的形式化语言就是有一定特征的句子的集合。也就是定义在词汇集合上的自由幺半群的一个子集[30]。

汉字是知识工程的最基本单元。也是中文所有文献与书籍的基本字符串。一切知识的表达，用数理语言学分析，都是字符串的流集合，可以像串并转换那样，转换为二叉或多叉树的集合，转化为多维向量运算。多国语言的字符串数据也可以映像到以概念为核心的映像。各国的语言词汇的字根反映的是本质，例如：苹果，在哪种语言中都是相同的客体。大型立体映像矩阵将应用于翻译工程实践中，其科学化的汉字系统将发挥重要作用。这也是对汉语和汉字未来百年的预测。

从这点分析，必须加强认知科学最基本的元素——汉字本体的研究。

6.2. 注重多学科协作的复杂网络技术研究

研究知识工程的基本单元——汉字与构成系统的知识复杂网络，就是研究科技知识及其描述的系统性。汉字及其表意文字与概念相关的固有属性，应该也有可能成为世界范围的知识组织系统工作平台核心。因此，构建基于汉字与中文的知识工程，把具有中国智慧，中国方案的知识推向世界。首先，知识库的设计框架是知识分类体系和基于汉字本体所构成词汇体系的彼此交叉，推进基于知识工程的广泛应用。

其次，知识库的核心层次结构是汉语的字、词(语)、句，通过对自然语言的逻辑描述和句法分析，推进以汉字为核心的世界各国技术语言互操作系统。实现新时代全球科学技术语言的交流的自动化、系统化。

6.3. 加强对中华民族大型古籍外译的研究

中华民族传统知识向全世界传播的工程，需要集中全国各专业的相关力量来实现各学科交叉条件下合作的艰巨任务，是中华民族传统文化走向世界的迫切需要。这个工程急需要总体的战略规划，也需要有各领域的科学家、技术工程专家的艰难探索。

古代前贤在以皇帝为首的国家大型古籍编辑整理的知识工程中，动用数百人，历经数十年的建设成果。迫切需要我们这一代继承那锲而不舍，经年勤奋努力的编辑加工的建设精神。

笔者作了多年的中医专家系统知识工程的探索实验。可以产生数十万医学医理方剂针灸医案等与疾病相关的知识卡片。已经为中医科技术语的抽取提供科研信息。并根据以上理念，力图为大型医学古籍人工智能检索、研究和外译做前期性研究提供各类重组数据，以期抛砖引玉，做出自己力所能及的一点贡献。

新时代需要发挥中国智慧、传递中国声音、传播中国精神，只有顽强拼搏和探索，才能使中国人工智能在知识工程方面的研究和应用提升到一个新高度。

6.4. 加强全球范围的知识服务互操作项目的建设工程

全球知识互操作将不断促进世界交流，极其迅速地更新的人类科技知识。建立基于知识工程需要的系统，需要加强以下三类研究：一是以概念(汉字本体为其符号表示形式)为基础的语义网的结构研究。二是术语与术语核心词汇之间构成的复杂网络的技术研究。三是面向系统化、结构化和工程化的知识服务体系研究。

机器人只是辅助人类生产生活的重要工具。各种交通工具是人脚功能的延伸。注入智慧思考各种算法的机器人永远是辅助人脑工作的机器。永远不能超越人类，人类大脑是数亿年演化的结果，把人类大脑与整个人体的控制结合起来。蚂蚁社会的通信，蜜蜂社会的通信和分群，将启示我们研究蚂蚁、蜜蜂和各种动物的大脑。脑科学与类脑科学将有极其漫长的路要走。只要看物理学关于光“波粒大战从牛顿到爱因斯坦持续 300 年”就可以了解人类对自然认知的艰难探索史，就可以从人类生命的有限性和认知客观世界的无穷性出发，认识与解决现有急需问题的迫切性。

用人工智能的理论解读引领知识工程，解决国家经济建设的实际应用，推动中国文化与科技走向世界，是当前的重点。

最后，仅以中国人工智能学会理事长涂序彦 2005 年为《智能科学技术著作丛书》序的论述作为本文的收尾。

涂序彦指出：“智能”是“信息”的精彩结晶，“智能科学技术”是“信息科学技术”的辉煌篇章，“智能化”是“信息化”发展的新动向、新阶段。

“智能科学技术”是关于“广义智能”的理论方法和应用技术的综合性科学技术领域，其研究对象如下：

“自然智能”，包括“人的智能”和其他“生物智能”。

“人工智能”，包括“机器智能”与“智能机器”。

“集成智能”，即“人的智能”与“机器智能”人机互补的集成智能。

“协同智能”，指“个体智能”相互协调共生的群体智能。

“分布智能”，如广域信息网、分散大系统的分布式智能。

如果说，当年“人工智能”学科的诞生是生物科学技术与信息科学技术、系统科学技术的一次成功的结合，那么，可以认为，现在“智能科学技术”领域的兴起是在信息化、网络化时代又一次新的多学科交融[31]。

基金项目

国家社科基金项目“中国学派背景下汉语术语学学术话语体系建构及俄译研究”(19BYY212)阶段性成果。

参考文献

[1] 秦陇纪. 徐匡迪院士之问揭开当下中国人工智能虚伪的面纱[EB/OL]. https://www.sohu.com/a/312151330_680938, 2019-05-06/2020-11-24.

[2] 人工智能是天使还是魔鬼？谭铁牛院士指取决人类自身[EB/OL]. http://news.ifeng.com/c/7dIwxsqBzWt, 2017-05-28/2021-02-04.

[3] (英)伊姆雷·拉卡托斯. 证明与反驳-数学发现的逻辑[M]. 康宏奎, 译. 上海: 上海译文出版社, 1987.

[4] 全国信息技术标准化技术委员会. GB/T 5271.28-2001 信息技术词汇第 28 部分: 人工智能基本概念与专家系统[S]. 北京: 中国标准出版社, 2004.

[5] (奥地利)赫尔穆特·费尔伯. 术语学知识论和知识技术[M]. 北京: 商务印书馆, 2011.

[6] 张钹、朱军、苏杭. 迈向第三代人工智能[J]. 中国科学: 信息科学, 2020, 50(9): 1281-1302.

[7] (美)约翰·奈斯比特, 著. 大趋势[M]. 北京: 新华出版社, 1984.

[8] 马占华, 侯汉清, 薛春香. 文献分类法主题法导论[M]. 北京: 国家图书馆出版社, 2009: 341.

[9] 诺意曼. 计算机与人脑[M]. 甘子玉, 译. 北京: 商务印书馆, 1965: 58.

[10] 自科在线. 重磅, 国家自然科学基金委第 9 个科学部来了! [EB/OL]. http://n.eastday.com/pnews/1604116203025721, 2020-10-29/2020-11-23.

[11] 陈洪澜. 论知识分类的十大方式[J]. 科学学研究, 2007, 25(1): 26-31. http://dx.chinadoi.cn/10.3969/j.issn.1003-2053.2007.01.006

[12] (汉)许慎撰. 说文解字[M]. (宋)徐铉, 校定. 北京: 中华书局出版, 2013: 321.

[13] (清)张廷玉 编. 骈字类编[M]. 北京: 中国书店出版, 1984 .

[14] 陶伯华, 朱亚燕. 灵感学引论[M]. 沈阳: 辽宁人民出版社, 1987: 272-237.

[15] 陈霖. 新一代人工智能的核心基础科学问题: 认知和计算的关系[J]. 中国科学院院刊, 2018, 33(10): 1104-1106. http://dx.chinadoi.cn/10.16418/j.issn.1000-3045.2018.10.011

[16] 宋培彦, 路青, 刘宁静. 一种从术语定义句中自动抽取知识单元的方法[J]. 情报杂志, 2014(1): 139-143.

[17] 齐熠, 都立澜, 李晓莉. 试论中医术语翻译中的翻译单位问题[J]. 中国科技术语, 2008, 20(5): 37-41.

[18] 苗夺谦, 王国胤, 刘清, 林早阳, 姚一豫. 粒计算: 过去、现在与展望[M]. 北京: 科学出版社, 2007: 6-7.

[19] 陆俭明. 亟需解决好中文信息处理和汉语本体研究的接口问题[EB/OL]. http://www.cips-cl.org/static/CCL2020/invited.html, 2020-10-31/2020-11-24.

[20] (英)尼古拉斯·奥斯特勒. 语言帝国[M]. 上海: 上海人民出版社, 2016: 143.

[21] 弗迪南·德·索绪尔. 普通语言学教程[M]. 高明凯, 译. 北京: 商务印书馆, 1985: 50-51.

[22] 刘洪波. 英文字根词源精讲[M]. 北京: 中国广播电视出版社, 2007: 序.

[23] 王宁著. 汉字构形学导论[M]. 北京: 商务印书馆, 2015: 219.

[24] 杨光治. 鲁迅曾称: 汉字不灭中国必亡[EB/OL]. https://cul.qq.com/a/20150729/023277.htm, 2015-7-29/2020-11-24.

[25] 马君花. 消失的部首字——从《说文》到现代汉字记号化进程的研究[J]. 图书馆理论与实践, 2014(12): 97

[26] 魏迈. 汉字部首解说[M]. 北京: 商务印书馆国际有限公司, 2015: 8.

[27] 杨炳儒. 图论概要[M]. 天津: 天津科学技术出版社, 1990: 184.

[28] 董金祥, 主编. 基于语义面向服务的知识管理与处理[M]. 杭州: 浙江大学出版社, 2009: 155.

[29] 杨福义. 云计算与相关术语概念的探讨[J]. 中国科技术语, 2018, 20(5): 47-51. http://dx.chinadoi.cn/10.3969/j.issn.1673-8578.2018.05.009

[30] 冯志伟. 数理语言学[M]. 北京: 商务印书馆, 2012: 68.

[31] 郑家恒, 张虎, 谭红叶, 钱揖丽, 卢娇丽. 智能信息处理: 汉语语料库加工技术及应用[M]. 北京: 科学技术出版社, 2010: 序.