



Collaborative Intelligent Computing System—Theoretical Model with Its Application

Xiaohui Zou and Shunpeng Zou

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

January 27, 2025

协同智能计算系统——理论模型及其应用 *

邹晓辉^{1,2} 邹顺鹏¹

1 中国地质大学（北京）高等教育研究所（原创） 北京 100083

2 北京大学跨学科知识建模课题组（2025 标读） 北京 100871

摘要：本研究工作目的是揭示计算机数据与自然人知识两类信息处理方式基础之上派生的协同智能观及其指导下的协同智能计算系统的理论模型及其应用。它涉及的可验证方法，一方面，在 n^2 矩阵范围内，以等价于 2^n 的**发散方式**枚举和以等价于 $1/2^n$ 的**收敛方式**搜索**两类基本算法**可用作处理纯数字计算的任务，其特征是满足 n 的取值不影响计算效率的**间接计算**任务；另一方面，在与 n^2 矩阵各个格子一一对应的范围对**单音节汉字**进行**间接形式化**处理，其特征，不仅在于**单音节字，即言，可间接计算**，而且，还在于**双音节和多音节的字组，即语，也可间接计算**，同时，**言和语的复用频率**均可以且便于**间接计算和统计**。其结果是：不仅中文的自然语言理解的双重技术路线被揭示，而且支配这类间接计算模型与间接形式化方法的信息基本定律假说也可被验证。最终可得出这样的**结论**，即：在前述两方面可验证的两种实证方法，远不仅仅是计算机数据信息处理方式与自然人知识信息处理方式这两类信息处理方式的简单相加，而是这两者合理分工、高度协作所产生的协同智能计算系统的理论模型或第三类信息处理方式及其应用，例如：国内外学术前沿的各类期刊及会议论文摘要、各种软件的常用问题解答，以及帮助文件、协同智能计算系统用户个性化记录、自然语言的有限符号及其多样化组合或重复使用过程中蕴含的有限规则、等各类双语信息的计算机辅助分析，该类云端计算主要服务对象是在创造性合作型生产式教研产学研用各类活动中需要计算机辅助双语知识信息数据处理服务的客户。

关键词：计算机；间接计算；间接形式化；软件

中图分类号：TP18

文献标识码：A

DOI: 10.3969/j.issn.1003-6970.2011.06.001

Collaborative Intelligent Computing System: Theoretical Model with Its Application

ZOU Xiaohui, ZOU Shunpeng

*Institute of Higher Education at China University of Geosciences
(Beijing), Beijing 100083, China*

【 Abstract 】 This study is aimed at revealing the view of Collaborative Intelligence derived from the mode of two types of Information Processing, namely computer data and human knowledge, as well as the theoretical model of Collaborative Intelligence Computing System with its application guided by this view. It involves a verifiable method. On the one hand, within the range of n^2 matrix, the two basic algorithms, which are enumeration based on divergence equivalent to the way of 2^n and search based on convergence equivalent to the way of $1/2^n$, can be used as tasks of pure digital computing. It is characterized by satisfying the condition that the value-taking of n does not influence computational efficiency of the Indirect Computational Tasks; On the other hand, within the range of n^2 matrix corresponding to each grid, Indirect Formalization Processing is used towards single character, which is characterized not only by that single-Zi-syllable, namely Yan, can be calculated indirectly, but also by that two-Zi-syllable and multi-Zi-syllable, namely Yu, can also be calculated indirectly. At the same time, the frequencies of the reuse of Yan and Yu can both be easily counted and calculated indirectly. Thus, this model can not only lead us to revealing the double technique route toward Chinese Information Processing or Natural Language Understanding, but also to verifying the hypothesis of Informatics Basic Laws, which dominate the Indirect Computing Model and Indirect Formalization Method. Eventually, it can be concluded that the previous two verifiable empirical methods are much more than a simple sum of these modes of the two types of Information Processing, namely computer data and human knowledge, but the theoretical model of Collaborate Intelligent Computing System or the third

kind of Information Processing Mode with its application generated by Rational Division of labor, highly Collaborative Synergy, such as academic forefront of various journals and conference abstracts home and abroad, answers to Frequent Answered Questions of a variety of software and their Help Files, individual records of Collaborative Intelligent Computing System users, limited symbols of natural language and its limited rules inherent in the diverse combination process or the reuse process and so on, which are the Computer-Aided Analysis of all kinds of Bilingual Information. The main target of such cloud-based computing is the customers who need Computer-Aided Bilingual Knowledge and Information Processing in various types of creative cooperative and productive activities, such as teaching, research, production, learning and using.

【Key words】 Computer ; Indirect Computing; Indirect Formalization; Software

* 基金项目：中美合作项目：双语信息处理 (No.20110126) *Foundation Items:
SINO-US Project: Bilingual Information Processing (No.20110126) 作者简介：邹晓辉
(1958-), 男, 研究员, 研究方向：双语信息处理, 协同智能计算系统 ; 通信联系人:
邹顺鹏 (1986-), 男, 研究生, 研究方向：教育领域知识本体 .

0 引言

本文旨在阐述协同智能计算系统的理论模型及其应用，其特征是间接计算模型与间接形式化方法^[1]的结合而产生第三类信息处理方式及其应用。它不仅仅是计算机数据信息处理方式与自然人知识信息处理方式两类信息处理方式的简单相加，而是，这两者相互之间的合理分工、高度协作所产生的协同智能^[2]计算系统的理论模型及其应用。

计算机科学技术界可以理解人们做这样的假设：如果人工智能^[3]是信息技术的皇冠，那么，自然语言理解^[4]就是该皇冠上的一颗明珠。进一步也可以理解我们做这样的假设，即：如果自然人的大脑智能是第一智能，而计算机的电脑智能是第二智能，那么，基于整体大于局部之和的系统科学原理，是否可以把前两种智能的结合称之为第三智能^[5]呢？

我们认为不仅可以这么说，而且还可以这么做的。本文的一个研究切入点同时也是一个原创点就是从这个思路来做自然语言理解的，因此，才可能发现并确信自然语言理解存在着双重技术路线，进而也才可能揭示第二路径的科学机理并发现其在双语信息处理上的妙用，这不仅涉及语言学基础研究^[6]一个突破，而且，也涉及信息学基础研究^[7]另一个突破，同时，还涉及教育学和管理学两个领域基础研究^[8-10]的又一个突破。

这怎么可能呢？真是让人难以置信！但是，经过近十几年在上述几个相关研究领域的探索、研究和广泛的国际国内交流，最终，得到了确切的研究结果、结论和具体的应用示例。

下面就把难以置信变为确信无疑的探索历程介绍给读者。

1 什么是协同智能计算系统的理论模型？

1.1 抽象理论模型可计算结果和应用方向的直观表示

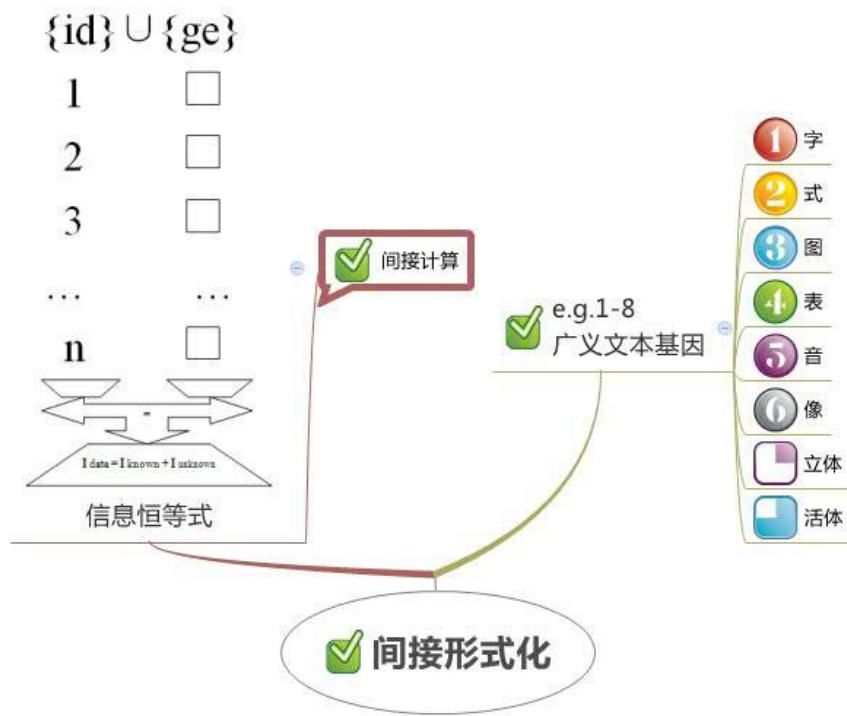


图 1 协同智能计算模型及其原理和应用的直观示意图 [11]

由图 1 可见，左边的信息恒等式 $(I_d = I_k + I_u)$ ^[12]，其中， I_d 指称目标域的总量数据信息、 I_k 指称已知域的知识信息和 I_u 指称未知域的未知信息或相当于人们常说的情报信息^[13]，其上 $\{id\} \cup \{ge\}$ 两个同义并列的形式化数据集合，则由一个虚拟的天平加以描述，其中在左列表内的十进制数 $\{id\}$ 与在右列表内的空格 $\{ge\}$ 之间的关系，相当于标准化的砝码与个性化的物品，两者均可放在一个虚拟天平上权衡称量。这个虚拟天平，与我们日常所见的实物的天平可被视为是同类但不同例。这个虚拟天平就是图 1 左边的这个双列表，它通过左列表 id 编号依次排序的标准数字（前台是十进制数 id ，后台是与其对应的二进制数）来称量右列表 ge 空格预留的非标对象，这些非标对象是图 1 右边的广义文本基因，即：“字、式、图、表、音、像、立体、活体” 八类对象的任何一类或多类的混合调用。这里所谓广义文本基因是指这八类对象均可通过解析的方式提取出其基因元素从而分析出其排列组合的基本特征。这就是间接计算模型与间接形式化方法及其结合而建立的基础。

协同智能计算系统的理论模型本质上是以下数学方程组。

1.2 计算序位关系的两个基本方程

协同智能计算模型就是计算序位关系的两个基本方程，即：

$$I_D = n^2 \quad (1)$$

$$I_V = I_D - I_K \quad (2)$$

其中，(1) 广义的信息方程，即：以纯数字的矩阵表达的抽象的形式信息方程。(2) 狭义的信息方程，即：与该矩阵格最优化算法和基本数据结构类型表达的序位关系之间有对应关系的虚拟矩阵范围之内可表达的内容信息方程。本文在以下将阐述的原理及其应用，均与上述两个基本方程及其应用有关。图 1 所示的可计算结果和应用方向，也只是这两个方程的变形及其直观表示。为什么可以这样说呢？其理由既很简单，又很复杂。其一，抽象的形式信息方程，可以什么意思都没有，也可以说，什么意思你都可以赋予给它。其二，内容信息方程，或者说，具体的内容信息方程，其实也就是，在具体的用户从图 1 所示的可间接形式化或间接数字化且可间接计算的任何一类或多类广义文本基因及其派生演绎的广义基因文本明确地选订某个部分放入图 1 所示的空格 (ge) 中称量的时候，所有可能蕴含的内容信息就已被相应的形式信息所确定了，也就是说，信息恒等式 ($I_D = I_K + I_V$) 已经预设了其最终可能的结果。对具体用户而言，不同的，只是他或她在限定目标域的总量数据信息 (I_D) 之内所掌握已知域的知识信息 (I_K) 和暂时没有掌握未知域的未知信息 (I_V) 之间的比值在数量上和结构上以及在性质上的具体差异罢了，这些，在理论上都是可计算的——至少是可间接计算的，而在实际上却要视用户所处的云端计算入口的情况，即搜索、检索、查询的访问权限设置而定。为搞清楚笔者的上述信息方程及其蕴含的信息概念的学术渊源，有必要补充适当的背景知识：哈特莱^[14]从指数计算对数化的策略这个角度得到的信息公式 ($I = H - 0 = N \log S$) 以及仙农^[15]从信息熵的概率分布这个角度得到的信息公式 [$I = H_s(p_1, \dots, p_n) - 0 = -K \sum p_i \log p_i$] 与邹晓辉从自然数及其矩阵分布这个角度所得到的信息公式 ($I_V = I_D - 0 = m n \approx n^2$) 是当 $I_K = 0$ 时 $I_V = I_D - I_K$ 的一个特例，就是在不考虑已知域的知识信息

时的情形)的区别,三者的联系则可视为当 $H = I_D$ 时的情形。这涉及:数学、通信和计算机科学,以及知识学习的交叉研究,属于**信息科学的基本理论研究领域**。在仙农信息论提出之后的其他研究几乎都忽略了这里所指出的一个情形,即:**广义信息方程**也可是**狭义信息方程**的特例。就是说,广义和狭义不过是站在不同角度来看待同一对信息方程及其所蕴含的**信息概念**而得到的**不同说法**罢了。可见,真正的经典论著并不会过时,相反更加值得读者反复地深究。笔者正是由此获得三个很关键的基本判断:

其一,如果仅从形式信息处理的角度来看,那么,信息恒等式与广义信息方程之间的关系可视为是完全等价的;

其二,如果仅从内容信息处理的角度来看,那么,信息恒等式所蕴含狭义信息方程所有可能的各个分布解。这怎么可能呢?真是**难以置信**,但是,这就是可能,而且,还可以让人**确信无疑**,只要遵循严格的形式推理。

其三,如果不得不**兼顾内容和形式两个方面的信息处理**,那么,信息恒等式就成了广义信息方程与狭义信息方程的**联系纽带**。于是,纯理性的形式化计算就与充满不确定性的**经验内容乃至两者结合的分析**和进一步的**解释**就都因为主体及其代理共同遭遇的时间和空间以及载体载能而**纠结到一起**了。笔者认为这是一个**误区**。这也就是为什么会有许多的人**长期地**把内容信息处理的**复杂性**以及主体选择的**多元化**乃至主体间的**社会约订**诸因素集群都统统**扯进了**广义的**意**(笔者在此,特意使用它“**广义的意**”以区别于:广义的**义**)的**旷日持久**的信息概念**纷争之中**,与此密切相关的,还有语言和知识乃至教育,这样的概念纷争,其理论研究领域**长期不能够完善其学科的内在建制**^[16]。笔者区分**取值与置信**,对此不失为是一个有效的策略^[17]。

1.3 协同智能计算系统的理论模型及其应用所依据的协同智能观

鉴于**广义文本基因**,及其派生演绎的**广义基因文本**均可直接**数字化**和**间接数字化**,笔者便可通过引入**间接计算模型**和**间接形式化方法**按照统一的**砝码或尺度**来称量它们。这也就给**协同通智能计算**提供了**切实可行的途径**。

也正因为如此，我们可称之为协同通智能计算模型。它不仅可在理论上很好地解释协同智能观 ($CI \geq HI+AI$)，而且，还可在实际中体现符合其要求的协同智能计算，并且在实践上建构出可应用于教学、研究和服务的协同智能计算系统。

简单地说，自然人类智力 (HI) 主体及计算机人工智能 (AI) 代理如按照合理分工和高度协作的方式来处理知识信息数据，并产生优势互补和优化互动的效果，那这样的协作产物就可被称之为协同智能 (CI) 计算系统。这种关系的表达被抽象化 (如: $CI \geq HI + AI$) 就可被视为是一种抽象的协同智能观。本文的任务就是用可计算的方式来举例说明并具体解释这个抽象的协同智能观及其依据它构造的一个符合 $CI \geq HI + AI$ 要求的协同智能计算模型，即：计算图 1 所示的一系列虚拟的天平描述的 $\{id\} \cup \{ge\}$ 序位关系的两个基本方程及其最终可获得其解的信息恒等式。这里所谓一系列的虚拟的天平也就是理论上假定的 n 个虚拟的天平。换句话说，这就是建立在图灵机^[18]基础之上的孪生图灵机的可计算上限。

1.4 协同智能计算系统及其基本特点

本文对计算机软件设计与计算机辅助应用两个方面构成的协同智能计算系统的理解或逻辑上的基本思路可用几个关系式加以概括，它们是：

(协同智能计算系统) 软件 \geq 程序 + 界面

(协同智能计算系统) 程序 \geq 算法 + 数据结构

(协同智能计算系统) 界面 \geq 知识本体 + 对象

其中蕴含本研究的另一个研究切入点同时也是一个原创点即：人机交互界面，依托于：计算机数据处理与自然人知识处理两类信息处理方式的结合以及在其基础之上派生的协同智能观及其指导下的协同智能计算模型，涉及：信息学基础研究的一个突破。

其特点就在于“ \geq ”这个特定数学符号的应用。因为，如仅使用“ $=$ ” (AI) 或仅使用“ $>$ ” (HI) 与之 (CI) 都不同。

2 协同智能计算模型怎样发挥作用？

下面笔者以间接形式化的中文云端计算实例从四个方面来回答上面这个问题，即应用广义文本基因中字的类和例来阐述。

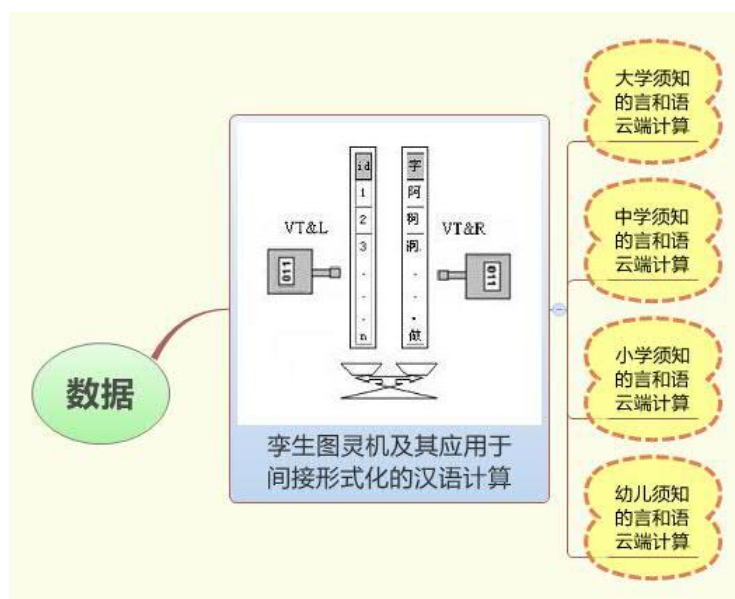


图 2 协同智能计算模型及其汉字汉语间接形式化计算实例示意图 [19]

由图 2 可见，直观天平上方的双列表恰似一对孪生图灵机，它本质上就是：孪生图灵机用于汉语形式化处理的实施例。因为，无论是**虚拟的左列表**（VT&L），还是**虚拟的右列表**（VT&R）都分别等价于两个单独的图灵机，所不同的就在于其受限的表格形态，即：在右列表内的单音节字即言的数量对计算机的记忆和计算等能力来说简直算不了什么。这就奠定了计算机辅助分析中文的言（单音节字）和语（双、多音节的字组）的应用理论基础。

细心的读者可能会发觉图 2 左端所示数据与右端所示知识之间似乎存在数据信息处理与知识信息处理上的某种不对称的情形。其实，这就是我们前面说到过的形式信息与内容信息这两方面的关系问题在技术层面的体现。从应用的角度来看，可把国内外学术前沿的各类期刊及会议论文摘要的计算机辅助分析，以及各种软件常用问题解答以及帮助文件的计算机辅助分析，乃至协同智能计算系统的用户个性化记录的计算机辅助分析，尤其是涉及自然语言的有限符号及其多样化组合或重复使用过程中蕴含的有限规则的计算机辅助分析，等等，做这样的处理，即：把它们均视为在各类活动中需要计算机辅助双语知识信息数据处理服务的客户将分别从该类云端计算服务系统中选择并有针对性地重复调用

相关形式信息以构成其各自独特组合或重组的内容信息的过程。第一类涉及学术研究层面主要在大学阶段来操作；第二类涉及不同程度的软件学习使用，可放在各个层面来操作，也就是说大中小学及幼儿园各阶段用户重复调用的数据量（形式信息）和知识量（内容信息）是不一样的；第三类主要涉及前述各类用户关注的具体内容信息，第四类涉及相应的形式信息，它们均可在融智学应用场景所述的创造性合作型生产式教研产学研用各类活动中得到最有条理且最有效的利用，尤其是对需要计算机辅助双语知识信息数据处理服务的客户而言，这一类云端计算实际上就是本文所述的协同智能计算系统的典型，其主要服务对象，是在创造性合作型生产式教研产学研用各类活动中的大中小学及幼儿园各阶段广大师生用户。善于思考的读者有可能会提出这样的问题：在图 2 中部的天平明明是对称的，可是在计算机擅长处理的数据与自然人擅长处理的知识这两个极端方面却为何偏偏是不对称的呢？

这个问题问得非常好。一语道破了“第三智能”（CI）同时具有对称与非对称的特点，以及可处理标准与非标的数据和知识的本领。这是“第一智能”（HI）和“第二智能”（AI）各自均欠缺的。进而，也就预示着基于“第三智能”（CI）的云端计算比主要依靠“第二智能”（AI）的云计算与主要依靠“第一智能”（HI）端计算乃至两者松散结合的云端计算具有更为优越的性能。

为什么会是这样呢？这还要从自然语言理解双重技术路线第二路径的发现这个角度来探究其成因。

2.1 揭示自然语言理解双重技术路线第二路径的奥妙

协同智能计算模型同时蕴含着语言学和信息学的基础研究成果所揭示的自然语言理解双重技术路线的第二路径的存在机理。

由图 3 可见，I 和 II 两条技术路线的区别与联系。路线 I 是：美国标准信息交换码（ASCII）^[21] 和程序语言（低中高）^[23] 均便于计算机底层适应于英文的信息处理；路线 II 是：在兼容于路线 I 且拓宽了计算机底层，本质上适应于各类语言或广义文本基因的信息处理新途径，其简捷性由图 3 下方的虚线所揭示，其本质特征是双语信息处理而非路线 I 一系列单语信息处理。为了详述第一和第二这两条形式化技术路线，不仅异曲同工、各有千秋，而且，后者比前者甚至还很有可能更胜一筹的实质。下面结合图 3，分别从（1）和（2）两个角度，来做

进一步说明:

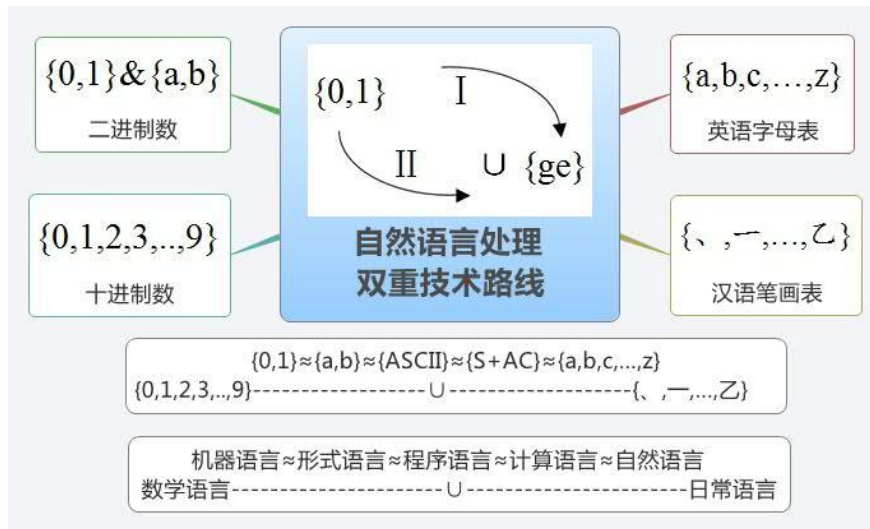


图 3 自然语言文本理解双重技术路线中第二路径的特征示意图 [20]

(1) 从拓广后的语言学视域看自然语言符号理解的双重路径

- I. 机器语言 \approx 形式语言 \approx 程序语言 \approx 计算语言 \approx 自然语言
- II. 数学语言 ----- U ----- 日常语言

以上第一路径的五种类型，从左到右，即：由低到高，可见其形式化程度，呈递减趋势，自然语言理解陷入了歧义泥潭；与之不同，第二路径的间接形式化的技术路线可走出歧义泥潭。

(2) 从拓广后的信息学视域看自然语言文本处理的双重路径

- I . $\{0,1\} \approx \{a,b\} \approx \{\text{ASCII}\} \approx \{S+AC\} \approx \{a,b,c,\dots,z\}$
- II . $\{0,1,2,3,\dots,9\} \cup \{、,一,\dots,乙\}$

其中，二进制数 $\{0,1\}$ 和十进制数 $\{0,1,2,3,\dots,9\}$ 及英语字母表 $\{a,b,c,\dots,z\}$ 和汉语字符笔画表 $\{、,一,\dots,乙\}$ 均可不做解释，即使是形式语言 [24] 基本符号 $\{a,b\}$ 和美国标准信息交换码 $\{\text{ASCII}\}$ 等符号集对计算机专业读者也不必说明或解释，但是，因本人依据乔姆斯基形式句法公式 ($S=NP+VP$) 以及语料库计算语言学^[25] 的标注集 (Annotated Corpora)，用首字母缩写相加用以指代或表示计算语言学符号集 $\{S+AC\}$ 却须在此说明，因为，这只是本人为描述或指代的方便而这样使用的，别无所指即无额外的意思。

联系图 1 的 $\{id\} \cup \{ge\}$ 示例和图 2 的 $\{id\} \cup \{字\}$ 用例中信息恒等式或天平

原理所蕴含的“同义并列、对应转换”法则，即：**第二信息基本定律假说**，再加上图 1 所示的**广义文本基因**所蕴含的“异义排列、序趣简美”法则，即：**第一信息基本定律假说**，可顺理成章地**解释**路线 II 的**简捷性、高效性**及其应用的**广普性**。

下面将分别结合从**计算机数据处理**和**自然人知识处理**来做进一步的阐述。

2.2 协同智能计算模型蕴含的数据信息处理方式

如果把一系列的左列表 {id} 以矩阵分布的形态展开分析，那么，我们就会发现它可相当方便地借用数学以及计算机科学技术的一系列现有成果来间接计算一系列右列表 {ge} 可记录且可**有针对性地**再现的**广义文本基因**及其衍生的**广义基因文本**。

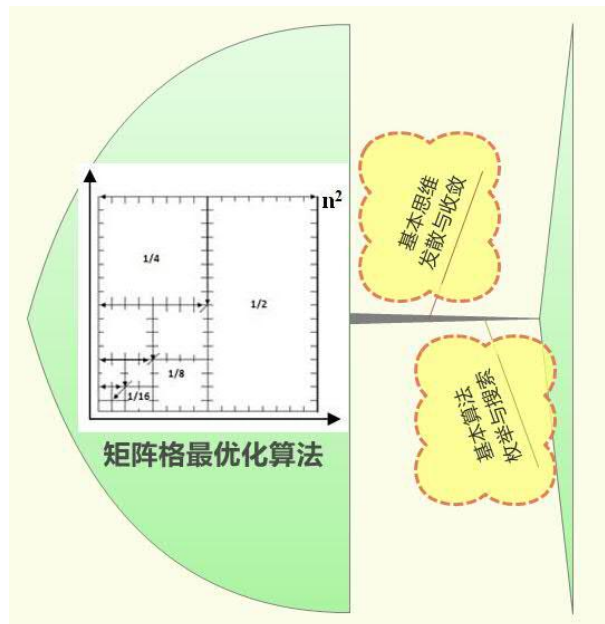


图 4 收敛与发散、搜索与穷举在 n^2 矩阵格中均可表达的示意图^[25]

由图 4 可见，在 n^2 个格子构成的方阵中，**搜索与枚举**的情形。每次搜索**总量减半**的区域**穷举式搜索**是一种**高效方法**。仅采用纯粹碰运气的**随机策略**，那是远远不够的，还需要用户不断地提供**反馈信息**，来引导：**收敛性搜索与发散性穷举**从而**减少盲目操作**。其它各种具体的搜索，如：**宽度优先、深度优先、启发式、遗传式**均可借助**反馈信息**的帮助在这个矩阵中进行。也就是说，

只要**设置的双列表**都遵循**同义并列对应转换**法则，而左、右两列表序位关系均满足**矩阵格最优化算法**所必须的数据结构，就能实现**间接计算**以及**间接**

形式化表达。左列表的标准化数据与右列表的个性化知识，均必须在间接计算模型和间接形式化方法结合的范式或格局之下，才能充分体现第二路径的一系列优势，即：既可**选取**更好的算法，又可**选取**最恰当的数据结构类型。

这和以往第一路径那种仅基于**小字符集串**的众多程序语言探寻的形式化途径的**最大不同**就在于：第二路径不仅可兼容于以往第一路径的各种程序语言，而且，还可做它们都做不到的事情——这就是可以让不懂或不喜欢它们那些程序语言的用户也能够**在一定的范围内随心所欲地编写自己的应用软件**，这就相当于：用户**仅仅用自己最为熟悉的母语和习惯的操作方式**，也就同样可像某些软件高手那样**编制针对性很强的应用软件**。通俗地讲就像做算术题或做泥塑乃至写字或绘画那样的方便。当然，其中，也有较复杂的设计，就像各行各业的专家那样做复杂的事。

2.3 协同智能计算模型蕴含的知识信息处理方式



图 5 以汉语或中文组字成语的特点为例的语言分析示意图 [26]

由图 5 可见，汉语**组字成语**的基本特点。结合图 2 左列表的 {id} 数与右列表的 {字} 即言（同时也是**基本语言结构单位**，即：**单音节形式**）描述**单音节字**的诸例是如何占据各个空格的。再强调一下，汉语的言，即**单音节字**，**语**即**双音节**，以及**多音节**的字组。英语的**词**即**混音节**，因此兼有言和语双重属性。换一句话说，英语只能区分语言和言语，而不能进一步区分言和语。已验证，无论是自然语言，还是形式语言，均已找到可区分的言和语的形式化依据。由此可见，汉语在**结构形式**

上有独特优势，即：汉语的言和语，均有明显的形式区分标记。

在图 5 下方不仅明确表示汉语的言是层面型结构（形字）与线串型结构（音字）而且还在图 5 中左方上部明确表示了言是组成辞的实字。这又是什么意思呢？这些其实主要是关于言在汉语实际应用中被人们赋予或可能赋予的含义以及用意。这是其形式标记以外被人们在使用中附加的意义。这也正是自然语言理解复杂的一个主要原因；相比之下前面提到汉语的形式特征，则是汉语在自然语言处理上一个有利条件。两者不能混淆。

笔者采用矩阵格最优化算法易于处理的方式排列汉语的字，就是因为其中搜索某个格子的过程正好就与从其中枚举所有格子的过程呈相反方向，即：在 n^2 矩阵范围内以等价于 2^n 的发散方式枚举乃至以等价于 $1/2^n$ 的收敛方式搜索，这两类基本算法在处理纯数字计算任务上的优势，其特征就在满足 n 的取值不影响计算效率。其间接计算任务，特指：言和语的间接计算。

2.4 协同智能计算模型蕴含的双语信息处理方式



图 6 以英汉汉英双语解释和双向机器翻译策略为例的语言处理示意图^[27]

由图 6 可见，双语信息处理方式与“解释 + 翻译”策略的联系。在图 6 中，两个序号①分别说明英汉双语各自的对象语言，即：词 (a_1) 与字 (b_1)，两个序号②分别说明英汉双语各自的解释语言：词组 (a_2) 与字组 1 (b_2)、句子 (a_3) 与字组 2 (b_3)。双方各自先完成从①到②即从对象语言到解释语言的转换之后，再进行从②到②即从一类解释语言到另一类解释语言的转换。其中，Start 标识说明双方各自语汇一级的两类解释语言之间的相互转换只是起步，打钩的方框标识说明双方各自语句一级的两类解释语言之间相互转换成功之后，才算完成了

翻译任务。

也就是说，在图 6 所示的“解释 + 翻译”策略模型中，通常是不考虑从①到①，即：从一类对象语言到另一类对象语言之间的双语直接转换，而是采取先完成从①到②，即：从一类对象语言到该类解释语言的解释转换后，再从②到②，即：从一类解释语言到另一类解释语言的翻译转换。其特征不在于：先进行从①到②即从一类对象语言的符号义项到该类解释语言的符号组合的同义并列对应转换，即在同义项的符号组合之间相互解释；再进行从②到②，即：从一类解释语言的具体符号组合到另一类解释语言的具体符号组合的同意并列对应转换，即：相当于彼此同意把约定俗成的双语解释相互转换或翻译。

在这两个步骤之中，第一步强调对语言基本结构单位（言）义项解释，第二步突出对不同层次的语言派生结构单位（语）相互翻译，两个步骤都建立“同意并列、对应转换”，即明确一系列“双语对”作为双语之间相互解释、相互翻译的基础。这就是“解释 + 翻译”策略之所以能够成为一个崭新的机器翻译策略在语言学基础理论上的支撑点。细心的读者会发现，新策略首先正视英汉两种语言在词与字之间通常不可直译的事实，即：正视 $a_1 \neq b_1$ 的情形，也不奢求 $a_1 \approx b_1$ 的奇迹。进而，寻求在先解释、后翻译的策略方面采用“解释 + 翻译”策略，即在 $a_2 \approx b_2$ 以及 $a_3 \approx b_3$ 均可成立的条件下，实施机器翻译的策略。

细心的读者会提出这样的问题：图 2 和图 6 描述的双语信息处理是一回事吗？显然不是。因为图 2 和图 6 描述的双语信息处理并非同类，前者是数学语言与日常语言的广义双语信息处理；后者是英语与汉语之间的狭义双语信息处理。笔者在此称前者为广义双语信息处理，称后者为狭义双语信息处理，且广义的双语信息处理以狭义的双语信息处理为基础。也就是说，“第二智能”（AI）以“第一智能”（HI）为基础，同理，“第三智能”（CI）也要以前两类智能为基础。

3 为什么需要协同智能计算模型？

要很好地回答这个问题，还要从广义的双语信息处理入手来揭示：协同智能计算模型的作用、意义、必要性及其可能性。

举例来说，假设做和信这两个字各有两个义项，如何处理？



图 7 以做和信两字为例揭示双语信息处理中取值与置信的关系示意图^[28]

由图 7 可见，广义和狭义的双语信息处理结合示例。最能体现协同智能计算模型的作用、意义、必要性和可能性。

首先，来看狭义的双语信息处理的情形，汉语的**做和信**这两个字均为实字，也就是说，它们均有可**取真值**的义项，这对普通人可能是一个很抽象的说法，但是，对词典编撰者而言，它就是一个术语，也就是我们通常所说的**词条**，在汉语中，也就用来体现该义项的实际用例，某些字组，如：做人，又如：做事；再如：信某人，又如：信某事。在汉语中这些都是很平常的用法。然而，一旦需要把它们翻译成外语时，比如：英语，就会遭遇这样的情形，即：同一个字，如：做或信，在英语中却要选用不同的词语才能把这个字的两个不同词条翻译到位，即：译**成较为地道的英语词语**，如：to be as sb. 即：做人，又如：to do sth. 即：做事；再如：(We) Trust sb. 即：(我们)信任某人，又如：(We) Believe sth. 即：(我们)相信某事。

当然，这种**狭义的双语信息处理或“解释 + 翻译”**的工作是需要熟悉汉英双语并具有相当背景知识的自然人才能做好的。

如何让本质上只懂二进制数 {0, 1} 的计算机也能**熟悉汉英双语**且也具有相当的背景知识呢？计算语言学家以及机器翻译工作者，经过半个多世纪努力**也没从根本上解决**这个国际难题。乐观的预计认为至少还需要几代人的努力才可能有较根本的突破【没想到 AI 大模型却基本上解决了其中主要的问题而且效果还不错！】。细心的读者**会联想到**图 3 第一路径描绘的情形，

即:

机器语言 \approx 形式语言 \approx 程序语言 \approx 计算语言 \approx 自然语言

这个链条看似在用一连串的单语来试图解决自然语言理解的难题，其实，则是在用一系列的不同种类、不同层次的语言来试图解决自然语言理解的难题。换句话说，表面简单（每个环节仅用单语）的情况背后却隐含了被忽略的深层复杂（整个链条却是多语）的实际困难。

为此，笔者试图另辟蹊径，图 3 第二路径描绘的情形，即：

数学语言 ----- U ----- 日常语言

如果能行【现在发现的确能行】，第二路径显然是比第一路径更为简捷的路径。

还是回到图 7 描述的广义和狭义的双语信息处理相结合的示例上来继续探讨前面的问题。就是说，一方面，接受图 7 左、右两边描述的广义和狭义的双语信息处理的范式，然后，让计算机来重复并代替自然人翻译已完成的上述工作，方式就是采取第二路径，即：在自然人完成的狭义的双语信息处理的基础上，继续做计算机可以更好参与的广义的双语信息处理；另一方面，进一步探讨图 7 中部描述的取值与置信相互影响的深层次的理论问题。如，把做与做人和做事的关系这类判断，称为主要是取值问题，虽然取值的结果也会影响再取值前的置信态度；又如：把信与信某人和信某事的关系这类判断，称为主要是置信问题，虽然置信的结果也会影响接下来的置信态度，从而，不仅将直接决定是否会继续取值的行为而且极有可能左右取值前后是否能正确取舍的问题。

由于涉及主观态度，而不仅仅只是理性判断或经验实证，因此，取值与置信的关系问题是一个高度复杂的问题，须留给自然人而且是受过复杂训练的跨学科专家来做。只有那些单纯的取值问题——也就是毋庸置疑的问题，才适合启用计算机来处理。至于许多连自然人甚至专家都难以置信的问题，如果不打算放弃，那么，最好就是采用协同智能计算系统来做。因为，就功能而言，协同智能计算系统 \geq 计算机+自然人。如果计算机擅长处理在 n^2 个格子构成的方阵中搜索与枚举，自然人，尤其是说汉语的人习惯于采用类似玩扑克牌的方式来处理言和语——就是重复选用已习惯的字与字组的各种排列组合，那么，协同智能计算系统就可基于计算机和自然人的双重优势来 兑现协同智能计算模型的基本设想：采取计算机与自然人之间合理分工并且高度协作的方式来

达到双方的优势互补和优化互动的效果。

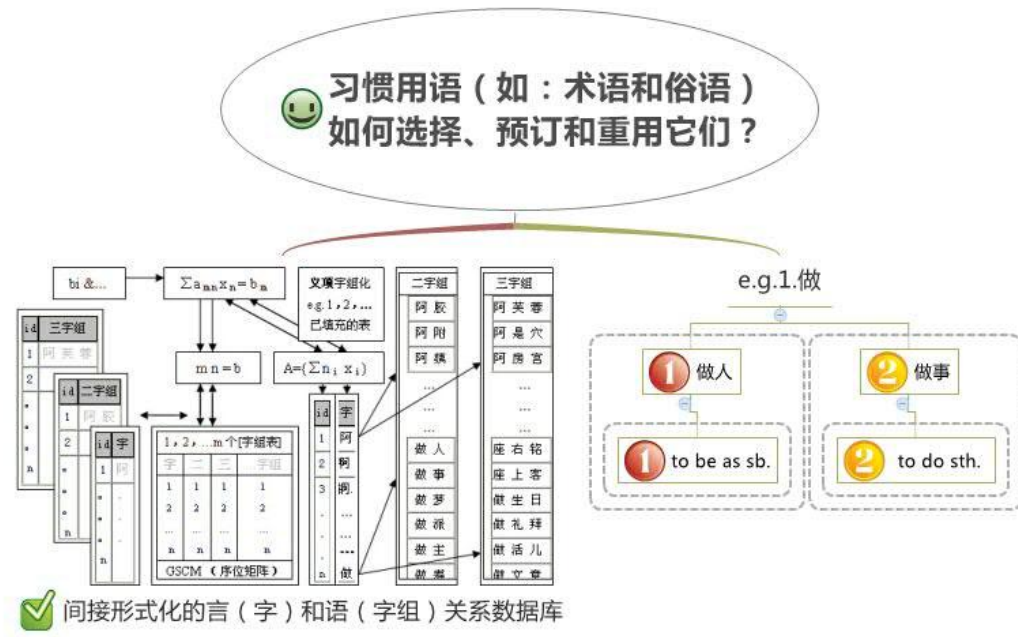


图 8 仅以做这个字为例来说明双语信息处理中如何取值计算的示意图^[29]

由图 8 可见, 经自然人专家完成了狭义的双语信息处理之后, 广义的双语信息处理就可交给计算机代理来做侍服处理, 例如: 图 2 所描述的为各类学校学生服务的云端计算数据查询服务器就可比普通人 (即: 各级学校的学生们) 都做得更好。

由此可见, 这样的结合才能体现协同智能计算的优点。

4 结 论

以上主要回答了三个问题:

1. 什么是协同智能计算系统的理论模型?
2. 它怎么通过序位关系计算而称量形式信息?
3. 它如何通过形式信息复用频率的统计而实现对内容信息的计量?

对此, 笔者给出一系列可验证且可重用的对策和方法, 如: 自然语言处理与理解的双重路径并行策略, 在 n^2 个格子构成的方阵中通过高效搜索与快速枚举而实现具体的言和语的间接形式化处理, 及其被重复使用情形的间接

计算——这主要是**机助人**的过程，可分别基于{笔画}和{音节}的计算或统计对言和语具体被重复使用的情形做计算机辅助分析——由于该过程可引入教育教学、科学研究、知识管理和数据查询等具体服务，因此，这又引入了**人助机**的过程，最后还可把上述**机助人**与**人助机**的**双向互动过程**分别以**形式信息方程**和**内容信息方程**的方式加以抽象提炼或概括，进而，获得理想的**协同智能计算模型**。这就排除了该装置左列表在设计上预留的一系列格子，所存在的**不确定性判定难题**（它是单纯的计算机所无法独立解决的）。

【人机互助新时代，凸显机助人与人助机的双向互动过程】

总而言之，上述**双重路径并行策略**的推行，可把**纯理论的置信假设问题**转换为**实践上的取值验证问题**，从而可开启**大小字符串**直接和间接的两种形式化计算范式结合^[30]的便利之门。

其具体应用至少可扩展到：国内外学术前沿的各类**期刊及会议论文摘要的计算机辅助分析**，各种软件的**常用问题解答**以及**帮助文件**的计算机辅助分析，**协同智能计算系统用户个性化记录**的计算机辅助分析，**自然语言有限符号及其多样化组合**或**重复使用过程中蕴含的有限规则**的计算机辅助分析，等等。 各类**双语信息**的计算机辅助分析该类**云端计算主要服务对象**是在**创造性合作型生产式教研产学研**各类活动中需要**计算机辅助双语知识信息数据处理服务**的客户即广大师生。

致 谢

感谢美国《科学》杂志执行总编（executive editor Monica M. Bradford）和高级编辑（Senior Editor Pamela J. Hines）在《Science》执行总编办公室耐心听取了笔者介绍自己**核心原创成果**的三个要点并给予了很好的建议！同时，还要感谢 UC Berkeley 提供笔者机会来介绍**汉语形式化数据库**即**字与字组的关系数据库设计方案**及其依据的“**言本位**”理论！ 有关单位涉及其信息学院（Technical Lead Howie X. Lan 在**认真**听取了笔者介绍自己**核心原创成果**之后还给予**三点书面评价**，并且认为该成果适合**放在云端**以供分享）、**东方语言系**（汉语中心张丽华**不仅两次认真**听取笔者介绍自己**核心原创成果**而且还**预订教室**并召集了老师们专门听取了笔者的详细报告和答疑）以及**高等教育研究中心主任**（C. Judson King **不仅认真**听取了笔者介绍的**原创思想**而且还**通过**

交谈举例表示了理解并给予了积极肯定)等。这些对本文的形成或尽快公开均具有积极的促进作用。

参考文献

- [1] 邹晓辉, 邹顺鹏. 间接计算模型与间接形式化方法 [J]. 软件
2011, 32 (5): p1-6.
- [2] ZOU Xiaohui. **Rational Standard Cooperating Intelligent Model** [J]. *Leading Science*, 2005,45 (6)
- [3] AAI. **Broad Discussions of Artificial Intelligence**.
[EB/OL][2011]<http://aaai.org/AITopics/AIOverview>
- [4] AAI. **Natural Language Understanding & Generating Text &Speech**.
[EB/OL][2011]<http://aaai.org/AITopics/NaturalLanguage>
- [5] ZOU Xiaohui, ZOU Shunpeng, WANG Ping . **Viewpoint and Method on Cooperating with Computer**[J]. *Leading Science*,2005,46 (7)
- [6] ZOU Xiaohui.**Formal Chinese and Zibenwei**[M]. Collected by XU Tongqiang, PAN Wenguo, YANG Zijian. *Study on Zibenwei Theory and Its Application*. Shan Dong Education Publisher,2008 .
- [7] ZOU Xiaohui.**Fundamental Research of Informatics**[M]. Collected by MA Aina, MIAO Dongsheng, JIANG Lu, YAN Xueshan . *Information Science: Interdisciplinary Research* . Zhe Jiang Education Publisher, 2 0 0 7.Edited by ZOU Xiaohui .*Science and Philosophy of Information* [J]. 2008(2)
[EB/OL]<http://survivor99.com/entropy/2008/20082Informatics.htm>
- [8] Zou Shunpeng; Zou Xiaohui; Liu Zhifang . **Focusing on Senior Talent Training and Profound Knowledge Management** [J]. *International Journal of Multidisciplinary Thought*.1 (1):356-369 (2010)
- [9] Zou Shunpeng; Zou Xiaohui; Liu Zhifang.**The Nature of College Education and College Management**[A].*Proceedings of 2010 Conference on Higher Education Development* [C]. YANG Deguang. London Science Press, 2010
- [10] ZOU Shunpeng.**The Theoretical Framework of Eight Disciplines**. AAAS

Student Poster Session.2011.2 .

- [11] The world's coolest brainstorming and mind mapping software and the best way to share ideas.XMind Ltd. [EB/ OL].[2011]
www.xmind.net/share/geneculture/indirect-formal/
- [12] ZOU Xiaohui, ZOU Shunpeng . Bad or Good for the Arithmetic of Information Formulas[J] . *Leading Science*, 2005,45 (6)
- [13]“**Intelligence information, often not differentiated from intelligence data**”[EB/OL]www.experts123.com/q/what-is-intelligence-information.html
- [14] Hartley, R.V.L .**Transmission of Information**[J].*Bell System Technical Journal*.Volume 7, Issue 3, July 1928
www.alcatel-lucent.com/bstj/vol07-1928/articles/bstj7-3-535.pdf
- [15] Shannon,C.E.**A Mathematical Theory of Communication** [J].*Bell System Technical Journal*. Volume 27, Issue 3, July - Issue 4, October 1948
www.alcatel-lucent.com/bstj/vol27-1948/articles/bstj27-3-379.pdf
www.alcatel-lucent.com/bstj/vol27-1948/articles/bstj27-4-623.pdf
- [16] ZOU Shunpeng, ZOU Xiaohui, LIU Zhifang. **The Overlapping Model of Education and Management**[C]. ISSN 2160-147X, MSE 2011
- [17] ZOU Xiaohui, ZOU Shunpeng. **Value-Taking and Confidence-building of Language**. [EB/OL] [2011]
www.paper.edu.cn/index.php/default/releasepaper/content/201108-218
- [18] A . M . Turing . **On Computable Numbers, with an application to the Entscheidungsproblem** [C], *Proceedings of the London Mathematical Society*. Second Series, 42, 230- 265, 1936.
- [19] The world's coolest brainstorming and mind mapping software and the best way to share ideas. XMind Ltd. [EB/ OL].[2011]
www.xmind.net/share/geneculture/data/
- [20] The world's coolest brainstorming and mind mapping software and the best way to share ideas. XMind Ltd. [EB/ OL].[2011]

www.xmind.net/share/geneculture/natural-language-processing-dual-technique

[21] American Standard Code for Information Interchange.1963 [EB/OL]

<http://wps.com/projects/codes/X3.4-1963/index.html>

<http://www.unicode.org/charts/PDF/U0000.pdf>

[22] History of Programming Language Conference (HOPL) .2007 [EB/OL]

<http://research.ihost.com/hopl/HOPL.html>

<http://research.ihost.com/hopl/HOPL-III.html>

<http://hopl.murdoch.edu.au/>

[23] James Power . Notes on Formal Language Theory and Parsing. 2002

[EB/OL] http://en.wikipedia.org/wiki/Formal_language

http://www.cs.umbc.edu/portal/help/theory/lang_def.shtml

[24] Douglas Biber, etc. **Corpus linguistics: investigating language structure and use**. Cambridge University Press,2004

[25] The world's coolest brainstorming and mind mapping software and the best way to share ideas. XMind Ltd. [EB/ OL].[2011]

www.xmind.net/share/geneculture/matrix-optimization-algorithms

[26] The world's coolest brainstorming and mind mapping software and the best way to share ideas. XMind Ltd. [EB/ OL].[2011]

www.xmind.net/share/geneculture/the-basic-model-of-linguistic-analysis-2

[27] The world's coolest brainstorming and mind mapping software and the best way to share ideas. XMind Ltd. [EB/ OL].[2011]

www.xmind.net/share/geneculture/a-brand-new-machine-translation-strategy

[28] The world's coolest brainstorming and mind mapping software and the best way to share ideas. XMind Ltd. [EB/ OL].[2011]

www.xmind.net/share/geneculture/principles-of-balance-and-cases

[29] The world's coolest brainstorming and mind mapping software and the best way to share ideas. XMind Ltd. [EB/ OL].[2011]

[www.xmind.net/share/geneculture/idioms-\(such-as-](http://www.xmind.net/share/geneculture/idioms-(such-as-)

terminology-and-colloquialisms)-how-to-selec/

- [30] ZOU Xioahui, ZOU Shunpeng. **The Indirect Formal Methods Based on the Bilingual Automatic Conversion**. 2010 Application, 2011 Public. (*Patent: 2010101752962*)