

支持创新型组织学习:基于任务情景的知识适用性管理

高 济

(浙江大学人工智能研究所 杭州 310027)

摘 要 面临商务环境的快速变迁和难以预测,创新型组织学习正在成为组织提高生存、适应和竞争能力的关键途径。然而,现有的 KM 技术对于如何支持创新型组织学习,没有进行深入、系统的研究;导致知识工作者在获取适用性好的信息体(从而促进业务创新)和主动参与 OM 开放性进化(以促进创新知识的传播和共享)方面缺乏有效的支持。文中提出基于任务情景的知识适用性管理方法 KAMTC,旨在以基于应用域本体的任务情景描述模式作为统一的语义基础,去确切、全面和便捷地描述业务工作的查询需求和受检信息体的适用性,以克服这些挑战性问题,进而有效支持知识创新和创新型组织学习的展开。

关键词 知识管理;组织学习;知识创新;适用性管理;任务情景;开放性进化

中图法分类号 TP311

Supporting Creative Organizational Learning: Knowledge Applicability Management Based on Task Circumstance

GAO Ji

(Institute of Artificial Intelligence, Zhejiang University, Hangzhou 310027)

Abstract In face of the business environments rapidly changing and difficulty to predict, creative organizational learning becomes the key approach for organizations to enhance ability of survival, adaptability, and competition. However, the present technology of KM (Knowledge Management) has not made the deep and systematic research work for supporting the creative learning in organizations, so that knowledge workers lack effective supports to capture applicability-good information bodies (hence able to advance business innovation) and to actively participate in the open evolution of OM (Organizational Memory) which can facilitate the communication and sharing of creative knowledge. This paper proposes a methodology, called KAMTC (Knowledge Applicability Management based on Task Circumstance), which aims at making domain ontology based circumstance description patterns of tasks as the uniform semantic basis to describe the query requirements for business work and the applicability of searched information bodies precisely, completely, and handily, in order to overcome those challenging problems and hence to support effectively knowledge creation and creative organizational learning.

Keywords knowledge management; organizational learning; knowledge creation; applicability management; task circumstance; open evolution

1 引 言

在相对稳定和可预测的传统商务环境下,组织

学习的主要模式就是让员工,尤其是知识工作者(运用密集知识的脑力劳动者)共享和重用组织积累的、使业务得以优化和高效运作的知识。为支持这种学习模式,KM(知识管理)系统都聚焦于应用信息处理

技术去“保存和传播这些可以长期不变的业务知识,并于适当的时间递送适用的知识给适当的人”^[1-2]。

随着全球经济一体化进程的加速和市场竞争的日趋加剧以及关于产品和服务的需求和技术的快速更新,商务环境越趋动态变化且难以预测。组织亟需变革学习模式,依赖知识创新——知识的按需创建和原有知识的及时更新,而非可能已过时知识的共享和重用,去应对商务环境的跳跃式甚至根本性的变化,并由此提高生存、适应和竞争能力^[3-6]。

尽管已存在支持知识自动创新的信息技术(如机器学习),但在可预见的未来,组织业务知识的创建和更新仍然主要依赖知识工作者的个人能力和精诚协作,信息处理技术难以取代。然而,现有的 KM 技术对于如何支持创新型组织学习,没有进行深入、系统的研究,导致知识工作者在获取适用性好的信息体(从而促进业务创新)和主动参与 OM 开放性进化(以促进创新知识的传播和共享)方面缺乏有效的支持,亟需革新。

本文提出基于任务情景的知识适用性管理方法 KAMTC,旨在以基于应用域共享本体的任务情景描述模式作为统一的语义基础,去确切、全面和便捷地描述业务工作的查询需求和受检信息体的适用性,以有效克服目前 KM 系统面临的这些挑战性问题,进而支持知识创新和创新型组织学习的展开。

2 创新型组织学习和知识适用性

首先引入知识分类和转变的概念。按照 Polanyi 的观点,知识可以划分为两大类: tacit(含蓄的)知识、explicit(明晰的)知识^[7]。前者含蓄于个人的内心世界,来自个人经验的积累,并体现个人的信念和价值观;尽管未表述(也往往难以明晰表述)出来,但最有价值——可直接导致个人的智能行为,从而成为个人和组织业务创新的基础。后者则通过某种人工物品(如文档、甚至视频)来加以明晰表示,建立的主要目的是促进个人间的知识传播和共享。在此基础上,Nonaka 制定了组织学习理论,聚焦于含蓄和明晰知识间的 4 类转变过程^[8]:

(1) Socialization (tacit to tacit)——社交化过程旨在通过个人间含蓄知识的传播来实现知识共享,并与社区及协作等理念紧密关联。

(2) Externalization (tacit to explicit)——外表化过程旨在通过概念化、抽取、表示等步骤,将个人拥有的含蓄知识转变为可供其他人取用的明晰知识。

(3) Combination (explicit to explicit)——化合

过程旨在通过对于明晰知识的存储、索引、分类、搜索和关联等手段,促进知识的共享和重用。

(4) Internalization (explicit to tacit)——内在化过程旨在通过理解和消化外来的明晰知识去促进个人创新含蓄知识的形成。鉴于知识创新的机遇取决于个人已有含蓄知识和外来知识的融合,获取适用的明晰知识至关重要。

显然,明晰知识的内在化是个人含蓄知识创新(进而业务创新)的动力和源泉,含蓄知识的外表化、外表化后明晰知识的化合以及含蓄知识的社交化,则可促进这些知识的传播、共享和重用(图 1)。这些转变过程的螺旋形循环渐进,再加上鼓励知识工作者间的精诚协作,使组织能够快速创建出足够的新知识去提升业务智能,以应对已发生根本性变化的商务环境。

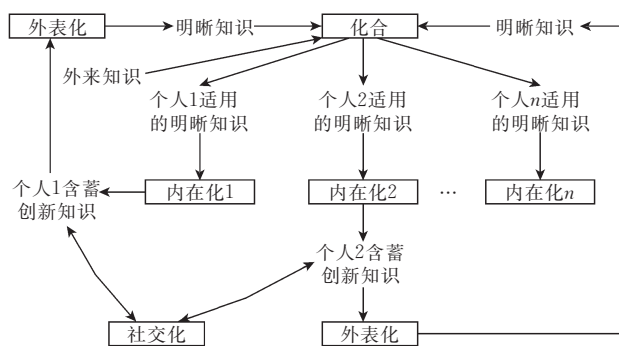


图 1 创新型组织学习

可以将明晰知识的内在化划分为两种类型:基础学习和问题求解。前者通过培训和自学方式,使员工培养起运作业务必须具备的基础性含蓄知识和基于这些知识的业务操作技能。后者则通过学习适用的明晰知识,融合或更新过去积累的含蓄知识,使员工能够高效创建解决当前问题所需的新含蓄知识。尽管含蓄知识的社交化也是含蓄知识创新的重要方式,但超出了本文的研究范围,将不予讨论。

本文聚焦于问题求解方式,并设想知识工作者已具备熟练开展业务所需的含蓄知识。尽管明晰知识的内在化过程中个人思维活动的细节不受外界控制,但适用性好(匹配问题求解需求)的明晰知识,有助于催化创新灵感的产生。这些知识中包含的有助于解决当前问题的概念、方法和技术,往往成为创新思维的源泉。

尽管获取适用明晰知识的技术属于已得到密集研究和开发的化合模式,但如何提高查询结果的适用精度仍然缺乏有效的办法。面临的困难可归纳为以下两个方面:

(1) 难以引导用户确切、全面和便捷地表述查

询需求。

其实, 提出查询需求也是含蓄知识的一种外表化方式, 只是这里的含蓄知识面向当前业务工作存在的问题, 而非解决问题的经验。显然, 人对于面临的问题往往只有模糊的认识, 这给确切、全面地表述查询需求带来困难。

(2) 难以确切、全面地表述受检信息体的适用性, 原因在于受检信息体的适用性评价存在固有模糊性:

① 受检信息体往往包含面向不同目标的、解决不同问题(任务)的知识。只有包含的知识有助于解决当前业务工作面临的问题(任务), 知识工作者才认为该信息体适用; 否则, 即便对别人很适用的信息体, 不包含解决当前问题的知识, 仍然是不适用的。

② 绝大多数电子信息体都以自然语言形式表述, 仅供人通过阅读时的理解(内在化过程)去领悟(创建和更新自己的含蓄知识)。人们知识背景(已积累的含蓄知识)的差异将导致领悟程度和适用性评价的迥异。

③ 目前, 指示信息体适用性的主流方法是建立基于关键词或词组的索引^[9-10], 但描述结构的缺乏、同义词和近义词的应用以及一词多义和语义隐含(内容隐含但未显式使用某关键词或词组)现象的大量存在, 使得信息体的适用性难以确切、全面地表述。

含蓄知识的外表化也是一件困难的任务。首先应做概念化工作, 使外表化后产生的知识能使潜在的使用者理解的概念、属性和关系明晰表示; 然后依据产生的概念化模型从含蓄知识抽取出可以明晰表示的知识; 最后以适当的语言和形式表示这些知识于信息体。显然, 外表化过程是人工密集型的, 尽管有信息技术工具作辅助。另外, 给信息体建立适用性描述作为索引, 以便 OM 俘获(使它们能被查询), 也是含蓄知识的一种外表化方式。

现今 KM 系统聚焦于化合模式的开发, 已在明晰知识的存储、索引、分类、搜索和关联等方面卓有成效; 支持社交化模式的信息技术也得到了开发, 包括 text-based chat, electronic meetings, bulletin boards, expertise location 等。尽管这些技术已涉及含蓄知识外表化后的存储和展示, 并为理解和分析这些知识提供辅助, 但尚缺乏有效的手段去支持明晰知识的内在化和含蓄知识的外表化。

如上所述, 尽管知识的内在化和外表化转变过程主要依赖个人能力, 但仍然可以通过提供适用性好(匹配业务工作需求)的明晰知识去促进个人含蓄知识的创新, 并通过即时建立信息体适用性索引去

加速创新知识的俘获和传播。为此我们指出, 革新 KM 系统去支持创新型组织学习的关键应该是实现有效的知识适用性管理, 而实现的前提则在于建立统一的语义基础, 去确切、全面和便捷地描述业务工作的查询需求和受检信息体的适用性。

3 任务情景描述模式的引入

从促进问题求解的角度, 我们限定 OM 俘获的信息体都包含解决业务工作中 1—多个任务(问题)的知识, 当前业务工作则面临 1—多个需解决的任务(问题)。以 E 指示当前业务工作或受检信息体, 并令这些任务分别属于 1—多个应用目标, 则 E 需求或包含的知识 $k(E)$ 为

$$k(E) = k(g_1) \cup k(g_2) \cup \cdots \cup k(g_n), \text{ 且}$$

$$k(g_h) = k(t_{h1}) \cup k(t_{h2}) \cup \cdots \cup k(t_{hm}),$$

其中 $g_h (h = 1, 2, \cdots, n)$ 为 E 面向的应用目标, $t_{hk} (k = 1, 2, \cdots, m)$ 为目标 g_h 涉及的任务。

从知识重用的角度, 可以将关于每个任务 t 的知识($k(t)$)划分为两个部分:

$$k(t) = k_c(t) \cup k_s(t),$$

$k_c(t)$ 指示任务情景; $k_s(t)$ 指示任务解决。 $k_c(t)$ 描述任务自身及其处境的特征。例如某信息体包含给电子文档作标注的知识, 标注任务自身的特征包括自动标注类型、并提供标注建立和浏览操作; 该任务处境的特征包括: 标注的内容是一些注释、标注的目标载体是 Web 页面、标注语言为 XML、标注技术依赖本体工程等。 $k_s(t)$ 则描述标注任务的实现方法和相关细节。

设想知识工作者旨在寻求当前业务工作 p 中关于任务 t 如何解决的知识 $k_s(t)$, 例如上述的标注任务如何实现, 但只有依赖关于目标(任务)分类和任务情景的知识作为适用性索引, 才易于搜索到。显然, 对于受检信息体 q , 只有存在某任务 s , 使得任务情景 $k_c(s)$ 与 $k_c(t)$ 相似, q 才可能有参考价值。建立目标(任务)分类体系 G , 并让抽象任务 T, S 分别隶属于底层子目标 g_x, g_y , 具体任务 t, s 分别作为 T, S 的示例, 即 $t \in T, s \in S$, 则有助于快捷确定当前业务工作 p 或受检信息体 q 面临哪些任务。

如此, 以目标-子目标-任务为线索, 只要能确切和全面地描述当前业务工作和受检信息体的任务情景, 并将前者作为查询需求(对解决当前任务所需知识的表述), 就可通过任务情景的匹配比较来精确地评价信息体的适用性。这种方式的价值又依赖于人们对业务处理的两个信念:

(1) 业务的规律性. 业务中情景类似的任务可以用相似的方式解决;

(2) 业务的重现性. 情景类似的任务往往会在将来的业务中重现.

尽管易于发生跳跃式甚至根本性的变化是商务环境的新特点,然而业务的规律性和重现性仍然存在.变化主要体现在新抽象任务(问题)的出现,或原有抽象任务的情景描述模式发生了变化;只要能够及时增加面向新抽象任务的情景描述模式或更新原有抽象任务的情景描述模式,使得 OM 包容面向这些抽象任务的新增信息体,并支持相应的查询需求,就可实现 OM 的开放性进化,以适应商务环境的快速变迁.

现在,如何确切、全面、便捷地描述任务情景,并能支持受检信息体任务情景的自动生成和确保适用性查询的精度,成为建立新型 KM 系统去推动创新型组织学习的关键.

如同图像可以从多个侧面(槽)来描述其特征那样,任务情景也需从多个侧面(槽)来加以描述,每个侧面指示任务自身或其处境的一个特征;若能给每个特征槽指派一个完备的可选填充值集合,就可使受检信息体和当前业务工作的任务情景具有统一的语义描述基础和强大的结构化语义表示手段.可以指派应用域术语(关键词或词组)作为特征槽的可选填充值,这些术语构成指派给相应特征槽的术语可选集.

我们定义抽象任务 T 的情景描述模式由某些特征槽及其术语可选集构成,任务示例 $t \in T$ 的情景即为该模式的示例,特征槽名和选取的槽值联合起来形成对任务情景 $k_c(t)$ 某个侧面的索引.任务情景描述模式和上述目标(任务)分类体系构成应用域本体论的核心部分.

定义 1(任务情景描述模式). 定义抽象任务 T 的情景描述模式 CDP^T 为特征槽 fs_i^T 的集合: $CDP^T = \{fs_1^T, fs_2^T, \dots, fs_n^T\}$, $fs_i^T = \langle fsn_i^T, tcs_i^T \rangle$;该序偶的前项 fsn_i^T 指示特征槽名,必须在 CDP^T 中唯一;后项 tcs_i^T 则指示提供槽值的术语可选集: $tcs_i^T = \{term_{i1}^T, term_{i2}^T, \dots, term_{im}^T\}$.

定义 2(任务情景). 任务情景 $k_c(t)$ 是对出现于受检信息体或当前业务工作 E 的任务示例 $t \in T$ 的情景描述, $k_c(t) = \{fs_1^t, fs_2^t, \dots, fs_n^t\}$, $fs_i^t = \langle fsn_i^T, tcs_i^t \rangle$, $tcs_i^t \subseteq tcs_i^T$.

理论上,抽象任务 T 的情景描述模式 CDP^T 应有完备的特征槽集合和对应每个特征槽的、完备的术语可选集,使其可以覆盖任务示例的所有可能的

情景描述;但由于上述受检信息体适用性评价的固有模糊性以及商务环境的变迁,绝对的完备性不可能达到.然而,只要应用域充分狭窄(如某企业的自身业务或某个细分行业),易于达成知识工作者的赞同一致,建立相对完备的任务情景描述模式是可能的;即使初始的任务情景描述模式不够完备,也可通过开放性进化,逐步达到相对完备.例如,上述对电子文档作标注的抽象任务,可以制定其初始的情景描述模式,如图 2 所示.对于目标分类体系和任务集的定义也可作类似处理.

Document Annotation
Type: Manual Annotation, Semi-automation Annotation, Automatic Annotation, Web Annotation
Content: Advice, Change, Comment, Example, Explanation, Question, See-also
Objective: Document, Web Page, Email, Video,
Language: Html, XML, RDF, Ontology Language, Annotation Language, Mark-up, Language
Technology: Ontology Engineering, Concept, Relationship, Lexical Analysis, Crawler, Data Modeling, Metadata
Operation: Creation, Browsing, Filtering

图 2 “文档标注”任务的情景描述模式

所以,一旦建立起应用域共享本体论,包括目标(任务)分类体系 G 和所有抽象任务的情景描述模式,受检信息体 q 的适用性评价就可转变为 q 涉及任务 s 的情景 $k_c(s)$ 与当前业务工作 p 关注任务 t 的情景 $k_c(t)$ 是否相容匹配的问题.可以制定相容匹配的操作化定义和匹配程度的计算公式,作为信息体适用性评价的量化标准.

4 知识适用性管理方法 KAMTC

以应用域共享本体论提供的任务情景描述模式和目标(任务)分类体系作为含蓄和明晰知识间转变过程的统一语义基础,KAMTC 方法从 3 个方面实现了知识适用性管理:当前业务工作的任务情景抽取、受检信息体任务情景的即时建立和基于任务情景的相容匹配.

4.1 当前业务工作的任务情景抽取

任务情景的抽取将对于当前业务工作(问题)的认识外表化,分 3 个步骤:

(1) 以应用域共享本体论提供的任务情景描述模式和目标(任务)分类体系作为抽取当前业务工作情景的概念模型.

(2) 引导知识工作者抽取任务及其情景描述,包括通过可视化目标(任务)分类体系去指导知识工作者便捷地确定面临的目标和任务,通过任务情景描述模式的图表化界面去引导知识工作者确切、全

面和便捷地外表面化任务的情景描述。

(3) KM 系统将任务的情景描述整理成规范的表现形式,作为查询需求。

任务的情景描述不一定会一次到位,尤其在知识工作者遭遇不熟悉的任务时,可以先建立不够确切的任务情景,获取匹配该任务情景的信息体;然后通过阅读和理解加强对当前业务工作的认识,再更新任务的情景描述做新一轮查询,以获取更适用的信息体。如此,可以形成一个对于业务工作的螺旋形认识提升过程,去推动知识的创新。

4.2 受检信息体任务情景的即时建立

组织应鼓励知识工作者将解决当前业务工作问题而建立的创新型含蓄知识外表面化。然后,给外表面化后建立的信息体即时抽取任务情景作为适用性索引,使其为 OM 俘获:

(1) 以应用域共享本体论提供的目标(任务)分类体系指导信息体作者确定面临的目标和任务。

(2) 针对这些任务的情景描述模式中各特征槽的术语可选集,搜索可选集每个术语及其同义词在该信息体中的出现频度,并由此自动建立相应于这些任务的情景描述。

(3) 信息体作者通过 KM 系统提供的图表化界面审查这些任务的情景描述,并在必要时人工增强任务情景的描述。

KM 系统分 4 个级别支持任务情景描述的增强:

(1) 信息体内容隐含了某术语可选集中的某术语,但未显式引用该术语,则将该术语加进任务情景的相应特征槽。

(2) 对于某特征槽,其术语可选集不足以描述该特征,则作者可以增加全新术语进该术语可选集。

(3) 特征槽的集合不足以描述任务情景,则作者可以增设新槽和该槽的术语可选集。

(4) 目标(任务)分类体系未覆盖信息体面向的问题求解任务,则作者可以设置新的任务(甚至目标)及其情景描述模式。

后 3 个级别更适合商务环境快速变迁甚至突变的情况。不过, KM 系统仅将这些增强操作视为对应用域共享本体论的试探性扩充,提交到扩充建议库,不会直接更新本体论。依据积累的扩充建议,应用域专家定期作审议,并据此扩充本体论,以支持本体论的开放性进化。

4.3 基于任务情景的相容匹配

由于受检信息体的适用性索引(描述)和当前业务工作的查询需求都以任务情景描述模式作为公共语义基础,可以应用基于任务情景的相容匹配去大幅度提高查询的准确度,使搜索引擎的注意力快捷

地聚焦到有参考价值的信息体。搜索机制 CMSM 依赖特征槽的相容匹配来决定任务情景的相似性,可表示为以下 4 元组:

$$CMSM = (SRange, CMM, TCBW, IBSE),$$

其中,

$SRange$ 为信息体搜索范围,其是 OM 包含的信息体全集 $IBody$ 的子集;

CMM 为受检信息体与业务工作的相容匹配策略;

$TCBW$ 为当前业务工作的任务情景,包含多个任务示例的情景描述;

$IBSE$ 为信息体搜索引擎。

$SRange$ 指定 $IBSE$ 的工作范围,大到整个 OM,小到仅限于涉及某个抽象任务 T 的信息体集合 $IBody^T \subset IBody$ 。 $CMM \in \{and, or, part-and, part-or\}$ 。 and 策略最严格:仅当抽象任务(即任务情景描述模式)相同,且所有对应的特征槽都相容匹配时,受检信息体 $q (\in IBody^T)$ 才与 p 相容匹配。 or 策略最宽松:不要求抽象任务(即任务情景描述模式)相同,且只需有一个对应的特征槽相容匹配, q 就与 p 相容匹配。 $part-and$ 策略处于前二种方式之间:允许知识工作者指定 1-多个必须相容匹配的特征槽,其它特征槽不必匹配;在用 and 策略搜索不到严格相容的信息体,而用 or 策略又会搜索到太多宽松相容的信息体场合下,这种策略最适合。 $part-or$ 策略将 or 策略局限于指定的几个特征槽; and 和 $part-and$ 策略只适用于 $SRange$ 为 $IBody^T$ 的场合。缺省情况下, CMM 取 and 策略, $SRange$ 取 $IBody^T$ 。

$IBSE$ 按照指定的 $SRange$ 和 CMM , 搜索任务情景与业务工作相容匹配的信息体。相容匹配的实现依赖于应用域基础本体论 O_{ib}^a 的建立(参见 5.2 节)。令 a, b 指示术语,且 $a, b \in O_{ib}^a$, 下面定义几个面向相容匹配的概念;这些概念与 O_{ib}^a 提供的术语分类体系和同义词一起,构成 $IBSE$ 实现相容匹配的操作化依据。

定义 3(术语的语义包容). a 语义包容 b , 若在相应于 O_{ib}^a 的术语分类体系中, 对应于 a 的节点是对应于 b 节点的祖先, 即 a 在 b 通往根节点的路径上, 记为 $b < a$ 。

定义 4(术语强相容). a 和 b 强相容, 若 a 和 b 相同、同义或 $b < a$ 。

定义 5(术语弱相容). a 和 b 弱相容, 若 $a < b$ 。

定义 6(术语间语义距离). b 到 a 的语义距离是 κ , 若 $a < b$, 且 b 是 a 通往根节点的路径上遇到的第 κ 个节点。

对于业务工作 p 关注的任务示例 $t \in T$, 记情景描述 $k_c(t)$ 的第 i 个特征槽为 fs_i^t , 槽名为 fsn_i^T , 槽内容为 tcs_i^t (参见定义 2); 对于 OM 中受检信息体 q 描述的某个任务示例 $s \in S$, 记情景描述 $k_c(s)$ 的第 j 个特征槽为 fs_j^s , 槽名为 fsn_j^S , 槽内容为 tcs_j^s , 定义以下概念.

定义 7 (特征槽名相容). 特征槽名 fsn_i^T 和 fsn_j^S 相容, 若 fsn_i^T 和 fsn_j^S 强相容或弱相容.

定义 8 (特征槽内容强相容). 特征槽内容 tcs_i^t 与 tcs_j^s 强相容, 若存在术语 $x_t \in tcs_i^t$ 和 $x_s \in tcs_j^s$, 使得 x_t 和 x_s 强相容 (即 $x_s < x_t$).

定义 9 (特征槽内容弱相容). 特征槽内容 tcs_i^t 与 tcs_j^s κ 级弱相容, 若 tcs_i^t 与 tcs_j^s 非强相容, 但存在术语 $x_t \in tcs_i^t$ 和 $x_s \in tcs_j^s$, 使得 x_t 和 x_s 弱相容 (即 $x_t < x_s$), 且 x_s 到 x_t 的语义距离 κ 是最小的 (可能存在多对弱相容术语).

定义 10 (特征槽强相容). 特征槽 fs_i^t 与 fs_j^s 强相容, 若槽名 fsn_i^T 和 fsn_j^S 相容, 且槽内容 tcs_i^t 与 tcs_j^s 强相容.

定义 11 (特征槽弱相容). 特征槽 fs_i^t 与 fs_j^s κ 级弱相容, 若槽名 fsn_i^T 和 fsn_j^S 相容, 且槽内容 tcs_i^t 与 tcs_j^s κ 级弱相容.

定义 12 (特征槽不相容). 特征槽 fs_i^t 与 fs_j^s 不相容, 若 fs_i^t 与 fs_j^s 既非强相容, 也非 κ 级弱相容.

鉴于特征槽的相容性取决于槽内容, 槽名仅用于区分任务情景描述的不同方面, 所以槽名的相容不必区分强、弱. 特征槽的强相容意味着 t 的特征描述在语义上包容 s , 可认为完全匹配知识工作者的需求; 而弱相容却说明 s 的特征描述仅在语义上部分匹配知识工作者的需求. 为量化特征槽相容匹配的程度, 引入参数 $v \in [0, 1]$, 指示 fs_i^t 与 fs_j^s 的相容性:

$v=1$, 若 fs_i^t 与 fs_j^s 强相容; $v=(1/2)^\kappa$, 若 fs_i^t 与 fs_j^s κ 级弱相容; $v=0$, 若 fs_i^t 与 fs_j^s 不相容.

CMSM 主要适用于当前业务工作 p 与受检信息体 q 关注同一抽象任务的情况, 即 $T=S$, 尤其适用于同一抽象任务积累了大量示例的企事业. 对于这种情况, 情景描述模式 $CDP^T \equiv CDP^S$, $Srange$ 就是 $IBody^T$, 槽的相容性只取决于槽内容的相容性. 设 CDP^T 有 h 个特征槽, 且令任务情景 $k_c(t)$ 和 $k_c(s)$ 的第 i 对槽的相容性为 v_i , CMSM 制定任务情景的相容匹配评价函数 F 如下:

$$F = \sum \omega_i \cdot v_i, v_i \in [0, 1], \omega_i \in [0, 1], \\ i = 1, 2, \dots, \gamma,$$

其中, ω_i 指示第 i 对槽在相容匹配检查中的重要性加权. CMM 取 and、or 或 part-and 策略时 $\gamma=h$, 取

part-or 策略时, $\gamma < h$ (γ 为知识工作者指定槽的个数). CMSM 将按 F 值 ($\in [0, 1]$) 从大到小排序相容匹配的信息体. 对于 and 或 part-and 策略, 所有槽或指定槽的相容性必须满足条件 $v > 0$. 显然, F 值最大的信息体具有最好的适用性.

CMSM 也用于当前业务工作 p 与受检信息体 q 包含的抽象任务不同的情况, 但 CMM 只能取 or 或 part-or 策略. 考虑到不同抽象任务的示例, 甚至不同目标的任务示例或许有某些值得借鉴之处, 可以在某目标或整个应用域范围内搜索有单一特征槽相容 (强、弱相容) 的信息体. 鉴于目标名、任务名、槽名和槽值 (术语列表) 都来自同一 O_{ib} , 评价函数 F 照样适用.

显然, and 和 part-and 策略适合于涉及同一抽象任务的信息体数量较大的 OM, 严格的相容匹配使得只有少量信息体能够满足查询需求, 并按 F 值排序适用性; or 和 part-or 策略则适合于满足查询需求的信息体很少的情况, 使得相容匹配程度较差 (但仍有参考价值) 的信息体有机会被检索到. 对于因商务环境突变而面临的新任务 (问题), 后者特别有用; 或许解决其它任务的知识能够激发处理新任务的灵感. 可以说, 正是引入了基于任务情景的适用性索引和相容匹配, IBSE 才有效避免了基于关键词或词组的搜索引擎 (如 google) 往往会出现的、要么返回成百上千个适用性不好的结果或者干脆无结果返回的困境.

4.4 支持创新型组织学习

KAMTC 方法抓住了支持创新型组织学习的 3 个关键环节: 抽取查询需求、获取适用知识和俘获创新知识 (图 3).

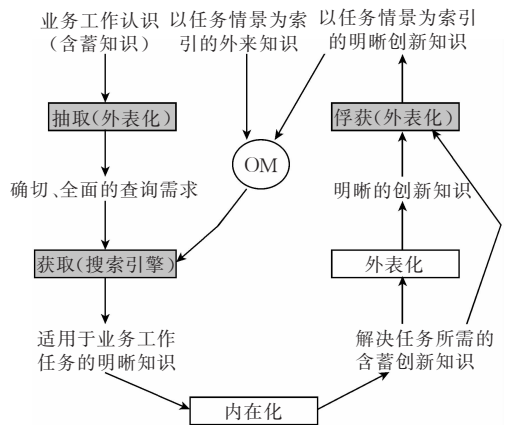


图 3 KAMTC 方法对创新型组织学习的支持

首先, 在“抽取查询需求”环节, 引导知识工作者便捷地抽取当前业务工作面临的创新任务 (如解决某个新颖问题) 及其情景描述, 作为查询需求. 然后

进入“获取适用知识”环节，假设组织已俘获了一些解决同类任务且具有相容任务情景的信息体，则获取和参阅这些（尤其是任务情景相容度最好的）信息体，显然可以促进知识工作者创新灵感的产生。鉴于为这些信息体建立的适用性索引基于同一任务情景描述模式和目标分类体系，KAMTC 方法有助于准确获取匹配当前业务工作需求的信息体。一旦知识工作者以这种方式高效创建了解决当前问题所需的新含蓄知识，并将其外表化为明晰的创新知识，就进入“俘获创新知识”环节，由 KM 系统自动给包含这些创新知识的信息体建立适用性索引（包括目标分类和任务情景），并征求作者的认可和可能的修改。

如此，通过这 3 个环节的无缝链接，KAMTC 方法可以高效地推动有针对性的知识创新，并由此促进个人和组织业务的及时创新。让信息体（包含明晰的创新知识）作者直接参与创新知识的俘获（建立任务情景），最大的意义在于使 OM 的进化能与个人在业务工作中的知识创新同步，以促进知识工作者之间通过及时共享创新知识的精诚协作，去优化创新型组织学习的效益。

5 KM 系统的实现

为实现基于任务情景的知识适用性管理，KM 系统的开发划分为两个层次（图 4）：

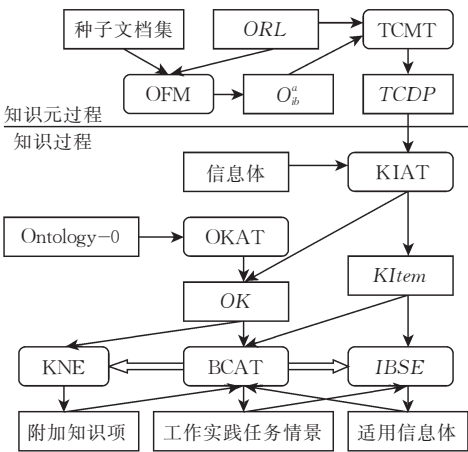


图 4 KM 系统的实现（划分为知识过程和知识元过程）

(1) 知识元过程。作为知识管理的上层，其主要工作是通过定义应用域共享本体论 O_i^a ($i=1,2,\dots,n$) 来建立面向 KM 的表示理论和描述框架，作为知识共享和重用的语义基础。

(2) 知识过程。作为知识管理的下层，其以应用域共享本体论提供的任务情景描述模式和目标（任务）分类体系作为概念模型，支持上述创新型组织学

习 3 个关键环节的实现。

作为知识元过程和知识过程的公共信息平台，OM 成为支持和沟通这 3 个关键环节的媒介，而任务情景描述模式则构成这些环节协调工作的核心语义。

5.1 OM 建模

OM 建模为包括概念层、知识层和信息层的三层体系结构。

(1) 概念层。为收集进 OM 的信息体建立表示和组织本体论，可表示为 3 元组：

$$(KItem-Def, OS, ORL).$$

其中，

$KItem-Def$ 为知识项描述模式，作为信息体适用性索引的表示规范；

OS 为本体论集合，包括组织业务相关的多个应用域共享本体论 $O_1^a, O_2^a, \dots, O_p^a$ 和描述组织自身构成的本体论 O^o ；

ORL 为本体表示语言，用于构建本体论。

(2) 知识层。作为概念层本体论的示例，存储描述 OM 内容的知识，可表示为 2 元组：

$$(KItem, OK),$$

其中，

$KItem$ 表示知识项（收集进 OM 的信息体适用性索引）集合：

$$KItem = \{kitem_1, kitem_2, \dots, kitem_n\};$$

OK 表示关于组织的知识（遵从 O^o ），包括部门、业务项目、知识工作者的描述。

(3) 信息层。收集进 OM 的、由知识项索引的信息体集合： $IBody = \{ib_1, ib_2, \dots, ib_m\}$ 。

术语知识项（Knowledge Item, $KItem$ ）取意于对信息体元知识的抽象和封装，用以屏蔽信息体内容（明晰知识）的细节和表示形式的异构性；使得 KM 系统的搜索引擎和知识联网机制工作时，只需取用包含于知识项的同构元知识。 $KItem-Def$ 的 BNF 定义表示如下：

$$\begin{aligned} \langle KItem \rangle &:= \langle KItem-ID \rangle \langle Application-Goal \rangle \\ &\quad \langle Content-Brief \rangle \langle Context \rangle \\ &\quad \{ \langle Task-Circumstance \rangle \}^+ \langle IBody-Link \rangle. \\ \langle Content-Brief \rangle &:= \langle Topic \rangle \langle Abstract \rangle \langle Keywords \rangle. \\ \langle Context \rangle &:= \langle Authors \rangle \langle Publication \rangle \langle Publication-Date \rangle \\ &\quad \langle Reference \rangle \langle Project \rangle \langle Department \rangle. \end{aligned}$$

其中， $KItem-ID$ 为知识项标识，并包含了信息体内容的多种分类指示。 $IBody-Link$ 是指向信息体的链接，以便 KM 系统在搜索或关联到适当知识项时取用链接的信息体。 $Context$ 和（上下文） $Task-Circumstance$ （任务情景）分别表示两种不同类型的适用性索引：信息体建立时的上下文（背景和动机）和

信息体内容的适用性特征(前景). Context 说明信息体的作者、所刊登的出版物、出版日期、参考资料、所属项目及其管理部门等,用于作为启发式知识去支持知识联网.

5.2 应用域共享本体论的构建

以支持任务情景的建模为主线,应用域共享本体论 O_i^a 定义为以下 6 元组:

$$O_i^a = (GOAL, TASK, TCDP, Mapping, Describing, O_{ib}^a),$$
$$Mapping: GOAL \rightarrow \mathbb{P}TASK (\mathbb{P} \text{指示 } TASK \text{ 幂集});$$
$$Describing: TASK \rightarrow TCDP;$$

其中,

$GOAL$ 为应用域中追求的应用目标集,目标层次分解,构成目标分类体系;

$TASK$ 为应用域关注的任务集: $TASK = \{t_1, t_2, \dots, t_k\}$;

$TCDP$ 为任务情景描述模式的集合: $TCDP = \{CDP^{t1}, CDP^{t2}, \dots, CDP^{tk}\}$;

$Mapping$ 为指示每个目标的达到可能涉及的任务子集,且目标分类体系中父目标的任务集为子目标任务集的并集;

$Describing$ 为指示每个任务都有一个对应的任务情景描述模式;

O_{ib}^a 为应用域基础本体论,包括术语集、同义词、术语分类体系:

$$O_{ib}^a = (TERM, CORE, Synonyms, Subsuming),$$
$$Synonyms: CORE \rightarrow \mathbb{P}(TERM - CORE);$$
$$Subsuming: CORE \rightarrow \mathbb{P}CORE;$$

其中,

$TERM$ 为应用域的术语集;

$CORE$ 为核心术语集,集中任意两个术语都不同义;

$Synonyms$ 为某些核心术语具有来自 $(TERM - CORE)$ 的同义词;

$Subsuming$ 为核心术语间存在语义包含关系,正是这些关系联合形成应用域中的术语分类体系.

O_i^a 由 KM 系统的开发者遵从领域专家的意见用本体表示语言 ORL 定义. ORL 支持对应用域的概念化描述,包括概念(对象类)、关系和特性,并支持应用域术语集和术语分类体系的建立. 概念的 BNF 形式表示如下:

$$\langle \text{Concept-Definitions} \rangle :=$$
$$\{ \langle \text{Concept } \langle \text{Concept-Name} \rangle [\text{Super}: \{ \langle \text{Super-Class-Name} \rangle \}^+;]$$
$$\langle \text{Slot-Name} \rangle; \text{type } [\text{list-of } \langle \text{Slot-Content-Type} \rangle \}^+ \}^+.$$

概念的 Super 槽用于建立概念间的包含关系(槽值是单一超类名)和对于概念的复合定义(槽值

是多个并列的超类名),进而支持概念分类体系的建立. 通过将槽值的类型定义为另一概念,可以建立概念间的任意自定义关联.

作为应用域本体论 O_i^a 的核心内容,任务情景描述模式就定义为概念,概念名即为任务名. 概念的槽名指示任务情景的特征槽,而给特征槽提供槽值的术语可选集则通过 ORL 的 $\langle \text{Type-Definitions} \rangle$ 元素定义为枚举数据类型. 例如,图 2 中的任务情景描述模式就可定义为取名“Document Annotation”的概念,其特征槽“Objective”定义成:

$$\text{Objective type list-of Objective-Term};$$

而“Objective-Term”则定义为指示术语可选集的枚举数据类型:

$$\text{Type Objective-Term: base-type string,}$$
$$\text{restriction enumeration}$$
$$(\text{"Document"}, \text{"Web page"}, \text{"Email"}, \text{"Video"}).$$

建立任务情景描述模式的关键在于确定适当的特征槽及其术语可选集,使得抽取的任务情景能够确切和全面地反映信息体内容或业务工作中相应任务及其处境的特征. 同时,目标名、任务名、特征槽名和术语可选集包含的术语都应取自应用域共享本体论 O_i^a 定义的术语集. 这个术语集构成上述应用域基础本体论 O_{ib}^a 的主要部分.

尽管特征槽及其术语可选集的确定必须依赖领域专家的决策,仍然可以提供辅助手段去加速确定过程. 那就是首先建立 O_{ib}^a , 并将其作为建立任务情景描述模式的必要基础.

如图 4 所示,基于本体表示语言 ORL,知识元过程提供建模工具 TCMT 和 OFM,分别用于支持任务情景描述模式 $TCDP$ (和目标分类体系)以及应用域基础本体论 O_{ib}^a 的构建、维护和进化.

依据由领域专家提供的种子文档集——从应用域精选出的有代表性的电子文档集,OFM 应用基于单词或词组的索引构建工具去自动建立初始的应用域术语集(图 5). 然后,辅助领域专家做术语筛选、同义词划分、并产生 $CORE$ 和确定核心术语间的包含关系,最后建立术语分类体系. $CORE$ 的规模取决于应用域的狭窄程度和业务工作的查询需求,并影响任务情景的描述精度以及建立 O_{ib}^a 的工作量.

基于 O_{ib}^a , 领域专家在 TCMT 的辅助下,只需从 O_{ib}^a 的核心术语集 $CORE$ 选取(通过可视化交互式人机界面)适当的术语作为目标名、任务名或槽名,并建立术语可选集,就可快捷地建立 $TCDP$ 和整个 O_i^a . 随着 $TCDP$ 和 O_i^a 的不断进化(参见 4.2 节),任务情景的描述精度将逐步提高,以更有效地支持信息体的适用性查询和 OM 的开放性进化. 由于篇幅

限制,知识元过程的实现细节不加赘述.

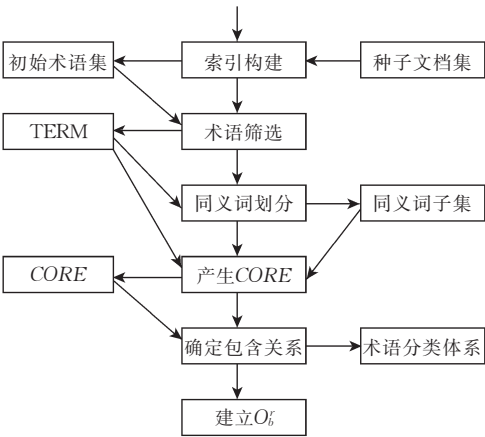


图 5 OFM 工作原理

5.3 知识过程

知识过程以 OM 的概念层为指导,支持上述创新型组织学习的 3 个关键环节的实现:抽取查询需求,获取适用知识和俘获创新知识.

如图 4 所示,知识过程提供知识项抽取工具 KIAT 和组织知识抽取工具 OKAT 去支持 OM 知识层(KItem 和 OK)的构建、维护和进化. KIAT 的核心功能是自动俘获创新知识——给包含创新知识的信息体建立知识项,包括信息体建立的上下文知识和面临任务的情景描述.

业务情景抽取工具 BCAT 引导知识工作者抽取当前业务工作面临任务(问题)的情景描述,作为精确的查询需求. 信息体搜索引擎 IBSE 搜索任务情景与业务工作相容匹配的知识项,以获取适用于当前业务工作的信息体. 知识联网引擎 KNE 则基于知识项提供的上下文知识和关于组织的知识(如部门、业务项目、知识工作者等实体)的结构化描述,支持隶属于相关实体的知识项之间的启发式关联,实现知识联网^[11],以辅助知识工作者获取与当前业务工作相关的附加知识项. 由于篇幅限制,知识过程的实现细节不加赘述.

6 应用分析

我们应用 KAMTC 方法开发了计算机学院科研知识管理系统原型,定义了面向 KM、DM(数据挖掘)、ERP(企业资源规划)、Agent 技术的应用域共享本体论以及包括研究室、科研项目 and 科研骨干描述模式的组织本体论. 以面向 KM 的应用域为例,我们精选了覆盖 KM 主流研究方向的代表性论文 80 篇,其中 13 篇面向信息体标注,22 篇面向知识元过程,20 篇面向知识过程,15 篇面向信息体定制,10

篇是本研究小组的论文. 面向 KM 的应用域目标分类体系如图 6 所示,其包括两个目标层和 1 个任务层;图 2 所示的任务“Document Annotation”就作为下层目标“Knowledge Process”的任务之一. KM 本体论的 CORE(核心术语集)收集术语 363 个,用于建立面向任务情景各特性槽的术语可选集.

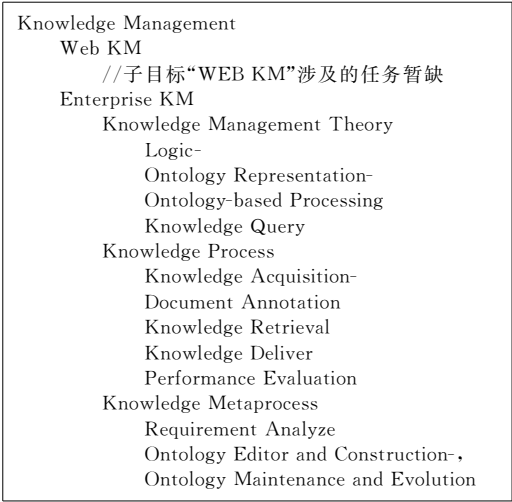


图 6 面向 KM 的应用域目标分类体系

实验结果表明,通过引进基于任务情景的相容匹配搜索机制,KAMTC 方法大幅度提高了搜索结果的适用性. 例如,应用传统的关键词搜索方法,在这个原型系统的 OM 中查寻包含关键词“Knowledge Management”的文献,会返回 41 篇;再附加关键词“Knowledge Annotation”,也会返回 12 篇文献;作为对比,应用基于任务情景的相容匹配搜索机制,依据任务“Document Annotation”的情景描述,采用“and”策略,只返回 2 篇文献. 由于任务情景以结构化方式确切描述了业务工作的查询需求,匹配需求的这 2 篇文献对于研究面向 KM 的知识标注具有较高的参考价值(适用性);作为关键词搜索结果的 12 篇文献,尽管也包括这 2 篇文献,但其它文献大多只是提到术语“Knowledge annotation”,并未阐述相关技术细节. 我们指出,对于知识密集型的业务工作,由于存在大量类同的文档资料,搜索结果适用性的提高及其效用将更为显著.

另一方面,在关键词较多(3 个以上)或关键词冷僻的情况下,传统的关键词搜索方法往往搜索不到相关的文献;作为对比,基于放松的相容匹配搜索策略(part-and, or, part-or)、同义词和术语分类体系以及知识联网机制,KAMTC 方法能够给知识工作者提供有潜在参考价值的文献或查询线索.

作为模拟实验,我们建立了 1 万个面向同一抽象任务的知识项,任务情景描述模式包括 6 个特征

槽,每个槽的术语可选集包括 8 个术语. 建立所有知识项和当前业务工作的任务情景:每个槽的填充值(两个术语)从相应的术语可选集随机选取. 以 1000 个查询例作统计,采用 and 策略,查询到任务情景与当前业务工作相容匹配的知识项稳定在 2.44 个. 作为对照,以任务情景描述模式包括的所有术语为索

引,采用常规的关键词查询法,则查询到的知识项随并列关键词个数的增加而指数级减少(表 1). 需要指出,由于缺少任务情景描述模式的指导,真实查询时,知识工作者列举的关键词(尤其是多于 3 个时)易超出索引范围,从而导致返回的知识项从许多骤降到 0.

表 1 查询方式和结果的比较

任务情景匹配方法		基于关键词方法				
查询策略	and 策略	1 个关键词	2 个关键词	3 个关键词	4 个关键词	5 个关键词
平均返回	2.44	1250	156.25	19.53	2.44	0.31

手工建立知识项及其任务情景是推动 KAMTC 方法实用化的最大障碍. 鉴于已给这二者的自动化建立提供了支持,知识工作者只需通过人机界面对自动建立的任务情景作审查和必要的修改,就可便捷完成知识项的设置和提交. 通过人机界面提供的应用域目标分类体系、任务情景描述模式、同义词和术语分类体系,KAMTC 方法不仅辅助知识工作者加速对于业务工作的认识,也允许他们只作快捷的选择操作,就能提出确切和全面的查询需求.

7 评价和结论

为促进创新型组织学习的实现,KM 系统应从 3 个关键环节加以支持:通过引导对于业务工作任务(问题)的认识,辅助知识工作者抽取确切、全面的查询需求;获取适用于这些任务的明晰知识(包含这些知识的信息体),去促进个人的知识创新和业务创新;通过即时建立适用性索引,让 OM 及时俘获个人的创新知识(包含创新知识的信息体),使 OM 进化和组织学习能与个人的知识创新同步.

尽管创新型组织学习已成为系统科学和工程学科研究的热点,但如何用 KM 技术来促进知识创新和组织学习,并没有得到深入、系统的研究. 如第 3 节所述,目前 KM 系统指示信息体适用性的主流方法是建立基于关键词或词组的索引,导致搜索引擎的适用性查询精度低下(参见第 6 节). 尽管抽取信息体建立时的上下文作为启发式知识^[1,12],可以支持知识联网,以辅助知识工作者增加俘获适用信息体的机会,但无助于适用性查询精度的直接提高. 文献^[13]开发了允许知识工作者参与 OM 开放性进化的 KM 系统,但未解决适用性查询精度低下的问题.

近年来,应用域共享本体论的研究和开发为信息体适用性和上下文的描述以及知识工作者主动参与的 OM 进化奠定了语义基础^[14-16]. 然而,受检信息体适用性评价的固有模糊性,使得基于本体的适

用性描述仍面临严重挑战.

本文提出基于任务情景的知识适用性管理方法 KAMTC,较好地解决了这一挑战性问题. 以应用域任务情景描述模式和目标分类体系作为统一的语义基础,去确切、全面和便捷地描述业务工作的查询需求和受检信息体的适用性,可以有效和高效地支持创新型组织学习的 3 个关键环节:抽取查询需求、获取适用知识和俘获创新知识;使得知识工作者能够便捷获取适用性好(匹配当前业务创新需求)的明晰知识,并便捷和即时地参与 OM 的开放性进化.

参 考 文 献

[1] Schwartz D, Te'eni D. Tying knowledge to action with kMail. IEEE Intelligent Systems, 2000, 15(3): 33-39

[2] Domingue J D, Motta E. PlanetOnto: From news publishing to integrated knowledge management support. IEEE Intelligent Systems, 2000, 15(3): 26-32

[3] Malhotra Y. Knowledge management & new organization forms: A framework for business model innovation. Information Resources Journal, 2000, 13(1): 5-14

[4] Marwick A D. Knowledge management technology. IBM Systems Journal (Knowledge Management), 2001, 40(4): 814-831

[5] McKeen J D, Zack M H, Singh S. Knowledge management and organizational performance: An exploratory survey//Proceedings of the 39th Hawaii International Conference on System Sciences. Hawaii, USA, 2006: 1-9

[6] Ashrafi N et al. A framework for implementing business agility through knowledge management systems//Proceedings of the 2005 7th IEEE International Conference on E-commerce Technology Workshops (CECW'05). Munich, Germany, 2005

[7] Polanyi M. The Tacit Dimension, Knowledge in Organizations. Woburn, MA: Butterworth-Heinemann, 1997

[8] Nonaka I, Takeuchi H. The Knowledge Creating Company. Oxford, UK: Oxford University Press, 1995

[9] Turney P. Learning to extract keyphrases from text. National Research Council, Institute for Information Technology, Canada, NRC Technical Report ERB-1057, 1999

- [10] Arampatzis A, Tsoris T, Koster C, van Der Weide T. Phrase-based information retrieval. *Information Processing and Management Journal*, 1998, 34(6): 693-707
- [11] Mika P, Akkermans H. Towards a new synthesis of ontology technology and knowledge management. Free University Amsterdam, Netherlands; Technical Report IR-BI-001, 2004
- [12] Rabarijaona A, Dieng R, Corby O, Ouaddari R. Building and searching an XML-based corporate memory. *IEEE Intelligent Systems*, 2000, 15(3): 56-63
- [13] Fischer G, Ostwald J. Knowledge management: Problems, promises, realities, and challenges. *IEEE Intelligent Systems*, 2001, 16(1): 60-72
- [14] Fensel D. Ontology-based knowledge management. *IEEE Computer*, 2002, 35(11): 56-59
- [15] Warren P. Knowledge management and the semantic Web: From scenario to technology. *IEEE Intelligent Systems*, 2006, 21(1): 53-59
- [16] Pandey S, Mishra R. Knowledge discovery and ontology-based services on the grid (a survey report)//*Proceedings of the 6th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT'05)*. 2005: 1033-1038



GAO Ji, born in 1946, professor and Ph.D. supervisor. His research work focuses on artificial intelligence, network computing, software engineering, and information intelligence.

Background

The research work presented in this paper is supported by the National Basic Research Program (973 Program) of China under grant No. 2003CB31700. This program researches the basic theories, models, methodologies, and key techniques of Semantic Grid, and includes the research task "The Theory and Technology of Knowledge Provision for Supporting the Creative Design of Products in Knowledge Grid Environment" as a subprogram. The authors desire that the proposed methodology KAMTC can facilitate the creative design effectively.

Although COL has become the research hotspot for system science and engineering subject, it has not gained in-depth and systematic investigation how to use KM technology to facilitate knowledge creation and organizational learning.

Since the query requirements for business work and the applicability of searched information bodies may be described in terms of the uniform semantic basis (i. e. domain ontology based circumstance description patterns of tasks), the KM systems adopting KAMTC can support effectively and efficiently the three key phases for realizing COL: abstracting query requirement, acquiring applicable knowledge, and capturing creative knowledge. Therefore knowledge workers can acquire handily applicability-good explicit knowledge to facilitate creative inspiration, and participate in the open evolution of OM agilely and in time.

The authors have published five papers for KM research work in *Journal of Computer-Aided Design & Computer Graphics*.