A Cooperative Multi-Agent Platform for Invention based on Ontology and Patent Document Analysis

Von-Wun Soo, Szu-Yin Lin, Shih-Yao Yang, Shih-Neng Lin, and Shian-Luen Cheng
{soo, stan, yao, goz, luen}@cs.nthu.edu.tw
Department of Computer Science
National Tsing Hua University
HsinChu, Taiwan

Abstract

We propose a cooperative multi-agent platform to support the invention process based on the patent document analysis. It helps industrial knowledge managers to retrieve and analyze existing patent documents and extract structure information from patents with the aid of ontology and natural language processing techniques. It allows the invention process to be carried out through the cooperation and coordination among software agents delegated by the various domain experts in the complex industrial R&D environment. Furthermore, it integrates the patent document analysis with the inventive problem solving method known as TRIZ method that can suggest invention directions based on the heuristics or principles to resolve the contradictions among design objectives and engineering parameters. We chose the patent invention for Chemical Mechanical Polishing (CMP) as our case study. However, the platform and techniques could be extended to most cooperative invention domains.

Index Terms — cooperative invention, patent document analysis, multi-agent cooperative platform, knowledge management, TRIZ method.

I. Introduction

In knowledge-based economy age, the management of intellectual properties has become more and more important for an industry. Many companies hold their patents not only as an invisible asset but also as a strategy for its development and competition against its competitors in the market. However, inventing a new patent does not seem to be an easy task that can be systematically and automatically carried out. In the past, the invention process in a company often involved complex decision-making from different domain experts such as design engineers, manufacture engineers, patent engineers, market engineers, etc. who collaborate together. The domain experts may contribute their expertise by proposing and modifying some candidate solutions until a solution that satisfies most constraints and design objectives or specifications is found. A final design is often a compromised solution after several iterations of cooperative communications and negotiations among different domain experts. Therefore, the cooperative invention process is very time consuming and can be a bottleneck for most companies. How to use a multi-agent cooperative platform to facilitate and speed up the invention process is one of the major goals of the research.

Another problem for invention is that most “novel” ideas often turn out to be a “re-inventing the wheel” or might impinge on some claimed rights protected by other’s patents. Before developing a new product and inventing a new patent, it had better ensure that it does not impinge other’s patents by consulting with some world patent database. In fact, the World Intellectual Property Organization revealed that 90% to 95% of world’s inventions are found in patented documents. The European Patent Office also disclosed that more than 80 percent of man’s technical knowledge is described in patent literature.” It can increase the efficiency of invention substantially and reduce the risk of impinging on the patent rights of others if we know how to extract and utilize effectively the information embedded in the large patent document base. Unfortunately, most studies in the patent document analysis are on document search and classification [3] [4] [5], very few discussed how to use the patent documents to further support the invention.

Our research objective is to integrate the patent document analysis techniques with the cooperative invention process. We propose a cooperative multi-agent (MA) platform for invention based on ontology and patent documents. The multiple agents include ontology agents and thesaurus agents, invention Agent (such as TRIZ agent), coordination agent, domain problem solving agents (such as design agent, material agent, manufacturing agent in the mechanical design problem domain), and patent Agents (including patent information retrieval agent and patent information extraction agent). The patent documents for a given domain can be retrieved from the U.S. Patent and Trademark Office (USPTO) database. We implemented these agents to assist engineers to exchange the needed design knowledge in the inventive problem solving based on the patent document analysis. The goal is to partially automate such tasks to be carried out by the design and patent engineers as (1) searching for existing patent documents similar or related to a potential new invention by domain ontology and patent knowledge base, (2) Analyzing related patent documents by semantically annotating the patent documents and extracting relevant design parameters and claims from them, and (3) using TRIZ method suggest new direction of invention based on the domain parameters and principles and the design objectives and specifications. The MA platform supports the cooperation and coordination among different agents delegated by corresponding engineers or experts who play some roles in the invention process.

II. BACKGROUND

2.1. MAS for Cooperative Invention

A multi-agent system (MAS) [8] consists of a collection of autonomous agents who can define their own goals and actions and can interact and collaborate among each other through communication. As mentioned in the introduction, the invention relies on collaborations among different domain experts. With proper ontology and workflow implementation, the multi-agent can achieve a piece of invention through cooperation and **c**oordination [9], [10].

2.2. Domain Thesaurus and Ontology

A thesaurus is a hierarchical structure which classifies domain terminologies into different concepts. It's very common that the domain-specific terms cannot be covered by common dictionaries. Therefore the domain thesaurus is needed for machines to process specific domain corpus to understand what are these terminologies mean.

Ontology is abstracted domain concepts and relations expressed in terms of a standard knowledge representation language that can be reused and shared by many users over internet. The standard knowledge representation languages have been defined as RDF (Resource Description Framework) and OWL (Web Ontology Language) by the semantic web consortiums in the internet. The Semantic Web provides a common framework that allows data to be shared and reused across application, enterprise, and community boundaries [14]. We implemented the schematic conceptual relations of the engineering design of a particular domain in terms of the standard ontology. The case study domain we chose is the patent invention in CMP (Chemical Mechanical Polishing). The CMP is a global planarization process that utilizes both the chemical slurry and mechanical polishing to remove the passivation layer of a wafer. The CMP device is very expensive and is an important technology that have existed many patents that claims different ways of achieving global wafer planarization at different costs and efficiencies. As shown in Figure 1, it illustrates a schematic CMP device that consists of such components as the polishing head, the polishing pad and polishing plate, the wafer, and the polishing slurry. Therefore we need to implement some fundamental concepts as the domain ontology for CMP in terms of OWL.

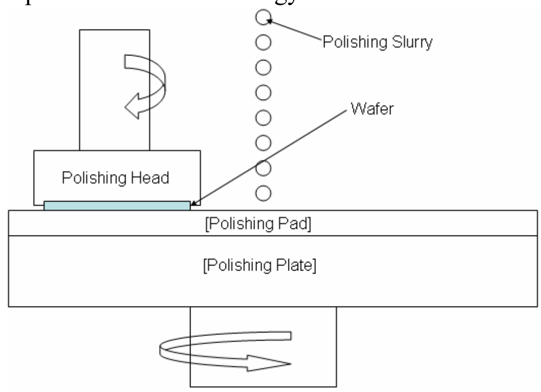


Figure 1: A schematic illustration of a CMP device

2.3. Inventive Problem Solving Using TRIZ

The common methods used for stimulating invention are the brain storming methods [6]. The main drawbacks for using the brain storming method as an invention technique are: 1) it can be affected by the limited knowledge and mental barriers of participated engineers, 2) it can be very time consuming, and 3) it lacks of a systematic process. The TRIZ method [7], the acronym for "Theory of Inventive Problem Solving" in Russian, can overcome the drawbacks. It was developed by Genrich Altshuller and his colleagues in 1946 and has become popular throughout the world. Many companies have been using TRIZ to find new solutions for many real world problems.

Altshuller extracted 39 standard technical characteristics that ever caused conflicts from over 1,500,000 worldwide patents. They are called the 39 engineering parameters such as Weight of Moving Object, Weight of Nonmoving Object, Length of Moving object, Length of Nonmoving Object, Area of Moving Object, etc. [11]. TRIZ method employs a concept of contradiction matrix in which the rows indicate the parameters needing improvement while the columns indicate the parameters that have been degraded as a result of improving the parameter in the row. TRIZ has 40 inventive principles as hints or heuristics to improve or even solve the contradiction such as Segmentation, Extraction, Local Quality, Asymmetry, Combining, Universality, Nesting, and etc. Users can obtain the principles from the entry in the contradiction matrix as a solution to the problem.

2.4. Patent documents

A patent document consists of many fields. A large number of these fields are small and not text-like information such as the application number, the patent number, the dates of application and issue, and figures. A few fields contain specific pieces of text information such as the names and addresses of authors, assignees, patent examiners, and patent attorneys. A few fields such as abstract, background summary, detailed description, and claims are also expressed in natural language text. The claims in a patent are written in legalistic language and must be analyzed intensively and converted into a structured format that can be used further in the purposes of patent documents classification, retrieval, comparison, and so on with the aid of ontology and natural language technology.

III. The Cooperative Multi-agent Platform for Invention

The overview of cooperative invention platform based on patent document analysis is shown in Figure 2.

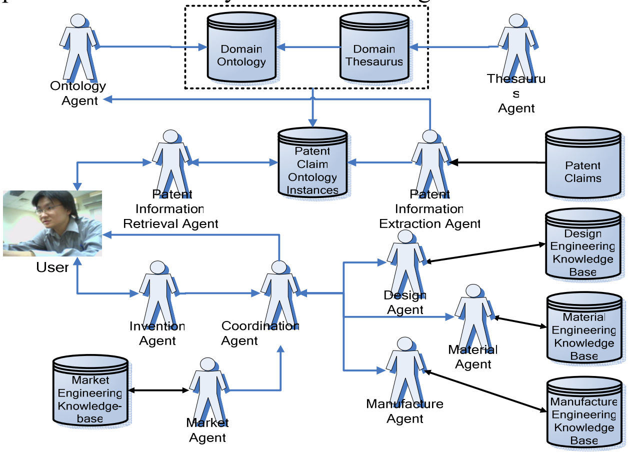


Figure 2: The MA cooperative platform for invention based on ontology and patent document analysis

There are six types of agents; Ontology Agents and Thesaurus Agents, Invention Agent (such as TRIZ Agent), Coordination Agent, Domain Agents (such as Market Agent, Design Agent, Material Agent, Manufacturing Agent in the mechanical design problem domain), and Patent Agents (including patent information retrieval agent and patent information extraction agent). Besides the ability to cooperate using FIPA Agent Communication Language (ACL) each agent has its own special capabilities.

1) Thesaurus Agent

The Thesaurus Agent assists users to extract domain-specific terminologies from patent documents. It utilizes NLP tools to get syntactic and semantic information from patent documents by a semantic and syntactic tagger, and pattern matching rules to extract candidate domain terminologies. Then the domain experts verify the candidate terminologies and classified them into correct semantic hierarchical tree in the thesaurus. The workflow of the thesaurus agent is shown in Figure 3.

To extract domain terminologies, it normally needs to extract proper noun phrase as a group instead of separate individual nouns. For example, the "rotating speed" should combine two words as a domain term, the separate 'rotating" and "speed" did not represent the correct domain concept. As shown in Figure 3, the term "rotating speed" has a semantic code "B1:2:2:1:1", and the concept "Rotational Speed" in the hierarchy has code "B1:2:2:1", so the agents could interpret the "rotating speed" as a "Rotational Speed" concept.

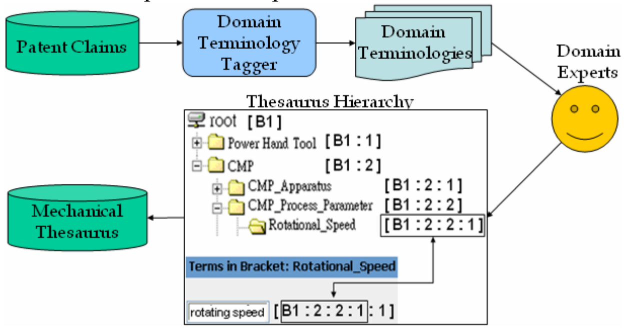


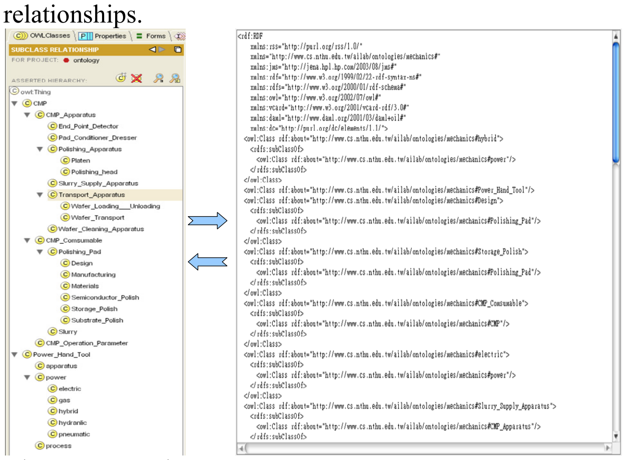
Figure 3: The workflow of thesaurus construction

At present, the domain terminology tagger can extract domain terminologies at 92.62% precision rate and 90.02% recall rate. The recall and precision are evaluated as following formula:

Precision :
# of expert_Marked_Terms / # of candidate_Domain_Terms
Recall :
(# of total_Words — # of candidate_Domain_Terms) / # total_Words

2) Ontology Agent

The ontology agent has two major goals to help domain experts. The first is to build the domain ontology; the second is to refine the domain ontology. The ontology agent can convert the thesaurus hierarchy structure into the OWL format, and then the ontology construction tool such as Protégé, as shown in Figure 4, can import the domain terminologies for domain experts to define their relationships.



Thesaurus Hierarchy at Protégé 3.0 OWL format

Figure 4: Thesaurus hierarchy is converted into the OWL format

After the ontology agent finishes the transformation from a thesaurus hierarchy into the OWL format, the domain experts can define term relations in the Protégé tool.

For example, the terms “Wafer” and “Polishing Pad” have the “Polish” relation in the physical world can be defined as a triple relation as shown in Figure 5.

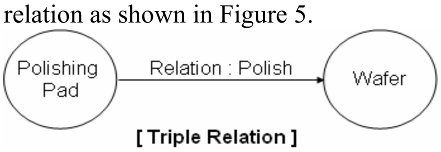


Figure 5: A triple relation

The corresponding OWL format of the domain triple relation is illustrated in Figure 6.

```
<owl:ObjectProperty rdf:ID="Polish">
  <rdfs:subPropertyOf rdf:resource="#hasRelationship"/>
  <rdfs:domain rdf:resource="#PolishingPad"/>
  <rdfs:range rdf:resource="#Wafer"/>
</owl:ObjectProperty>
```

Figure 6: The triple relation in OWL format

After many domain triple relations are defined in Protégé, the domain ontology in OWL format will be constructed.

3) Patent Information Extraction Agent

The workflow for the Patent Information extraction agent is shown in Figure 7. In order to convert a patent claim corpus into machine readable format, the Patent Information Extraction Agent receives each patent claim and transforms its statements into so called ontology instances. The agent first annotates the terms in the claim with the semantic/syntactic information, and then it extracts the design structure in the patent claims using techniques of regular expression and NLP technologies. The design structure is mapped into domain ontology in OWL that describes the relationships and attributes of design components mentioned in the claims. The OWL instances of the given patent claims are stored in the “Claim Ontology Instances” for user queries.

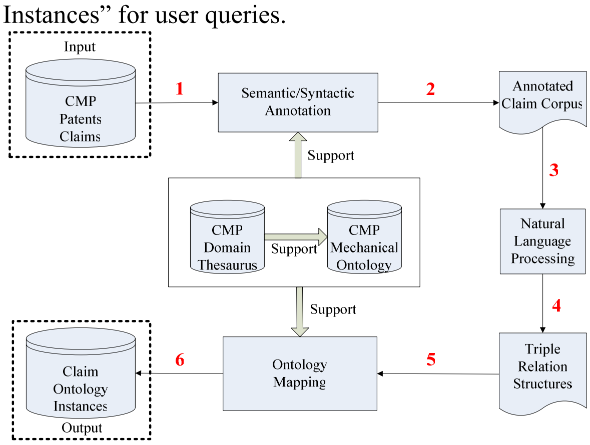


Figure 7: Workflow of Patent Information Extraction Agent

For example, an original claim from USPTO patent 6524176 shown in Figure 8 is converted into an ontology instance in Figure 9.

What is claimed is:
1. A polishing pad comprising:
a first layer;
a second layer;
a hole formed in the polishing pad, the hole having:
a first section in the first layer of the polishing pad; and
a second section in the second layer of the polishing pad; and
a hollow plug embedded in the hole;
wherein the hollow plug has an upper portion and a lower portion, and the upper portion of the hollow plug fits into the first section of the hole, and the lower portion of the hollow plug fits into the second section of the hole.

Figure 8: USPTO Patent 6524176, claim-1

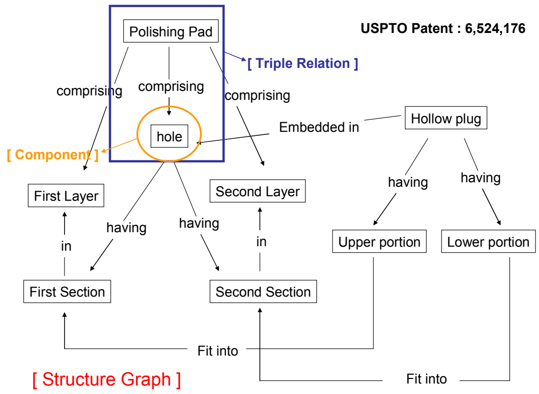


Figure 9: An ontology instance in Patent 6524176

4) Patent Information Retrieval Agent

The Patent Information Retrieval Agent retrieves relevant patents and makes a brief report according to queries given by the user. The queries can be expressed in terms of not only keywords but also triple relations. For example, a query can be a set of triples like “(polishing_pad,contain, hollow_plug),(hollow_plug,contain,upper_portion),(hollo w_plug, containt, lower_portion)” as shown in Figure 10.

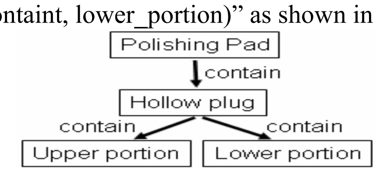


Figure: 10: An example query in triples

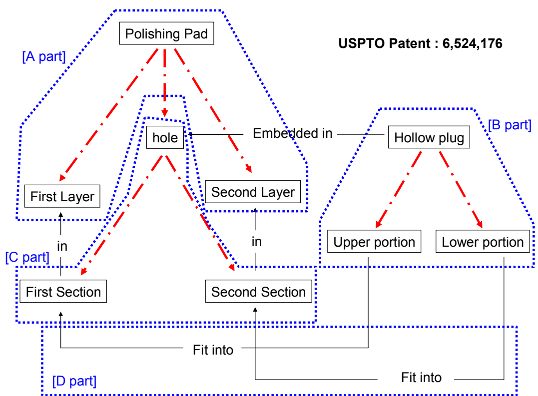


Figure 11: An instance structure in Patent 6524176

The patents containing the proper keywords, component properties, or similar mechanical structure will be returned by the patent information retrieval agent. The answers of the query include patent 6524176 because the agent found that the structure similarity of B part in Figure 11 is similar to the query structure.

5) Invention Agent

Invention agent uses TRIZ in the invention process to suggest principles to improve or solve the problem. So we call it a TRIZ agent. The first step of an invention process is to identify the problem. We allow users to search a patent document from the engineering patent knowledge base and express their problem to be solved in terms of physical contradictions. The user has to express a technical characteristic that he/she wants to improve and a technical characteristic that will be worsen when he/she solves the problem. The user enters the improved technical characteristic to be imprived, the technical characteristic to be worsen, and the patent document with the domain ontology at the initial step of the invention process. The appropriate inventive principles suggested by TRIZ agent are then mapped with the domain ontology in order to find the related attributes and possible problems. The related attributes and possible problems are responsed to the user in terms of Agent Communication Language (ACL) messages to the coordination agent. The coordination agent also sends the related attributes and possible problems to different domain solving agents such as design agent, material agent, and manufacture agent for further analysis.

We illustrate a case of invention process of a CMP that uses invention agent. The problem to be solved is stated as: when the rotation speeds at the outer diameter and inner diameter of the polishing pad are different, the outer diameter of the polishing pad will polish and remove more particles from the surface of a wafer than the inner diameter and cause an uneven polishing problem. Therefore the problem to be improved belongs to the aspect on the rotation speed.

At first step, the User inputs two engineering parameters that are required by TRIZ and a patent claim ontology instance in OWL format to Invention Agent. The improving engineering parameter in this case is the rotational speed, so we choose #9 (Speed) and the worsen parameter is #15 (Durability of Moving Object) because to the field of speed is changed.

At the second step, Invention Agent returns the results in terms 4 principles (#03: Local quality, #19: Periodic action, #35: Transformation of physical and chemical states of an object, #05: Combining or integration) in the TRIZ Contradiction table to User.

At the third step, the User considers principles #03, #05, #35 is less relevant to the problem of the field of speed and thus chooses principle #19 to solve the problem. Invention Agent also lists that principle #19 includes 3 rules for the User (1: Replace a continuous action with a periodic (pulsed) one. 2: If an action is already periodic, change its frequency. 3: Use pulsed between impulses to provide additional actions). In this case, the User chooses rule 1 because it is more relevant to the problem.

At the fourth step, Invention Agent maps the Domain Ontology to the Patent Claim Ontology instances. Currently, all components in the Patent Claim Ontology instances are polishing particles, shape of polishing pad, shape of groove, structural parameters and Invention agent suggests to the Coordination Agent according to rule 1 of principle #19 that the two engineering parameters “Shape of polishing pad” and “Shape of groove” are to be modified and further analysis.

6) Coordination Agent and Domain Agents

The role of the Coordination Agent (CA) is to coordinate and cooperate with the User and such Domain Agents (DA) as Design agent, Material Agent, Manufacture Agent and Market agent to obtain a feasible solution as the new invention patent.

The DAs contain domain knowledge to find the best solution. They can conduct engineering analysis to verify if the solutions proposed by other agents violate the domain constraints or design objectives.

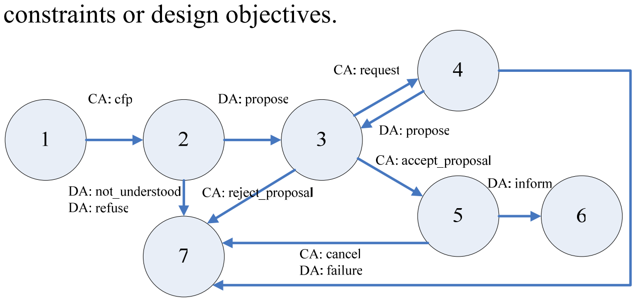


Figure 12: The Agents’ Coordination Model

The interaction protocol in our platform is shown in Figure 12.

In State 1, the CA sends a “cfp” (call for proposal) ACL message to all DAs and waits for DAs to propose. Each DA sends a “propose” ACL message to CA in State 2.

Here is an example of a FIPA ACL message in our coordination protocol:

```
(request
  :sender CA
  :receiver  DA1
  :content   (uspto6524176, shape of polishing pad, circle)
  :ontology domainontology)
```

The “propose” ACL message comprises initial proposal for invention engineering parameters. In the State 3, CA integrates all initial proposals. According to the initial proposal, it sends “request” ACL messages to other DAs for further analysis (State 4). After CA collects proposals and suggestions from all DAs, it sends an “accept_proposal” (State 5) ACL message or a “refuse_proposal” message (State 7), which the content of proposal is unworkable to the

DAs who send the original proposal. In State 4, the DA modifies and adds additional suggestions to the initial proposals depending on their knowledge domains (design, material, manufacture, or market), and return to State 3. In State 5, the protocols end when the accepted DA sends an "inform" ACL message to confirm the acceptation (State 6).

For example, the CA receives the suggestion to modify two CMP engineering parameters (e.g. "Shape of polishing pad" and "Shape of groove") from Invention Agent. The CA sends a cfp ACL message for "Shape of polishing pad" and the other cfp ACL message for "Shape of groove" to all DAs. After several iterations of communication protocols by collecting solutions from DAs, CA finds a solution to change the shape of polishing pad to a circle. The solution is delivered to the User as a useful invention design.

IV. Discussion and Conclusion

We select JADE as the tool to implement the multi-agent infrastructure. JADE provides a conceptual framework for implementing specific agents that can support multi-threading and message passing across distributed environment. We need to implement the internal behaviors for specific agents that contribute directly to the contexts needed for the analysis and processing of patent documents and the invention problem solving. Without proper thesaurus and domain ontology, it is impossible for the agent to conduct intelligent work. The question is "How to develop domain thesarus and ontology quickly so that it can be easily applying the platforms to a new invention domain other than CMP. Our thesaurus and ontology building tools support the reduction of the effort of such a shift.

The paper contributes to the speed up of the cooperative invention process based on a multi-agent platform and the patent document analysis. It could reduce the communication efforts of human experts by using autonomous agents and standard communication protocols. In the future, we need to further augment the natural language processing techniques to extract ontological instances from patent documents and design automatic mapping techniques to map the TRIZ principle specifications into the domain problems. The coordination methods proposed is only very primitive at current state, the module can be further refined in order to cope with more sophisticated coordination problems such as dead locks and live lock (e.g. endless cycle) in the invention process.

V. Acknowledgment

This research is supported by Ministry of Economic Affairs of ROC under grant number 93-EC-17-A-05-S1-030. We would like to thank our collaborator from the group of professor Zone-Ching Lin at National Taiwan University of Science and Technology.

VI. References

[1] P. Stone and M. Veloso. Multiagent systems: A survey from a machine learning perspective. *Autonomous Robots*, 8(3):345–383, July 2000.
[2] Fall, C. J., Törcsvári, A., Benzineb, K., Karetka, G (2003). Automated categorization in the international patent classification. ACM SI IR Forum 37.
[3] Larkey, Leah S. (1999) A Patent Search and Classification System In Digital Libraries 99 - The Fourth ACM Conference on Digital Libraries (Berkeley, CA, Aug. 11-14 1999) ACM Press, pp. 79-87
[4] Athanasios Tsakalidis, Konstantinos Markellos, Katerina Perdikuri, Penelope Markellou, Spiros Sirmakessis, George Mayritsakis, "Knowledge Discovery in Patent Databases", in the proceedings of the 11th ACM Conference on Information and Knowledge Management (ACM-CIKM 2002), pp 672 674, November 4 9, 2002, McLean VA, US.
[5] G. Pahl and W. Beitz, Engineering Design: A Systematic Approach, Springer, London, 1996.
[6] Ideation International Inc., Tools of Classical TRIZ, Ideation International Inc., Southfield, MI, 1999.
[7] Gilles DUBOIS, Danielle BOULANGER , A Multi-Agent System using Semantic Metadata for the Cooperation among Multiple Information Sources, 4th European Conf. on Principles and Practice of Knowledge Discovery in Databases, Workshop Knowledge Management Theory and Applications, Lyon, France, Sept. 2000.
[8] Alf Inge Wang, Reidar Conradi and Chunnian Liu, A Multi-Agent Architecture for Cooperative Software Engineering, Twelfth International Conference of Software Engineering and Knowledge Engineering (SEKE'2000), Chicago, 6-8 July 2000.
[9] C. Liu and R. Conradi. Process View of CSCW. In Proc. of ISFST98, Ocon Technology Application, pages 46--51, Bremen, Germany, 15-17 September 1998. International Workshop on Intelligent Agents in Information and Process Management.
[10] Terninko, J., Zusman, A., and Zlotin, B., Systematic Innovation – An, Introduction to TRIZ (Theory of Inventive Problem Solving), CRC, Press LLC, 1998.
[11] Glenn Mazur, Theory of Inventive Problem Solving (TRIZ), http://www.mazur.net/triz, 1996.
[12] http://www.uspto.gov, The United States Patent and Trademark Office.
[13] http://www.w3.org/2001/sw/, W3C Semantic Web
[14] Federico Bergenti and Alessandro Ricci, Three Approaches to the Coordination of Multiagent Systems, Proceedings of the 2002 ACM Symposium on Applied Computing (SAC), March 10-14, 2002, Madrid, Spain. ACM

V.-W Soo is with Department of Computer Science, National Tsing Hua University, Hsin chu Taiwan ROC (phone: 886-3-5731068; fax: 886-3-5731068; e-mail: soo@ cs.nthu.edu.tw). Dr. Soo was born in Kaohsiung, Taiwan in 1953 and had got his PhD degree from Department of Computer Science, Rutgers, the State University of New Jersey in 1987. His major study interests are in artificial intelligence, intelligent agents and multi-agent negotiations, machine learning and bioinformatics.

Szu-Yin Lin, Shih-Yao Yang, Shih-Neng Lin, and Shian-Luen Cheng are the students with Artificial Intelligence Lab, Department of Computer Science, National Tsing Hua University.