ภาคผนวก

ภาคผนวก ก

ผลการทดสอบโปรแกรม

จากการทดสอบโปรแกรมโดยทำการจัดกลุ่มบทคัดย่อด้วยชุดข้อมูลทดสอบที่เป็นแฟ้มข้อมูลเอกสารบทคัดย่อทางวิชาการที่เกี่ยวกับฐานข้อมูลจำนวนทั้งหมด 100 บทคัดย่อ ใช้แฟ้มข้อมูลคำสำคัญที่เกี่ยวกับฐานข้อมูล โดยมีคำสำคัญจำนวนทั้งหมด 80 คำ และระบุจำนวนกลุ่มที่ต้องการแบ่งเป็น 10 กลุ่ม ได้ผลการจัดกลุ่มบทคัดย่อดังตาราง ก.1 ซึ่งมีรายละเอียดของรายชื่อบทคัดย่อในแต่ละกลุ่ม และคำสำคัญที่ปรากฏในบทคัดย่อของกลุ่มนั้น ๆ ที่มีความถี่มากที่สุดสามอันดับแรก พร้อมทั้งจำนวนของคำสำคัญนั้นด้วย โดยชื่อบทคัดย่อที่มีเครื่องหมาย *** หมายถึงบทคัดย่อที่เป็นศูนย์กลางของกลุ่ม

ตาราง ก.1 ผลการจัดกลุ่มบทคัดย่อ

กลุ่ม	ชื่อบทคัดย่อ	คำสำคัญ		
1		method	object-oriented	schema
	1. An extensible object-oriented database testbed	0	5	6
	2. An object-oriented prototype for a geophysical database	0	5	3
	3. Disk management for object-oriented databases	0	6	0
	4. Method-induced partitioning schemes for object-oriented databases	5	4	0
	5. Serializability in object-oriented database systems***	2	6	0
2		access	mining	web
	1. A Tool For World-Wide Web Access Log Analysis***	5	0	5
	2. Mobile agents for World Wide Web distributed database access	4	0	4

ตาราง ก.1 (ต่อ) ผลการจัดกลุ่มบทคัดย่อ

กลุ่ม	ชื่อบทคัดย่อ	คำสำคัญ		
2		access	mining	web
(ต่อ)	3. Data mining for Web intelligence	0	6	11
	4. Performance modelling and metrics of database-backed Web sites	0	1	8
	5. Supporting dynamic interactions among Web-based information sources	5	0	5
	6. Web Database and Its Applications in Teaching Database	0	0	7
3		model	secure	security
	1. Data and applications security developments and directions	0	1	6
	2. Identifying and representing the security semantics of an application	1	1	7
	3. Multilevel database security using information clouding***	1	0	5
	4. Multilevel secure databases a new approach	0	2	5
	5. Proceedings of the 1988 IEEE Symposium on Security and Privacy	3	2	4
	6. Providing security in a phone book database using triggers	0	1	3
	7. Security model consistency in secure object-oriented systems	10	2	7
4		algorithm	mining	model
	1. Data mining an industrial research perspective***	7	6	3
	2. Data types generalization for data mining algorithms	8	10	0
5		mining	technique	tool
	1. A fuzzy search method for rough sets in data mining	5	1	0
	2. A primer for understanding and applying data mining	9	0	1
	3. Data mining applications in BT	5	2	0

ตาราง ก.1 (ต่อ) ผลการจัดกลุ่มบทคัดย่อ

กลุ่ม	ชื่อบทคัดย่อ	คำสำคัญ		
5		mining	technique	tool
(ต่อ)	4. Autonomous decision-making a data mining approach	4	2	2
	5. Data mining for short-term load forecasting	6	0	0
	6. Data mining for the enterprise	7	0	3
	7. Data mining making data meaningful	4	1	1
	8. Data mining of printed-circuit board defects	4	0	0
	9. Data mining problems in medicine***	4	0	0
	10. Data mining with an ant colony optimization algorithm	4	0	0
	11. Data mining with Clementine	7	2	1
	12. Database technology for decision support systems	3	0	2
	13. Decomposition in data mining an industrial case study	5	0	1
	14. Graph-based data mining	2	1	0
	15. Mining very large databases	5	1	1
6		query	optimization	queries
	1. A query algebra for program databases***	10	1	3
	2. Associative query answering via query feature similarity	11	0	1
	3. Dynamic query re-optimization	7	2	2
	4. Query folding	7	1	5
	5. Query planning with limited source capabilities	17	1	0
	6. Querying source code using an algebraic query language	9	1	0
	7. The Kendall Square Query Decomposer	7	0	4
	8. Visual feedback in querying large databases	12	0	0
7		algorithm	warehouse	tool
	1. A new dynamic voting algorithm for distributed database systems	6	0	0
	2. An architecture for data warehouse systems	0	4	1

ตาราง ก.1 (ต่อ) ผลการจัดกลุ่มบทคัดย่อ

กลุ่ม	ชื่อบทคัดย่อ	คำสำคัญ		
7		algorithm	warehouse	tool
(ต่อ)	3. Competitive analysis of caching in distributed databases	6	0	0
	4. Consistency in data warehouse dimensions	1	4	0
	5. Data warehouse for EIS some issues and impacts	0	10	1
	6. Data warehouse tools	0	5	6
	7. Efficient maintenance of temporal data warehouses	0	10	0
	8. Evaluating data warehouse toolkits	0	5	3
	9. Evolving materialized views in data warehouse	2	3	0
	10. Modeling a faster data warehouse	0	8	0
	11. Multidatabase performance evaluation	5	0	0
	12. Practical lineage tracing in data warehouses***	2	3	0
8		algorithm	recovery	web
	1. Database recovery using redundant disk arrays	0	4	0
	2. Incremental recovery in main memory database systems	3	4	0
	3. Main memory database recovery algorithms and their performance	2	4	0
	4. Mobile user recovery in the context of Internet transactions	0	6	4
	5. Persistent applications using generalized redo recovery***	0	4	0
	6. Recovery guarantees for general multi-tier applications	0	6	2
9		distribute	multidatabase	technique
	1. A type system for an object-oriented database system	0	0	0
	2. A uniform model for temporal object-oriented databases	0	0	0
	3. Allocating data and operations to nodes in distributed database design	3	0	0

ตาราง ก.1 (ต่อ) ผลการจัดกลุ่มบทคัดย่อ

กลุ่ม	ชื่อบทคัดย่อ	คำสำคัญ		
9		distribute	multidatabase	technique
(ต่อ)	4. A World Wide Web Resource Discovery System	0	0	3
	5. An algebra and calculus for relational multidatabase systems	0	4	0
	6. Authorization and revocation in object-oriented databases	0	0	0
	7. Automating compensation in a multidatabase	0	4	2
	8. Clustering techniques for object-oriented database systems	0	0	2
	9. Data warehousing and the Internet's impact on ERP	0	0	0
	10. Executing multidatabase transactions	1	3	0
	11. Healthcare data warehousing and quality assurance	0	0	0
	12. Image retrieval by examples	3	0	1
	13. Implementation of interoperability in large multidatabases	0	1	0
	14. Introduction to the minitrack on data warehousing***	0	0	0
	15. Lineage tracing in a data warehousing system	0	0	0
	16. Materialized views in data warehousing environments	1	0	1
	17. Multidatabase language requirements	0	3	0
	18. Multidatabase management in Pegasus	0	5	0
	19. Partitioning schemes for object oriented databases	2	0	0
	20. Proceedings. The Computer Security Foundations Workshop III	0	0	0
	21. Reliable transaction execution in multidatabase systems	0	3	0
	22. Rethinking integrity [distributed databases]	2	0	1
	23. RQL a recursive query language	0	0	0
	24. Storage efficient and secure replicated distributed databases	3	0	0

ตาราง ก.1 (ต่อ) ผลการจัดกลุ่มบทคัดย่อ

กลุ่ม	ชื่อบทคัดย่อ	คำสำคัญ		
9		distribute	multidatabase	technique
(ต่อ)	25. Strategic planning for data warehousing in the public sector	0	0	0
	26. Subqueries in SQLf, a fuzzy database query language	0	0	0
	27. Supporting Web-based database application development	0	0	0
	28. The CenSSIS image database	0	0	0
	29. The Pegasus heterogeneous multidatabase system	1	3	0
	30. The Raid distributed database system	5	0	0
	31. User authentication in multidatabase systems	0	4	0
	32. User interface tools for object-oriented database systems	0	0	0
10		design	image	search
	1. A Characterization of Visual Appearance Applied to Image Retrieval	1	1	0
	2. A measure of recognition difficulty for a character image database	0	5	0
	3. An image speech relational database and its application	0	4	1
	4. Japan image database for utilization	0	9	1
	5. Origins of Scaling in Natural Images	0	4	0
	6. The effect of expressions in a database of face images***	0	6	0
	7. The Sowerby Image Database	1	6	1

ภาคผนวก ข

เนื้อหาของบทคัดย่อทางวิชาการ

แฟ้มข้อมูลเอกสารบทคัดย่อทางวิชาการที่เกี่ยวกับฐานข้อมูลจำนวนทั้งหมด 100 บทคัดย่อที่ได้นำมาใช้ในการทดสอบโปรแกรม มีเนื้อหาดังตารางที่ ข.1 ถึงตารางที่ ข.10 โดยแสดงตามกลุ่มที่ได้จากการทดสอบโปรแกรมในภาคผนวก ก ดังต่อไปนี้

ตาราง ข.1 เนื้อหาของบทคัดย่อในกลุ่มที่ 1

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
1	An extensible object-oriented database testbed	The authors describe the object-oriented design and implementation of an extensible schema manager for object-oriented databases. The open class hierarchy approach has been adopted to achieve the extensibility of the implementation. In this approach. the system meta information is implemented as objects of system classes. A graphical interface for an object-oriented database scheme environment, GOOSE, has been developed. GOOSE supports several advanced features which include schema evolution, schema versioning, and DAG (direct acyclic graph) rearrangement view of a class hierarchy. Schema evolution is the ability to make a variety of changes to a database scheme without reorganization. Schema versioning is the ability to define multiple scheme versions and to keep track of schema changes. A novel type of view for object-oriented databases, the DAG rearrangement view of a class hierarchy, is also supported.
2	An object-oriented prototype for a geophysical database	Database management systems (DBMSs) are being used in a wide variety of domains to handle many types of data. Scientific data pose a special challenge to DBMSs due to their volume and complex nature. The object-oriented model has many additional capabilities over the relational model that enables the database designer to capture the semantics of the data better. The paper describes the process of re-engineering a part of a relational database to an

ตาราง ข.1 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 1

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
2 (ต่อ)	An object-oriented prototype for a geophysical database	object-oriented database. The transformation of the relational schema to an object-oriented schema, the implementation of the new schema using an object-oriented DBMS, and a user interface for the new database are described.
3	Disk management for object-oriented databases	An object-oriented database provides persistent storage for a large number of objects. These objects may be very small, and the access patterns are likely to be not as uniform as the mostly sequential reads and writes seen in file-systems. For example, the 007 benchmark for object-oriented databases specifies a number of traversals that follow pointers around a graph of objects. Given these differences between file-systems and object-oriented databases, disk management techniques used in file-systems will not perform well if naively applied to object-oriented databases. This paper proposes three disk management strategies for object-oriented databases. These strategies are based on earlier work on file-systems. They differ from this earlier work in their support for a large number of small objects and non-sequential access patterns.
4	Method-induced partitioning schemes for object-oriented databases	Object-oriented database systems are becoming popular and are being used in a large number of application domain. Many of these application domains are inherently distributed. The focus of this work is on articulating the concepts of method induced partitioning schemes in object-oriented databases by understanding and classifying the object behavior embodied by the methods. We provide a solution for supporting fragmentation transparency by using method transformation. Finally, we present guidelines for method induced partitioning in object-oriented databases.
5	Serializability in object-oriented database systems	In an object-oriented database the objects are encapsulated, that is, objects are only accessible by methods defined in the database system. The definition of object-oriented serializability takes advantage of the semantics and of the nesting of the methods. Therefore, a lower rate of conflicting accesses than with the conventional definition of serializability is achieved. Transactions of an object-oriented database are defined as open-nested transactions. Depending on the semantics of operations, actions can be serialized independently of the calling transactions. The techniques already used in multilayer transaction systems are extended to object-oriented systems.

ตาราง ข.2 เนื้อหาของบทคัดย่อในกลุ่มที่ 2

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
1	A Tool For World-Wide Web Access Log Analysis	Various programs have emerged that provide statistical analysis of World-Wide Web (WWW) access logs. These programs typically detail the number of accesses for a file, the number of times a site has visited the database, and some programs even provide temporal analysis of requests. However, these programs are not interactive nor do they Provide visualizations of the local database. WebViz was developed with the intention of providing WWW database maintainers and designers with a graphical view of their local database and access patterns. That is, by incorporating the Web-Path paradigm into interactive software, users can see not only the documents (represented visually as nodes) in their database but also the hyperlinks travelled (represented visually as links) by users requesting documents from the database. WebViz further enables uses to selectively filter the access log (i.e. restrict the graphical view by specifying the desired domain names or DSN numbers, directory names.
2	Mobile agents for World Wide Web distributed database access	The popularity of the Web as a universal access mechanism for network information has created the need for developing Web-based DBMS client/server applications. However, the current commercial applet-based approaches for accessing database systems offer limited flexibility, scalability, and robustness. We propose a new framework for Web-based distributed access to database systems based on Java-based mobile agents. The framework supports lightweight, portable, and autonomous clients as well as operation on slow or expensive networks. The implementation of the framework using the aglet workbench shows that its performance is comparable to, and in some case outperforms, the current approach. In fact, in wireless and dial-up environments and for average size transactions, a client/agent/server adaptation of the framework provides a performance improvement of approximately a factor of ten. For the fixed network, the gains are about 40 percent and 30 percent, respectively. We expect our framework to perform even better when deployed using different implementation platforms as indicated by our preliminary results from an implementation based on Voyager.

ตาราง ข.2 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 2

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
3	Data mining for Web intelligence	Searching, comprehending, and using the semistructured HTML, XML, and database-service-engine information stored on the Web poses a significant challenge. This data is more sophisticated and dynamic than the information commercial database systems store. To supplement keyword-based indexing, researchers have applied data mining to Web-page ranking. In this context, data mining helps Web search engines find high-quality Web pages and enhances Web click stream analysis. For the Web to reach its full potential, however, we must improve its services, make it more comprehensible, and increase its usability. As researchers continue to develop data mining techniques, the authors believe this technology will play an increasingly important role in meeting the challenges of developing the intelligent Web. Ultimately, data mining for Web intelligence will make the Web a richer, friendlier, and more intelligent resource that we can all share and explore. The paper considers how data mining holds the key to uncovering and cataloging the authoritative links, traversal patterns, and semantic structures that will bring intelligence and direction to our Web interactions.
4	Performance modelling and metrics of database-backed Web sites	Currently, Web database systems are widely used to construct Web sites for their excellent capability of providing online information. The authors analyse the workflow of a Web database system dealing with a Web page request. We have classified the different cases and given an approximate method to model a Web database system. As the service time of database servers is a primary factor in determining the input characteristics of Web servers, it is very important to investigate the relationship between database servers and Web servers. The performance metrics are introduced on the basis of the analysis of the relationship.
5	Supporting dynamic interactions among Web-based information sources	The ubiquity of the World Wide Web offers an ideal opportunity for the deployment of highly distributed applications. Now that connectivity is no longer an issue, attention has turned to providing a middleware infrastructure that will sustain data sharing among Web-accessible databases. We present a dynamic architecture and system for describing, locating, and accessing data from Web-accessible databases. We propose the use of flexible organizational

ตาราง ข.2 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 2

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
5 (ต่อ)	Supporting dynamic interactions among Web-based information sources	constructs service links and coalitions to facilitate data organization, discovery, and sharing among Internet-accessible databases. A language is also proposed to support the definition and manipulation of these constructs. The implementation combines Java, CORBA, database API (JDBC), agent, and database technologies to support a scalable and portable architecture interconnecting large networks of heterogeneous and autonomous databases. We report on an experiment to provide uniform access to a Web of healthcare-related databases.
6	Web Database and Its Applications in Teaching Database	The paper considers the issues of Web database technology and its application in teaching database classes. It discusses the usefulness of Web databases in business and in teaching. The discussion on the implementation of a Web database with Java programming language is also given. The ease of learning Java, its reliability, and its portability make this programming language a convenient tool to implement a Web database. As an assistant in classroom teaching, Web databases have the flexibility needed to develop classroom demonstrations and Internet enabled client-server applications. The objective of this paper is to discuss the implementation of Web databases in classroom teaching.

ตาราง ข.3 เนื้อหาของบทคัดย่อในกลุ่มที่ 3

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
1	Data and applications security: developments and directions	This paper first describes the developments in data and applications security with a special emphasis on database security. Then it discusses direction for data and applications security which includes secure semantic web, XML security, and security for emerging applications such as bioinformatics, peer-to-peer computing, and stream information management.
2	Identifying and representing the security semantics of an application	The author approaches database security from the semantic level. He identifies the need to classify outputs from multilevel secure database systems at a level which accurately reflects the contents. Specifically, he addresses the question of what really makes information classified, that is, the security semantics of an application. A multidimensional taxonomy of generic secrecy constraints is presented with examples that illustrate application-specific security semantics.

ตาราง ข.3 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 3

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
2 (ต่อ)	Identifying and representing the security semantics of an application	Using labels to represent security semantics is shown to be ambiguous and therefore ineffective. Representing security semantics external to the database is proposed and several approaches are discussed. Finally, the use of a semantic data model, on top of a logic-based representation, is proposed to explicitly represent the security semantics of an application.
3	Multilevel database security using information clouding	The author employs fuzzy sets in a multilevel model for general-purpose database security. Sensitive information in database relations is meaningfully clouded by fuzzy sets. This is accomplished by broadening the possibility distributions constraining the values of sensitive attributes. The technique promotes the use of data and also maintains database security. Clouding with fuzzy sets is the middle ground between information release and information hiding/falsification. It nicely supplements the two security techniques and helps strike the right balance between user convenience and database security.
4	Multilevel secure databases: a new approach	An examination is given of the relationship between the commercial INGRES relational database management system (DBMS) and the Unix operating system with a view toward determining its applicability to supporting the requirements for a class B1 system as defined in the trusted computer system evaluation criteria. The authors propose a security policy that provides database security in a multilevel environment and aims at a secure DBMS through the retrofit of security controls into the existing system. They also detail some of the fundamental security requirements and functions of INGRES that can be exploited to enhance its security features.
5	Proceedings of the 1988 IEEE Symposium on Security and Privacy	The following topics are dealt with: formal security models; security in distributed systems; emerging issues; database security; analysis of secure systems; applying and implementing integrity models; verification; and models of secure distributed systems. Abstracts of individual papers can be found under the relevant classification codes in this or other issues.
6	Providing security in a phone book database using triggers	It is proposed that all database security controls except those which provide information flow security can be built using a suitable trigger mechanism. The implementation of an example application, which has a variety of requirements for confidentially, integrity and accountability, is shown to illustrate the

ตาราง ข.3 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 3

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
6 (ต่อ)	Providing security in a phone book database using triggers	technique. The trigger mechanism is to be implemented using query modification as part of the SWORD secure database management system project.
7	Security model consistency in secure object-oriented systems	Examines three techniques for evaluating the logical consistency of an object-oriented Database security model. The first technique consists of judging the model with respect to a set of general consistency properties for database security models. The second technique compares the SODA model against two other database security models. The third technique consists of defining a set of entities and mechanisms fundamental to the object-oriented model and considering the effect on them by the security model. Each of these techniques are applied to the Secure Object-Oriented Database (SODA) security model and are evaluated with respect to their applicability difficulty and usefulness. Using the results of this analysis the authors characterize the consistency of the SODA security model.

ตาราง ข.4 เนื้อหาของบทคัดย่อในกลุ่มที่ 4

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
1	Data mining: an industrial research perspective	Just what exactly is data mining? At a broad level, it is the process by which accurate and previously unknown information is extracted from large volumes of data. This information should be in a form that can be understood, acted upon, and used for improving decision processes. Obviously, with this definition, data mining encompasses a broad set of technologies, including data warehousing, database management, data analysis algorithms, and visualization. The crux of the appeal for this new technology lies in the data analysis algorithms, since they provide automated mechanisms for sifting through data and extracting useful information. The analysis capability of these algorithms, coupled with today's data warehousing and database management technology, make corporate and industrial data mining possible. The data representation model for such algorithms is quite straightforward. Data is considered to be a collection of records, where each record is a collection of fields. Using this tabular data model, data mining algorithms are designed to

ตาราง ข.4 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 4

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
1 (ต่อ)	Data mining: an industrial research perspective	operate on the contents, under differing assumptions, and delivering results in differing formats. The data analysis algorithms (or data mining algorithms, as they are more popularly known nowadays) can be divided into three major categories based on the nature of their information extraction: predictive modeling (also called classification or supervised learning), clustering (also called segmentation or unsupervised learning), and frequent pattern extraction
2	Data types generalization for data mining algorithms	With the increasing use of database applications, mining interesting information from huge databases becomes of great concern and a variety of mining algorithms have been proposed in recent years. As we know, the data processed in data mining may be obtained from many sources in which different data types may be used. However, no algorithm can be applied to all applications due to the difficulty of fitting data types to the algorithm. The selection of an appropriate data mining algorithm is based not only on the goal of the application, but also the data fittability. Therefore, to transform the non-fitting data type into a target one is also important in data mining, but the work is often tedious or complex since a lot of data types exist in the real world. Merging the similar data types of a given selected mining algorithm into a generalized data type seems to be a good approach to reduce the transformation complexity. In this work, the data type fittability problem for six kinds of widely used data mining techniques is discussed and a data type generalization process, including merging and transforming phases is proposed. In the merging phase, the original data types of the data sources to be mined are first merged into the generalized ones. The transforming phase is then used to convert the generalized data types into the target ones for the selected mining algorithm. Using the data type generalization process, the user can select an appropriate mining algorithm just for the goal of the application without considering the data types

ตาราง ข.5 เนื้อหาของบทคัดย่อในกลุ่มที่ 5

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
1	A fuzzy search method for rough sets in data mining	This paper proposes a technique that combines a fuzzy search method called ?-connected search and rough sets, in data mining. ?-connected searching was originally proposed to search seismic layers in seismic data processing. Although ?-connected searching is designed for digital spaces, or numerical data analysis, it can be used for any domain, as long as the domain can be described by a graph. ?-connectedness is an equivalence relation, therefore all searched components form a partition of the base domain. Rough sets, a new methodology in data mining, is based on a classification R on a base set U (the universal set). Then, any subset of U can be represented by an approximation based on the union of certain classes with respect to R. For data processing, U usually is a digital space. The value of each point is often a vector of real/rational numbers. Base domain classification is the key to a rough set system. Theoretically, any equivalence relation R can be defined by a ?-connected classification. In order to use the concept of ?-connectedness in rough sets and data mining, this paper proposes a limited multi-level ?-connected search. In addition, some properties of rough sets using ?-connectedness and their applications to data mining are investigated.
2	A primer for understanding and applying data mining	Data mining is such a hot topic that it has become an obscured buzzword. Data mining can be a powerful tool for extracting useful information from tons of data. But it can just as easily extract erroneous and useless information if it's not used correctly. Key to avoiding the pitfalls is a basic understanding of what data mining is and what things to consider in planning a data mining project. The steps in a data mining project include: integrating and cleaning or modifying the data sources, mining the data, examining and pruning the mining results, and reporting the final results
3	Data mining applications in BT	Presents a number of case studies demonstrating how data mining is being used within BT to discover valuable knowledge. The case studies highlight the use of these techniques, the wealth of information contained in databases and learning points encountered. The studies presented are in the following areas: (1) identifying faults on printed circuit boards; (2) discovering the organisational structure of groups of criminals; and (3) predicting outcomes of

ตาราง ข.5 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 5

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
3 (ต่อ)	Data mining applications in BT	credit assessment and litigation. Data mining techniques have been shown to provide significant benefits in terms of cost savings and detection of fraud against the company. What has also been shown, particularly in the case of fraud, is that the combination of data mining, data visualisation and human expertise is highly effective. A number of lessons have been learnt from these studies. First, simply throwing a machine learning system at a database is unlikely to yield good results. A significant amount of effort is required to pre-process data and understand its meaning in the problem domain Specialist domain knowledge will almost certainly be required. Second, a good deal of problem simplification is likely to be needed if high-accuracy results are to be obtained. This inevitably requires an element of compromise between overall business goals and what is practically achievable. Lastly, data mining alone will not yield business benefits. To be successful, it is necessary that business processes are changed to deliver them. The mind-set which views data as something to be archived has to be changed to one which views it as a valuable resource.
4	Autonomous decision-making: a data mining approach	The researchers and practitioners of today create models, algorithms, functions, and other constructs defined in abstract spaces. The research of the future will likely be data driven. Symbolic and numeric data that are becoming available in large volumes will define the need for new data analysis techniques and tools. Data mining is an emerging area of computational intelligence that offers new theories, techniques, and tools for analysis of large data sets. In this paper, a novel approach for autonomous decision-making is developed based on the rough set theory of data mining. The approach has been tested on a medical data set for patients with lung abnormalities referred to as solitary pulmonary nodules (SPNs). The two independent algorithms developed in this paper either generate an accurate diagnosis or make no decision. The methodology discussed in the paper depart from the developments in data mining as well as current medical literature, thus creating a variable approach for autonomous decision-making.

ตาราง ข.5 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 5

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
5	Data mining for short-term load forecasting	Short-term load forecasting plays a key role in power system operation and planning. This paper presents a method for data mining for short-term load forecasting in power systems. This paper makes use of a data mining method to clarify the nonlinear relationship between input and output variables in short-term load forecasting. Data mining discovers useful knowledge and rules in large data bases. Data mining is more attractive because of difficulty in understanding large data bases. The obtained model structure explains the importance of input variables. It may be classified into the classification and the regression trees. This paper handles the regression tree since load forecasting corresponds to the quantitative problem. This paper presents three strategies: hybrid model of CART and multi-layer perceptron (MLP); optimal structure with Tabu search; and fuzzy data mining.
6	Data mining for the enterprise	The emergence of comprehensive data warehouses which integrate operational data with customer, supplier and market data have resulted in an explosion of information. Competition requires timely and sophisticated analysis on an integrated view of that data. However, there has been a growing gap between more powerful data warehousing systems and the users' ability to effectively analyze and act on the information they contain. Data mining tools and services are providing the leap necessary to close this gap. Data mining offers automated discovery of previously unknown patterns as well as automated prediction of trends and behaviors; its technologies are complimentary to existing decision support tools and provide the business analyst and marketing professional with a new way of analyzing the business. After a general introduction of the knowledge discovery lifecycle and the data mining lifecycle, this article examines the data mining issues and requirements within an enterprise. A comprehensive architectural overview proposes data mining integration solutions for data warehouses, application servers, thick clients, and thin clients. This article concludes with an analysis of current trends relevant to enterprise usage of data mining tools and methodologies.

ตาราง ข.5 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 5

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
7	Data mining: making data meaningful	Research groups, large businesses, government agencies and other organizations are using improved data mining technologies and techniques to discover meaningful patterns in huge databases, and now, data mining has been refined to the point where even people who aren't highly trained statisticians can use this complex data analysis tool. Data mining's increased popularity is due partly to technological improvements that permit faster, more effective analyses of databases.
8	Data mining of printed-circuit board defects	This paper discusses an industrial case study in which data mining has been applied to solve a quality engineering problem in electronics assembly. During the assembly process, solder balls occur underneath some components of printed circuit boards. The goal is to identify the cause of solder defects in a circuit board using a data mining approach. Statistical process control and design of experiment approaches did not provide conclusive results. The paper discusses features considered in the study, data collected, and the data mining solution approach to identify causes of quality faults in an industrial application
9	Data mining problems in medicine	The principle of any retrospective on patient data-based investigation is searching the patients by problem or sign, but not by name. With a proper problem-encoded archival database, the data mining process would be easy. One would only need to input the request and obtain the proper data in a short time. Medical archives are frequently based on paper records only, with the patient name as the entry key. To find the proper record in such an archive, a detection strategy is needed. The process continues with collecting the usually enormous amount of papers, finding the appropriate records within them, and finally encoding and arranging them in a table. The whole process can be separated into patients, paper and data mining. Because of their slowness, these phases can be the most time-consuming part of a medical data-based investigation. The author describes his data mining experience.
10	Data mining with an ant colony optimization algorithm	The paper proposes an algorithm for data mining called Ant-Miner (ant-colony-based data miner). The goal of Ant-Miner is to extract classification rules from data. The algorithm is inspired by both research on the behavior of real ant colonies and some data mining concepts as well as principles. We compare the

ตาราง ข.5 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 5

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
10 (ต่อ)	Data mining with an ant colony optimization algorithm	performance of Ant-Miner with CN2, a well-known data mining algorithm for classification, in six public domain data sets. The results provide evidence that: 1) Ant-Miner is competitive with CN2 with respect to predictive accuracy, and 2) the rule lists discovered by Ant-Miner are considerably simpler (smaller) than those discovered by CN2
11	Data mining with Clementine	Data mining is the extraction of useful information or knowledge from bodies of data. Data mining is also sometimes referred to as knowledge discovery in databases. Clementine is a comprehensive, integrated toolkit which provides active support for data mining in the form of neural network and rule induction learning techniques, passive support in the form of visualisation, statistical and browsing facilities, and peripheral support for data access and manipulation. Clementine's visual programming interface provides an environment which is easy to use for technological experts and non-experts alike, and provides a convenient organising framework for any data mining technique. Clementine makes machine learning accessible to non-experts. Clementine also goes beyond simple organisational advantages because it reduces the time taken to perform data mining experiments by one to two orders of magnitude. This means that experiments which would previously have been impractical in any realistic project are made possible by Clementine, effectively opening up new areas of exploration. The first commercial version of Clementine is currently on the market and is attracting a great deal of interest. Future versions will extend the machine learning facilities and provide many other new features. The utility of Clementine will continue to grow, and to provide new possibilities for data mining
12	Database technology for decision support systems	Decision support systems form the core of business IT infrastructures because they let companies translate business information into tangible and lucrative results. Collecting, maintaining, and analyzing large amounts of data, however, involves expensive technical challenges that require organizational commitment. Many commercial tools are available for each of the three major data warehousing tasks: populating the data warehouse from independent operational databases, storing and managing the data, and analyzing the data to

ตาราง ข.5 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 5

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
12 (ต่อ)	Database technology for decision support systems	make intelligent business decisions. Data cleaning relates to heterogeneous data integration, a problem studied for many years. More work must be done to develop domain-independent tools that solve the data cleaning problems associated with data warehouse development. Most data mining research has focused on developing algorithms for building more accurate models or building models faster. However, data preparation and mining model deployment present several engaging problems that relate specifically to achieving better synergy between database systems and data mining technology
13	Decomposition in data mining: an industrial case study	Data mining offers tools for discovery of relationships, patterns, and knowledge in large databases. The knowledge extraction process is computationally complex and therefore a subset of all data Is normally considered for mining. In this paper, numerous methods for decomposition of data sets are discussed. Decomposition enhances the quality of knowledge extracted from large databases by simplification of the data mining task. The ideas presented are illustrated with examples and an industrial case study. In the case study reported in this paper, a data mining approach is applied to extract knowledge from a data set. The extracted knowledge is used for the prediction and prevention of manufacturing faults in wafers
14	Graph-based data mining	Using databases represented as graphs, the Subdue system performs two key data mining techniques: unsupervised pattern discovery and supervised concept learning from examples. Applications to large structural databases demonstrate Subdue's scalability and effectiveness
15	Mining very large databases	Established companies have had decades to accumulate masses of data about their customers, suppliers, products and services, and employees. Data mining, also known as knowledge discovery in databases, gives organizations the tools to sift through these vast data stores to find the trends, patterns, and correlations that can guide strategic decision making. Traditionally, algorithms for data analysis assume that the input data contains relatively few records. Current databases however, are much too large to be held in main memory. To be efficient, the data mining techniques applied to very large databases must be highly scalable. An algorithm is said to be scalable if (given a fixed amount of

ตาราง ข.5 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 5

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
15 (ต่อ)	Mining very large databases	main memory), its runtime increases linearly with the number of records in the input database. Recent work has focused on scaling data mining algorithms to very large data sets. The authors describe a broad range of algorithms that address three classical data mining problems: market basket analysis, clustering, and classification

ตาราง ข.6 เนื้อหาของบทคัดย่อในกลุ่มที่ 6

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
1	A query algebra for program databases	Querying source code is an essential aspect of a variety of software engineering tasks such as program understanding, reverse engineering, program structure analysis and program flow analysis. In this paper, we present and demonstrate the use of an algebraic source code query technique that blends expressive power with query compactness. The query framework of Source Code Algebra (SCA) permits users to express complex source code queries and views as algebraic expressions. Queries are expressed on an extensible, object-oriented database that stores program source code. The SCA algebraic approach offers multiple benefits such as an applicative query language, high expressive power, seamless handling of structural and flow information, clean formalism and potential for query optimization. We present a case study where SCA expressions are used to query a program in terms of program organization, resource flow, control flow, metrics and syntactic structure. Our experience with an SCA-based prototype query processor indicates that an algebraic approach to source code queries combines the benefits of expressive power and compact query formulation
2	Associative query answering via query feature similarity	Associative query answering provides additional relevant information to the queries that is not explicitly asked, but is of interest to the user. For a given query, associative information may be derived from past user query cases based on the user type and the query context. A case-based reasoning approach that matches query features is proposed. The query feature consists of the query topic, the output attribute list, and the selection constraints. The similarity of the query feature is defined and can be evaluated from the semantic model that is derived from the database schema. Query feature is presented

ตาราง ข.6 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 6

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
3	Dynamic query re-optimization	Very long running queries in database systems are not uncommon in non traditional application domains such as image processing or data warehousing analysis. Query optimization, therefore, is important. However, estimates of the query characteristics before query execution are usually inaccurate. Further, system configuration and resource availability may change during long evaluation period. As a result, queries are often evaluated with sub-optimal plan configurations. To remedy this situation, we have designed a novel approach to re-optimize suboptimal query plan configurations on-the-fly with Conquest, an extensible and distributed query processing system. A dynamic optimizer considers reconfiguration cost as well as execution cost in determining the best query plan configuration. Experimental results are presented
4	Query folding	Query folding refers to the activity of determining if and how a query can be answered using a given set of resources, which might be materialized views, cached results of previous queries, or queries answerable by other databases. We investigate query folding in the context where queries and resources are conjunctive queries. We develop an exponential time algorithm that finds all complete or partial foldings, and a polynomial time algorithm for the subclass of acyclic conjunctive queries. Our results can be applied to query optimization in centralized databases, to query processing in distributed databases, and to query answering in federated databases
5	Query planning with limited source capabilities	In information-integration systems, sources may have diverse and limited query capabilities. We show that because sources have restrictions on retrieving their information, sources not mentioned in a query can contribute to the query result by providing useful bindings. In some cases we can access sources repeatedly to retrieve bindings to answer a query, and query planning thus becomes considerably more challenging. We find all the obtainable answers to a query by translating the query and source descriptions to a simple recursive Datalog program, and evaluating the program on the source relations. This program often accesses sources that are not in the query. Some of these accesses are essential, as they provide bindings that let us query sources, which we could

ตาราง ข.6 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 6

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
5 (ต่อ)	Query planning with limited source capabilities	not do otherwise. However, some of these accesses can be proven not to add anything to the query's answer. We show in which cases these off-query accesses are useless, and prove that in these cases we can compute the complete answer to the query by using only the sources in the query. In the cases where off-query accesses are necessary, we propose an algorithm for finding all the useful sources for a query. We thus solve the optimization problem of eliminating the unnecessary source accesses, and optimize the program to answer the query
6	Querying source code using an algebraic query language	Querying and analyzing source code interactively is a critical task in reverse engineering and program understanding. Current source code query systems lack sufficient formalism and offer limited query capabilities. We introduce the formal framework of Source Code Algebra (SCA), and outline a source code query system based on it. SCA provides a formal data model for source code, an algebraic expression-based query language, and opportunities for query optimization. An algebraic model of source code addresses the issues of conceptual integrity, expressive power, and performance of a source code query system within a unified framework
7	The Kendall Square Query Decomposer	The Kendall Square Query Decomposer works in conjunction with the underlying ORACLE7 relational database management system to greatly speed the execution of decision support queries. Based on the data access strategy selected for a query by the ORACLE query optimizer, the Query Decomposer generates a number of subqueries to match the underlying physical data partitions of a table on disk. These subqueries are submitted in parallel to ORACLE over multiple connections. The subqueries are executed in parallel, each accessing only its own part of the driving table. Finally, the Query Decomposer assembles subquery results and returns the full result
8	Visual feedback in querying large databases	In this paper, we describe a database query system that provides visual relevance feedback in querying large databases. The goal of our system is to support the query specification process by using each pixel of the display to represent one data item of the database. By arranging and coloring the pixels according to their relevance for the query, the user gets a visual impression of

ตาราง ข.6 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 6

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
8 (ต่อ)	Visual feedback in querying large databases	the resulting data set. Using sliders for each condition of the query, the user may change the query dynamically and receives immediate feedback by the visual representation of the resulting data set. By using multiple windows for different parts of a complex query, the user gets visual feedback for each part of the query and, therefore, will easier understand the overall result. The system may be used to query any database that contains tens of thousands to millions of data items, but it is especially helpful to explore large data sets with an unknown distribution of values and to find the interesting hot spots in huge amounts of data. The direct feedback allows to visually display the influence of incremental query refinements and, therefore, allows a better, easier and faster query specification

ตาราง ข.7 เนื้อหาของบทคัดย่อในกลุ่มที่ 7

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
1	A new dynamic voting algorithm for distributed database systems	We consider the problem of keeping a distributed database system that has been partitioned because of site or communication link failures partially operable while ensuring data consistency. A dynamic-voting-consistency algorithm is proposed, and its correctness is demonstrated. The proposed algorithm results in improved efficiency in executing read requests by not requiring a read quorum. This algorithm is effective in environments where the majority of user requests are “read” types of requests. Furthermore, the proposed algorithm results in efficient recovery by avoiding updating those data objects that are still current. Under the proposed algorithm, the majority partition would be available even if changes in the network topology take place at a higher rate than the update rate, as long as simple partitioning takes place
2	An architecture for data warehouse systems	The growing demand for decision support tools has resulted in data warehousing as an important database research area. We propose an architecture for data warehouse systems. The architecture involves client/server properties and deductive database features. It can provide good properties of generality, extendibility, efficiency, scalability, and intelligence. We detail information flow in warehouse systems and functions of warehouse components. We also review and highlight the research problems in this area

ตาราง ข.7 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 7

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
3	Competitive analysis of caching in distributed databases	This paper makes two contributions. First, we introduce a model for evaluating the performance of data allocation and replication algorithms in distributed databases. The model is comprehensive in the sense that it accounts for I/O cost, for communication cost, and, because of reliability considerations, for limits on the minimum number of copies of the object. The model captures existing replica-management algorithms, such as read-one-write-all, quorum-consensus, etc. These algorithms are static in the sense that, in the absence of failures, the copies of each object are allocated to a fixed set of processors. In modern distributed databases, particularly in mobile computing environments, processors will dynamically store objects in their local database and will relinquish them. Therefore, as a second contribution of this paper, we introduce an algorithm for automatic dynamic allocation of replicas to processors. Then, using the new model, we compare the performance of the traditional read-one-write-all static allocation algorithm to the performance of the dynamic allocation algorithm. As a result, we obtain the relationship between the communication cost and I/O cost for which static allocation is superior to dynamic allocation, and the relationships for which dynamic allocation is superior
4	Consistency in data warehouse dimensions	Data warehouses present a powerful framework for storing and analyzing huge amounts of data. In this context analyses focus on data that has been gathered over long periods of time, often between one and five years. A data warehouse can therefore be regarded as a specialized historical database. However not only the data kept in a data warehouse has to be seen in a temporal context, but also the fact that dimension data may undergo changes during such a time period needs to be taken into consideration. In this paper we focus on update operations on dimensions and establish a notion of consistency for guiding such operations. We devise algorithms for executing update operations that can be shown to preserve consistency, and we study their time complexity.
5	Data warehouse for EIS: some issues and impacts	The data warehouse is one of the most rapidly growing areas in management information systems. With this approach, data for EIS and DSS applications is separated from operational data and stored in a separate database called a data

ตาราง ข.7 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 7

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
5 (ต่อ)	Data warehouse for EIS: some issues and impacts	warehouse. Some of the advantages of this approach are improved performance, better data quality, and the ability to consolidate and summarize data from heterogeneous legacy systems. A data warehouse is part of a larger infrastructure that includes legacy data sources, external data sources, a repository, data acquisition software, and user interface and related analytical tools. A powerful form of data analysis, called multidimensional data analysis, is often performed by users of a data warehouse. Data warehouses can be organized into two-tier or three-tier client/server systems. Despite the complexity of the data warehouse environment little academic research has been performed in this area. This paper identifies a number of issues that arise in the context of developing and using a data warehouse. It develops a proposed research model to determine the impact of factors such as organizational factors, warehouse infrastructure, and management support on user satisfaction and development characteristics of a data warehouse
6	Data warehouse tools	Many large Government software projects are designed and implemented as if they were satisfying requirements never seen before. There is a common tendency to start “from a blank sheet of paper”. This is often because software engineers view projects in terms of specialized mission-critical applications that appear to have unique requirements. The common result is the constant reinvention of existing software and the rejection of sophisticated tools as not applicable to mission requirements that appear unique. The Oracle data warehouse approach has made a clear break from this expensive, application-oriented approach. In this new analytical environment, there is wide-spread recognition that existing application data is a potential information source for existing commercial off-the-shelf (COTS) analytical tools whose value has been established by previous successful implementations. Critical legacy data can be accessed and transformed into a data warehouse using COTS development tools, the data COTS warehouse can be accessed by user-friendly analytical tools, and the data warehouse maintained by proven COTS administration tools

ตาราง ข.7 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 7

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
7	Efficient maintenance of temporal data warehouses	A temporal data warehouse can be defined as a set V of materialized views over non-temporal sources. We present a self-maintainable temporal data warehouse that, besides a set of temporal views V, has two kinds of auxiliary views: a set T of auxiliary relations containing only temporal information necessary to maintain the data warehouse and a set C of complements allowing the warehouse maintenance without consulting the source databases. Several features of our approach make it especially attractive as a maintenance method for warehouses: the entire history of source databases is not stored in the warehouse; the warehouse keeps enough information for its self-maintenance; and the warehouse maintenance is achieved by executing only relational algebra operators . Although temporal views in V are specified in a temporal algebraic query language we translate them into relational algebra expressions containing only auxiliary relations in T and complements in C. We claim that our method which extends some new techniques, can simplify considerably the implementation of a temporal data warehouse
8	Evaluating data warehouse toolkits	Historically, building a data warehouse has been a risky proposition. According to the Gartner Group and similar sources, the failure rate for data warehouse projects runs as high as 60 percent. The dilemma facing businesses today is this: although implementing an enterprise-wide database is risky, an organization that does so gains a significant business advantage over competitors who do not have data warehouse capability or who continue to rely on localized, group-specific data marts.
9	Evolving materialized views in data warehouse	A data warehouse contains multiple views accessed by queries. One of the most important decisions in designing a data warehouse is the selection of materialized views for the purpose of efficiently implementing decision making. The search space for the selection of materialized views is exponentially large, therefore, heuristics have been used to search a small fraction of the space to get a near optimal solution. In this paper, we explore the use of a genetic algorithm for the selection of materialized views based on multiple global processing plans for many queries. Our experimental studies indicate that the genetic algorithm delivers better solutions than some heuristics

ตาราง ข.7 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 7

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
10	Modeling a faster data warehouse	One key area involved in implementing the data warehouse is the transfer of information from the Source Databases (SDBs) into the data warehouse and making it available to the user. In looking at the overall intent of the data warehouse as a repository for vast amounts of diverse information, changes made to the SDBs, can be critical to the correctness of the query responses provided by the warehouse. The paper presents an alternative to the state of the art data warehouse environment by providing an auxiliary structure to accept these updates and store them until the data warehouse has available cycle time to incorporate the new data and update the necessary views. Using the Integrated Definition (IDEF) modeling process, we have created models of both the state of the art and alternative data warehouse structures to validate our design and performance characteristics
11	Multidatabase performance evaluation	A model of a multidatabase transaction management system is presented, and the performance of two concurrency control algorithms operating in such an environment is analyzed. One algorithm does not impose any restrictions on either the structure of the concurrency control mechanisms used by local database management systems (DBMSs) or the type of multidatabase transactions. The other algorithm assumes that each local DBMS uses the two-phase locking protocol. The performance results presented demonstrate that in both cases the concurrent processing of global transactions provides a better throughput than their serial execution. However, the first algorithm may cause a significant number of global transaction rollbacks for some combinations of local and global transactions. For the second algorithm it is shown that the number of global transaction rollbacks is quite small for reasonable multiprogramming levels and that it performs almost as well as a distributed homogeneous database system
12	Practical lineage tracing in data warehouses	We consider the view data lineage problem in a warehousing environment: for a given data item in a materialized warehouse view, we want to identify the set of source data items that produced the view item. We formalize the problem and we present a lineage tracing algorithm for relational views with aggregation. Based on our tracing algorithm, we propose a number of schemes

ตาราง ข.7 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 7

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
12 (ต่อ)	Practical lineage tracing in data warehouses	for storing auxiliary views that enable consistent and efficient lineage tracing in a multi-source data warehouse. We report on a performance study of the various schemes, identifying which schemes perform best in which settings. Based on our results, we have implemented a lineage tracing package in the WHIPS data warehousing system prototype at Stanford. With this package, users can select view tuples of interest, then efficiently “drill through” to examine the exact source tuples that produced the view tuples of interest

ตาราง ข.8 เนื้อหาของบทคัดย่อในกลุ่มที่ 8

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
1	Database recovery using redundant disk arrays	The authors propose a method for using redundant disk arrays to support rapid recovery from system crashes and transaction aborts in addition to their role in providing media failure recovery. A twin-page scheme is used to store the parity information in the array, making it possible to keep the old version of the parity along with the new version. The old version of the parity is used to undo updates performed by aborted transactions or by transactions interrupted by a system failure. Using an analytical model, it is shown that the proposed method achieves a significant increase in the throughput of database systems using redundant disk arrays by reducing the number of recovery operations needed to maintain the consistency of the database
2	Incremental recovery in main memory database systems	Recovery activities, like checkpointing and restart, in traditional database management systems are performed in a quiescent state where no transactions are active. This approach impairs the performance of online transaction processing systems, especially when a large volatile memory is used. An incremental scheme for performing recovery in main memory database systems (MMDBs), in parallel with transaction execution, is presented. A page-based incremental restart algorithm that enables the resumption of transaction processing as soon as the system is up is proposed. Pages are recovered individually and according to the demands of the post-crash transactions. A method for propagating updates from main memory to the backup database on disk is also provided. The emphasis is on decoupling the I/O activities related to the propagation to disk from the forward transaction execution in memory.

ตาราง ข.8 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 8

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
2 (ต่อ)	Incremental recovery in main memory database systems	The authors also construct a high-level recovery manager based on operation logging on top of the page-based algorithms. The proposed algorithms are motivated by the characteristics of large MMDBs, and exploit the technology of nonvolatile RAM
3	Main memory database recovery algorithms and their performance	In main memory systems recovery becomes more complex mainly due to the volatility of main memory and the elimination of the separation of data storage (disk) and data processing locations (main memory). In a disk based database system, it is the disk copy that represents the stable database restored after a system failure; but in MDBSs, the main memory copy of the database must be recovered. The paper discusses the recovery issues in MDBSs and presents a somewhat detailed performance study of two recovery algorithms, one using a shadow approach and the other update-in-place
4	Mobile user recovery in the context of Internet transactions	With the expansion of Web sites to include business functions, a user interfaces with e-businesses through an interactive and multistep process, which is often time-consuming. For mobile users accessing the Web over digital cellular networks, the failure of the wireless link, a frequent occurrence, can result in the loss of work accomplished prior to the disruption. This work must then be repeated upon subsequent reconnection - often at significant cost in time and computation. This "disconnection-reconnection-repeat work" cycle may cause mobile clients to incur substantial monetary as well as resource (such as battery power) costs. In this paper, we propose a protocol for "recovering" a user to an appropriate recent interaction state after such a failure. The objective is to minimize the amount of work that needs to be redone upon restart after failure. Whereas classical database recovery focuses on recovering the system, i.e., all transactions, our work considers the problem of recovering a particular user interaction with the system. This recovery problem encompasses several interesting subproblems: (1) modeling user interaction in a way that is useful for recovery, (2) characterizing a user's "recovery state", (3) determining the state to which a user should be recovered, and (4) defining a recovery mechanism. We describe the user interaction with one or more Web sites using intuitive and familiar concepts from database transactions. We call this

ตาราง ข.8 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 8

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
4 (ต่อ)	Mobile user recovery in the context of Internet transactions	interaction an Internet transaction (iTX), distinguish this notion from extant transaction models, and develop a model for it, as well as for a user's state on a Web site. Based on the twin foundations of our iTX and state models, we finally describe an effective protocol for recovering users to valid states in Internet interactions
5	Persistent applications using generalized redo recovery	We describe how to recover applications after system crashes using database recovery. Earlier efforts, based on frequent application checkpoints and/or logging values read, are very expensive. We treat application state as a cached object and log application execution as operations in the recovery framework of D. Lomet and M. Tuttle (1995). Logging application execution does not require logging the application state. Further logged application reads are mostly logical operations in which only the data source identity is logged. We describe a cache manager that handles the flush order dependencies introduced by these log operations and a recovery process that restores application state by replaying the application
6	Recovery guarantees for general multi-tier applications	Database recovery does not mask failures to applications and users. Recovery is needed that considers data, messages and application components. Special cases have been studied, but clear principles for recovery guarantees in general multi-tier applications such as Web-based e-services are missing. We develop a framework for recovery guarantees that masks almost all failures. The main concept is an interaction contract between two components, a pledge as to message and state persistence, and contract release. Contracts are composed into system-wide agreements so that a set of components is provably recoverable with exactly-once message delivery and execution, except perhaps for crash-interrupted user input or output. Our implementation techniques reduce the data logging cost, allow effective log truncation, and provide independent recovery for critical server components. Interaction contracts form the basis for our Phoenix/COM project on persistent components. Our framework's utility is demonstrated with a case study of a web-based e-service

ตาราง ข.9 เนื้อหาของบทคัดย่อในกลุ่มที่ 9

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
1	A type system for an object-oriented database system	A type system is presented with strong typing and static type checking that is not yet well supported in most existing object-oriented database systems. A subtyping rule is defined for correctly structuring the inheritance hierarchy of the types. Based on the subtyping, a number of type inference rules are defined. They can be used by the type system to statically determine the types of the query results and whether a given application program is type correct
2	A uniform model for temporal object-oriented databases	A temporal object-oriented model and query language that supports the modeling and manipulation of complex temporal or versioned objects is developed. The authors show that the approach not only provides a richer model than the relational for capturing the semantics of complex temporal objects, but also requires no special constructs in the query language. Consequently, the retrieval of temporal and non-temporal information is uniformly expressed. By allowing variables and quantifiers to range over time, can be formulated that require special operators in other languages. Temporal aggregation queries, which are not easily expressed in other models, are expressed using the same aggregation operators as for nontemporal data
3	Allocating data and operations to nodes in distributed database design	The allocation of data and operations to nodes in a computer communications network is a critical issue in distributed database design. An efficient distributed database design must trade off performance and cost among retrieval and update activities at the various nodes. It must consider the concurrency control mechanism used as well as capacity constraints at nodes and on links in the network. It must determine where data will be allocated, the degree of data replication, which copy of the data will be used for each retrieval activity, and where operations such as select, project, join, and union will be performed. We develop a comprehensive mathematical modeling approach for this problem. The approach first generates units of data to be allocated from a logical data model representation and a characterization of retrieval and update activities. Retrieval and update activities are then decomposed into relational operations on these fragments. Both fragments and operations on them are then allocated to nodes using a mathematical modeling approach.

ตาราง ข.9 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 9

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
4	A World Wide Web Resource Discovery System	As the WWW grows at an increasing rate, efforts to make the technology more manageable are highly in demand. Applying advanced information retrieval techniques is one approach to such efforts. Despite the potential benefit of these techniques in reducing users information overload and improving the effectiveness access to on-line information, little research has been done on applying them to WWW environment. In this paper we present our attempt to apply the vector space retrieval model, relevance feedback mechanisms and a hypertext mapping technique as an integrated resource discovery system to the WWW. This paper discusses some design issues involved, as well as practical issues such as retrieval effectiveness, usability and scalability.
5	An algebra and calculus for relational multidatabase systems	With the existence of many autonomous databases widely accessible through computer networks, users will require the capability to jointly manipulate data in different databases. A multidatabase system provides such a capability through a multidatabase manipulation language. The authors propose a theoretical foundation for such languages by presenting a multirelational algebra and calculus based on the relational algebra and calculus. The proposal is illustrated by various queries on an example multidatabase.
6	Authorization and revocation in object-oriented databases	Few studies of object-oriented databases deal with their security, a fundamental aspect of systems with complex data structures. Most authorization systems give users who own resources only some basic control over them; here, we provide users with more direct control over their resources by associating with each grant propagation numbers. Propagation numbers govern the grantability and exercisability of the privileges. Of particular interest in our study of authorization in an OO environment is the combination of inheritance and granting of privileges. Diverse policies are discussed and implemented in a test-bed system
7	Automating compensation in a multidatabase	Compensation is the process by which a committed transaction in a database is undone by running the semantic inverse of that transaction on the database. Compensation has been proposed as a technique for undoing committed work in various situations where strict atomicity cannot be maintained (H. Garcia-Molina, K. Salem, 1987; J.G. Mullen, 1991). We discuss compensation in long-

ตาราง ข.9 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 9

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
7 (ต่อ)	Automating compensation in a multidatabase	running multidatabase transactions. We define the step approach to integrating local database schemas into a multidatabase. In the step approach, each local database is encapsulated by a set of procedures (steps). Steps can be grouped into atomic global transactions. Each step also has an associated compensating step, which is called if the compensating transaction is run. We examine two areas of multidatabase transaction management where compensation is required. The first is implementing compensation as a recovery technique when an open, nested transaction is aborted. The second is in backing out the effects of an atomic mutlidatabase transaction when some local database transaction commits before a global abort decision is made
8	Clustering techniques for object-oriented database systems	The authors conduct an analysis to compare three clustering strategies, Cactis, ORION, and CK, in terms of space and time overhead. They also examine the level, page, or segment at which clustering should take place. The dynamic clustering technique CK is found to be best in exploiting the structural relationships between objects and inheritance semantics to identify an efficient storage scheme. However, it creates high overhead and is best suited for applications in which the read/write ratio is high. To remove this limitation, the authors show how segment clustering, instead of page clustering, could reduce the number of cases where a page split is necessary
9	Data warehousing and the Internet's impact on ERP	Advances in data warehousing and Internet technology have spurred ERP development, but these systems still need some improvements. The information era has dramatically changed the business environment, and keen competition puts tremendous demands on companies to reduce total cost and maximize return on investment. This is why more businesses are seeing systems for enterprise resource planning (ERP) as key to remaining competitive. To keep pace, ERP systems themselves have undergone considerable change in the past 10 years (1990-2000). To help your organization make the right ERP choices, it is helpful to know how two key technologies: data warehousing and the Internet, are transforming these complex systems

ตาราง ข.9 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 9

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
10	Executing multidatabase transactions	In a multidatabase environment, the traditional transaction model has been found to be too restrictive. Therefore, several extended transaction models have been proposed in which some of the requirements of transaction, such as isolation or atomicity, are optional. The authors describe one of such extensions, the flexible transaction model and discuss the scheduling of transactions involving multiple autonomous database systems managed by heterogeneous DBMS. The scheduling algorithm for flexible transactions is implemented using L.0, a logically parallel language which provides a framework for concisely specifying the multidatabase transactions and for scheduling them. The key aspects of a flexible transaction specification, such as subtransaction execution dependencies and transaction success criteria, can be naturally represented in L.0. Furthermore, scheduling in L.0 achieves maximal parallelism allowed by the specifications of transactions, which results in the improvement of their response times. To provide access to multiple heterogeneous hardware and software systems, they use the Distributed Operation Language (DOL). DOL approach is based on providing a common communication and data exchange protocol and uses local access managers to protect the autonomy of member software systems. When L.0 determines that a subtransaction is ready to execute, it hands it through an interface to the DOL system for execution. The interface between L.0 and DOL provides the former with the execution status of subtransactions
11	Healthcare data warehousing and quality assurance	Healthcare data warehousing presents unique challenges. The industry is rife with often incompatible medical standards and coding schemes that require careful translation. Healthcare data comes from many sources and is delivered in many forms, including published books, individual spreadsheets, and several tape or data formats. Results derived from a healthcare data warehouse must be delivered in accessible form to diverse stakeholders, including healthcare regulators, physicians, hospital administrators, consumers, community activists, and members of the popular press. The industry's widely decentralized and largely autonomous data collection efforts make data quality a significant challenge. Finally, the sensitivity of healthcare data makes privacy

ตาราง ข.9 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 9

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
11 (ต่อ)	Healthcare data warehousing and quality assurance	and security issues paramount. Healthcare data warehousing will make rigorous, quantitative information available to healthcare decision makers. The authors describe a fully functional healthcare data warehouse used to produce several reports for communities throughout Florida. Building on this work, they're actively pursuing a research agenda to enhance technical data warehousing capabilities while investigating innovative community and clinical healthcare applications
12	Image retrieval by examples	A currently important research field in information sciences is the management of nontraditional distributed multimedia databases. Two related key issues are to achieve an efficient content-based query by example retrieval and a fast response time. This paper presents the architecture of a distributed image retrieval system which provides novel solutions to these key issues. In particular, a method to quantify the effectiveness of low level visual descriptors in database query tasks is presented. The results are also used to improve the system response time, which is an important issue when querying very large databases. A new mechanism to adapt system query strategies to user behavior is also introduced in order to improve the effectiveness of relevance feedback and overall system response time. Finally, the issue of browsing multiple distributed databases is considered and a solution is proposed using multidimensional scaling techniques
13	Implementation of interoperability in large multidatabases	Sharing information among databases has traditionally been at the data level. In a large network of databases and with the current technology, direct data sharing, i.e. sharing actual data in one single step, is no longer tractable or even possible. Another layer of data sharing has to be supplied before direct data sharing can take place. The size issue is a major factor in complicating issues of autonomy and heterogeneity. In this new environment, users need to be incrementally and dynamically informed about available information and where it is located. A static approach to user education cannot be reasonable in such a large environment. This approach relies on building a new layer that is responsible for organizing databases in suitable and flexible conglomerations. This layer is implemented using special-purpose databases (co-databases).

ตาราง ข.9 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 9

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
14	Introduction to the minitrack on data warehousing	Data warehousing continues to grow in importance. Surveys consistently rank it as one of the most important strategic initiatives in companies today. Customer relationship management, performance management, e-commerce, and connecting companies along the supply chain all rely on data warehousing. Data warehousing has moved from being an innovation in a few leading-edge companies, to being a requirement for doing business in a large number of firms.
15	Lineage tracing in a data warehousing system	Some commercial data warehousing systems support schema-level lineage tracing, or provide specialized drill-down and/or drill-through facilities for multi-dimensional warehouse views. Our lineage tracing system supports more fine-grained instance-level lineage tracing for arbitrarily complex relational views, including aggregation. At view definition time, our system automatically generates lineage tracing procedures and supporting auxiliary views. At lineage tracing time, the system applies the tracing procedures to the source tables and/or auxiliary views to obtain the lineage results and to illustrate the specific view data derivation process
16	Materialized views in data warehousing environments	We discuss the use of materialized views in the construction and maintenance of data warehousing environments. Although materializing views represents the most effective technique to improve OLAP (on-line analytical processing) query performance, it introduces particular challenges into these environments: (i) identification of which views to materialize, considering the trade-off among performance, scalability and view maintenance restrictions; and (ii) incremental maintenance of these views. In addition to analyzing and comparing some current proposals aimed at solving these two challenges, we also identify restrictions in the surveyed proposals to a Web distributed data warehousing environment, in which the warehouse data are fragmented, replicated and allocated through several sites. Some suggestions are made to overcome such restrictions
17	Multidatabase language requirements	The integration transparency of a multidatabase system depends heavily on the diversity of local language interfaces and models. These parameters are not predetermined but application dependent. The flexibility of a new concept

ตาราง ข.9 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 9

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
17 (ต่อ)	Multidatabase language requirements	called megaprogramming can help to solve the multidatabase problem. The author evaluates this idea based on logic programming and discusses how such a language can serve simultaneously as global query language, for transaction specification and execution, and for the representation of static integration information
18	Multidatabase management in Pegasus	Pegasus is an object-oriented multidatabase system being developed at Hewlett-Packard Laboratories. The goal of the system is to provide facilities for multidatabase applications for accessing and manipulating multiple autonomous heterogeneous object-oriented, relational and other databases. This paper gives an overview of the architecture, describes the common object model and gives the justification for adapting the model for a multidatabase system. It describes the database environment of the system, discusses how to manage schemas consisting of types and functions from multiple databases. It briefly describes the features of the data language and gives an example. It discusses a method for dealing with object identification issues in a multidatabase system, reviews schema integration issues and briefly mentions schema mapping to relational models
19	Partitioning schemes for object oriented databases	In order to support homogeneous distributed object oriented database systems we need to have a clear understanding of what is meant by partitioning a class and what are the different ways of doing so. In this paper, we concentrate an studying the different types of class partitioning schemes that can arise in object oriented databases. By laying down the foundation, by articulating the concepts, representation, and implementation approaches for partitioning object databases, we facilitate further work on partitioning algorithms, query decomposition, optimization and transaction management for distributed OODBMS
20	Proceedings. The Computer Security Foundations Workshop III	The following topics are dealt with: logic and protocol analysis; models of information flow; information flow in abstract machines; modeling new properties; covert channel detection; and database security

ตาราง ข.9 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 9

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
21	Reliable transaction execution in multidatabase systems	The issue of executing concurrent transactions reliably in a multidatabase system is addressed. The authors' formalization assumes the most pessimistic scenario where individual database managers that make up the multidatabase system are totally autonomous. The authors present algorithms that ensure transaction atomicity and crash recovery. The atomicity algorithms are based on the emulation of the well-known two-phase commit protocol. It is applied among autonomous participants that neither know of each other's existence nor how to communicate
22	Rethinking integrity [distributed databases]	In practice, no real database enjoys complete integrity, a variety of factors conspire to make some information stale, missing, or just plain wrong. System design and analysis should recognize that real systems lack integrity to some degree, most, if not all, of the time. Significant benefits flow from such recognition. Risk management techniques can identify the severity of different integrity-loss scenarios, thereby focusing scarce resources on critical areas. A designer can deliberately sacrifice nonessential integrity under carefully controlled conditions to achieve other design objectives, such as performance, autonomy, availability, or security. Designers can achieve these objectives and still preserve essential aspects of integrity. The authors discuss the factors that undermine integrity in distributed databases
23	RQL: a recursive query language	Different classes of recursive queries in the relational databases are identified. It is shown that existing proposals to extend the relational query languages are either not powerful enough to express queries in many of these classes or use nonfirst normal form constructs. RQL, a recursive database query language that can be used to express recursive queries on all the classes identified, is presented. RQL is based on the relational algebra. In addition to functions that correspond to the standard and extended relational algebra operators, RQL supports functions required to express general recursive queries. The elements of RQL and the ways in which they are used to formulate complicated, but useful, recursive queries are described. The effects of the extensions embodied in RQL on the termination of recursive query evaluation are discussed

ตาราง ข.9 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 9

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
24	Storage efficient and secure replicated distributed databases	Data availability and security are two important issues in a distributed database system. Existing schemes achieve high availability at the expense of higher storage cost and data security at the expense of higher processing cost. We develop an integrated methodology that combines the features of some existing schemes dealing with data fragmentation, data encoding, partial replication, and quorum consensus concepts to achieve storage efficient, highly available, and secure distributed database systems
25	Strategic planning for data warehousing in the public sector	The City of St. Louis, Missouri, is a large municipality with an extremely fragmented organizational structure. The City recently undertook a major strategic planning study for development of its municipal information systems. In the course of the study, data warehousing emerged as a strategy for consolidating and sharing information among many City departments with different informational needs and a variety of computing platforms. We describe processes and analytical methods that proved useful for shaping a data warehousing strategy and for determining the contents of the data warehouse. We also present some cautionary conclusions about the development of monolithic data warehouses for decision support
26	Subqueries in SQLf, a fuzzy database query language	This paper is mainly concerned with the extension of database management systems querying capabilities, so that users may address queries involving preferences and get discriminated answers. The use of flexible (gradual) predicates interpreted in the framework of the fuzzy set theory is advocated for defining a query language, called SQLf. SQLf enlarges the functionalities offered by SQL and the various roles of subqueries are presented from both semantic and query processing points of view
27	Supporting Web-based database application development	This paper discusses our experiences of designing and implementing a pure Java database proxy server with a JDBC compatible driver for intranet/Internet database application development. In particular, we present a shared database connection strategy and flexible caching facilities to address the scalability problem. Web clients with the same access privilege can maintain their logical connections with shared physical connections to the database. Thus, not only a large number of users can be entertained by limited physical resource, the

ตาราง ข.9 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 9

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
27 (ต่อ)	Supporting Web-based database application development	connection cost is also not indulged for each individual client. Web clients can also express their cache requirements explicitly in the JDBC protocol so that a large number of clients can be served with improved response time. The effectiveness of such strategies is demonstrated through a set of experiments
28	The CenSSIS image database	The CenSSIS image database is a scientific database that enables effective data management and collaboration to accelerate fundamental research. This paper describes the design and use of a state-of-the-art relational image database management system, accessible through a standard Web-browser interface. The application utilizes a robust security architecture and is designed for efficient data submission. Our database query engine provides complex query capabilities to facilitate fast and efficient data retrieval. The system offers a highly extensible metadata schema, with the option of storing data within a hierarchical format.
29	The Pegasus heterogeneous multidatabase system	Pegasus, a heterogeneous multidatabase management system that responds to the need for effective access and management of shared data across in a wide range of applications, is described. Pegasus provides facilities for multidatabase applications to access and manipulate multipole autonomous heterogeneous distributed object-oriented relational, and other information systems through a uniform interface. It is a complete data management system that integrates various native and local databases. Pegasus takes advantage of object-oriented data modeling and programming capabilities. It uses both type and function abstractions to deal with mapping and integration problems. Function implementation can be defined in an underlying database language or a programming language. Data abstraction and encapsulation facilities in the Pegasus object model provide an extensible framework for dealing with various kinds of heterogeneities in the traditional database systems and nontraditional data sources
30	The Raid distributed database system	Raid, a robust and adaptable distributed database system for transaction processing, is described. Raid is a message-passing system, with server processes on each site. The servers manage concurrent processing, consistent replicated copies during site failures and atomic distributed commitment. A

ตาราง ข.9 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 9

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
30 (ต่อ)	The Raid distributed database system	high-level, layered communications package provides a clean, location-independent interface between servers. The latest design of the communications package delivers messages via shared memory in a high-performance configuration in which several servers are linked into a single process. Raid provides the infrastructure to experimentally investigate various methods for supporting reliable distributed transaction processing. Measurements on transaction processing time and server CPU time are presented. Data and conclusions of experiments in three categories are also presented: communications software, consistent replicated copy control during site failures, and concurrent distributed checkpointing. A software tool for the evaluation of transaction processing algorithms in an operating system kernel is proposed
31	User authentication in multidatabase systems	The aspect of security needs more consideration in the area of architectures for multidatabase systems. Particularly, the authentication of users which is a main prerequisite for a successful authorization is not considered sufficiently in current architectures. Due to the autonomy and heterogeneity of the component database systems, the problem of authentication in multidatabase systems is more complex than in traditional database systems. In this paper we discuss the foundations and prerequisites for architectures of authentication in multidatabase systems. We present several approaches with respect to different degrees of autonomy and heterogeneity.
32	User interface tools for object-oriented database systems	The authors describe an architecture for a user interface for an OODB which is general in nature, independent of variations in the details of an underlying object model, and capable of supporting a variety of user interface requirements. The prototype employs a default interface style based on hybrids of forms and control panels. Within this framework, a tool is built which allows users to declaratively define applications at a high-level of abstraction. Such user interface tools will be a requirement for OODBs in order to permit users to exploit their powerful data and computational models while hiding their complexity. It is also likely that such tools will permit users to specify application-specific user interface behaviour

ตาราง ข.10 เนื้อหาของบทคัดย่อในกลุ่มที่ 10

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
1	A Characterization of Visual Appearance Applied to Image Retrieval	A system to retrieve images using a description of visual appearance is presented. A multi-scale invariant vector representation is obtained by first filtering images in the database with Gaussian derivative filters at several scales and then computing low order differential invariants. The multi-scale representation is indexed for rapid retrieval. Queries are designed by the users from an example image by selecting appropriate regions. The invariant vectors corresponding to these regions are matched with those in the database both in feature space as well as in coordinate space and a match score is obtained for each image. The results are then displayed to the user sorted by the match score. From experiments conducted with over 1500 images it is shown that images similar in appearance and whose viewpoint is within 25 degrees of the query image can be retrieved with a very satisfactory average precision of 57.4%
2	A measure of recognition difficulty for a character image database	A character image database plays an important role not only in the development stage but also for the evaluation of a handwritten character recognition system. Such a database is obtained from outside or customly made. At this point there is no measure which tells the level of recognition difficulty of a given database. If such a measure is available, we can use it in many useful ways. Especially, it will be valuable when we compare and evaluate the performance results of various systems since different databases whose recognition difficulties are unknown are usually used. In this paper we propose such a measure. We first define the entropy of a point of an image in the database. Then we obtain the measure by applying some normalizing factors to the entropy mentioned above. Note that such a measure should be used to compare the recognition difficulties of databases only of the same character set. We show that the proposed measure can be used for databases not only of different numbers of images per class but also of different image sizes. Finally we confirm that the proposed measure really reflects the relative recognition difficulties of databases by using real databases

ตาราง ข.10 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 10

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
3	An image/speech relational database and its application	There are many reasons to investigate the simultaneous analysis of corresponding speech and image information. For example, in the case of video telephone/conferencing there is clearly a strong connection between the sound phonemes of voiced speech data and the mouth shape of the speaker. Also it is known that most verbal communications use cues from both the visual and acoustic modalities to convey messages. During the production of speech, the visible information provided by the external articulatory organs can influence the understanding of the language by interpreting the combined information into meaningful linguistic expressions. Although the belief is that only the hearing impaired make use of the visual stimuli in percepting the speech, reports have shown that normal hearing people use all the available visual information that accompany speech, especially when there is degradation of speech. Therefore the objective of this project is to investigate and quantify the relationship between speech and image data such that the knowledge gained will assist in longer term multimedia and videophone research. To achieve the above, a statistical database of key parameters derived from corresponding speech and image information channels is discussed
4	Japan image database for utilization	Using the NOAA-HRPT data received at Tohoku University, a satellite image database for researchers is called the Tohoku Image Database (TIDAS). TIDAS is open for the public, and can be accessed via the Internet. The albedo and brightness temperatures processed from the AVHRR channel 2 and channel 4 are geocoded. This database starts from April 1990. The area is Tohoku on the east side of Japan. From 1996, the database includes the west side of Japan called the Japan Image Database (JAIDAS) to the present. The image data are 1024/spl times/1024 pixel for every channel data (Ch1 - Ch5) as the original. And also GIF files are presented as 1/4 size images in the catalog. The output images from the database are on the increase. In 1994 there were 7,834, in 1996 42,776, in 1998 74,241 and in 2000 219,740 images. We counted over 520,000 in only a half period of 2001. The results mean requests for environmental studies and education. We developed JAIDAS into the Tohoku University NOAA Image Database for more contributions to Earth environmental studies.

ตาราง ข.10 (ต่อ) เนื้อหาของบทคัดย่อในกลุ่มที่ 10

ลำดับ	ชื่อบทคัดย่อ	เนื้อหา
5	Origins of Scaling in Natural Images	One of the most robust qualities of our visual world is the scaleinvariance of natural images. Not only has scaling been found in different visual environments, but the phenomenon also appears to be calibration independent. This paper proposes a simple property of natural images which explains this robustness: They are collages of regions corresponding to statistically independent "objects". Evidence is provided for these objects having a power-law distribution of sizes within images, from which follows scaling in natural images. It is commonly suggested that scaling instead results from edges, each with power spectrum 1/k^2. This hypothesis is refuted by example.
6	The effect of expressions in a database of face images	This paper present the results of some face recognition experiments in which 3 different databases are used, each database giving a different representation of the same 100 people. The first database consists of 4 images per person, where each image shows a blank facial expression. The second database also contains 4 images per person, but each image shows a different facial expression: blank, smile, angry, and surprised. The third database is a combination of the other two databases and consists of 8 images per person. It is found that the database consisting of 4 different facial expression gives better classification results than the database with the 4 blank prototypes, and performs only slightly worse than the much larger database of 8 prototypes per person
7	The Sowerby Image Database	The development of image processing/computer vision algorithms is often limited by the data available to the researcher, in terms of both quality and quantity. This paper describes an extensive image data base, the Sowerby Image Database (SID), which has been specifically designed to facilitate rigorous algorithm development and testing in the areas of image segmentation and image labelling. Details are given of the database content, the process by which it was constructed and some preliminary results from its use

ภาคผนวก ค

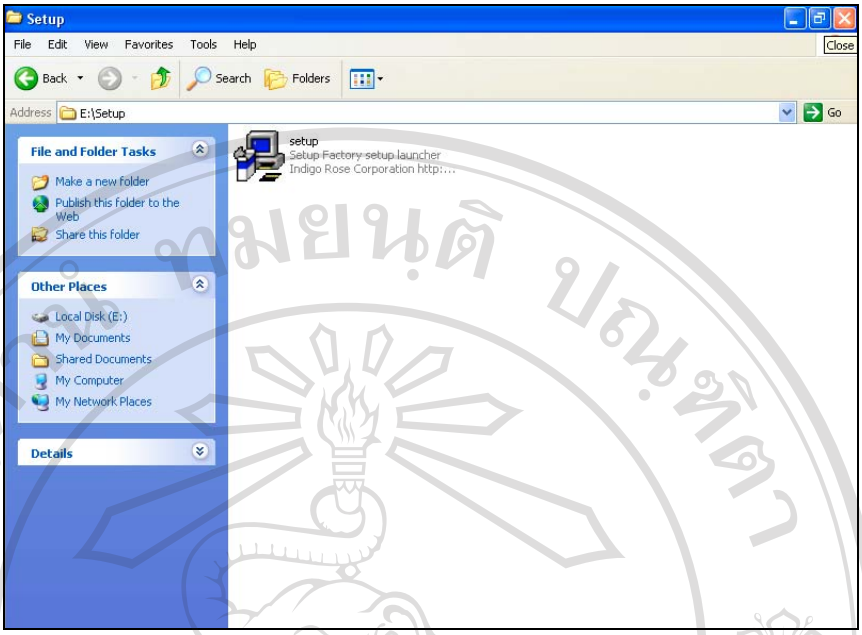
คู่มือการใช้โปรแกรม

คู่มือการใช้โปรแกรม เป็นส่วนที่อธิบายวิธีการใช้งานโปรแกรม เพื่อให้ผู้ใช้สามารถใช้งานโปรแกรมได้อย่างถูกต้อง และมีประสิทธิภาพมากที่สุด โดยรายละเอียดของคู่มือการใช้โปรแกรมจะแบ่งเป็น 2 ส่วน ได้แก่ การติดตั้งโปรแกรม และการใช้งานโปรแกรม ซึ่งมีรายละเอียดดังต่อไปนี้

ค.1 การติดตั้งโปรแกรม

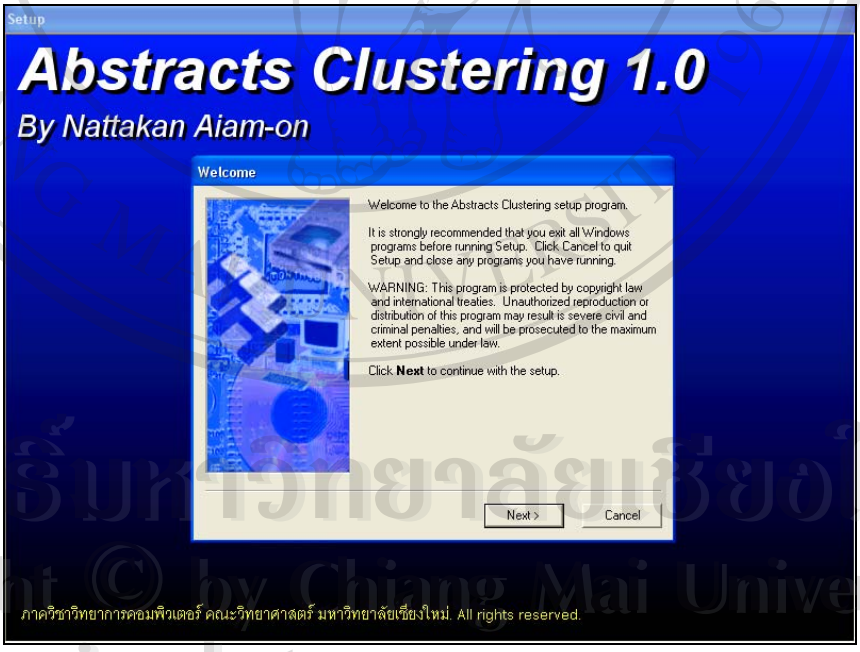
ในการติดตั้งโปรแกรมควรทำการติดตั้งโปรแกรมลงบนเครื่องคอมพิวเตอร์ส่วนบุคคลที่มีระบบปฏิบัติการไมโครซอฟต์วินโดว์ 98 (Microsoft Windows 98) หน่วยประมวลผลกลางเพนเทียม 2 ความเร็ว 400 เมกกะเฮิรตซ์ และมีหน่วยความจำ 64 เมกกะไบต์ เป็นอย่างน้อย เพื่อให้โปรแกรมสามารถทำงานได้อย่างมีประสิทธิภาพมากที่สุด ซึ่งการติดตั้งโปรแกรมมีขั้นตอนดังต่อไปนี้

1. ทำการเลือกที่ตัว Set up ดังรูปที่ ค.1



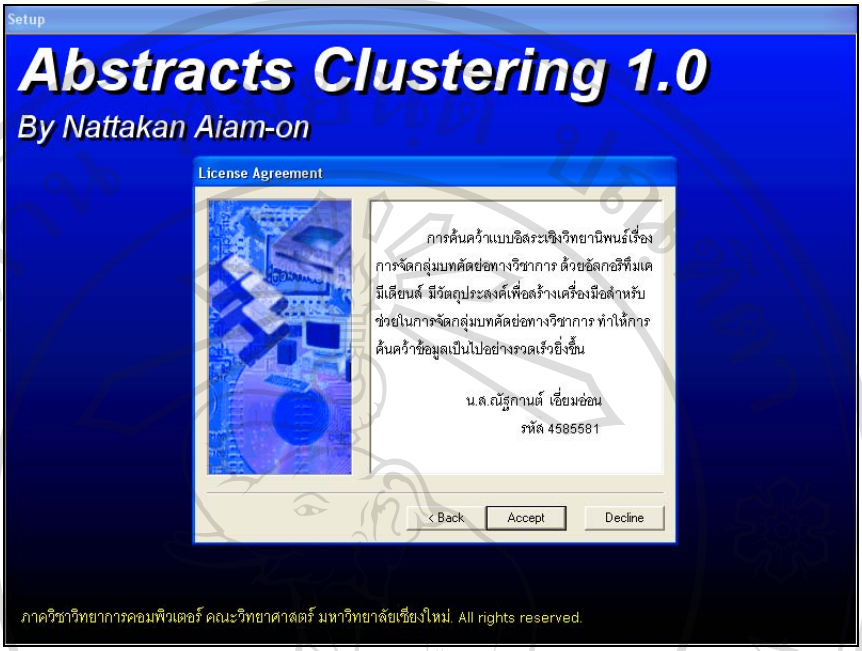
รูป ค.1 การเลือกตัวติดตั้งโปรแกรม

2. เมื่อเลือกตัว Set up แล้วจะปรากฏจอภาพ Welcome ให้ทำการคลิกที่ Next ดังรูปที่ ค.2



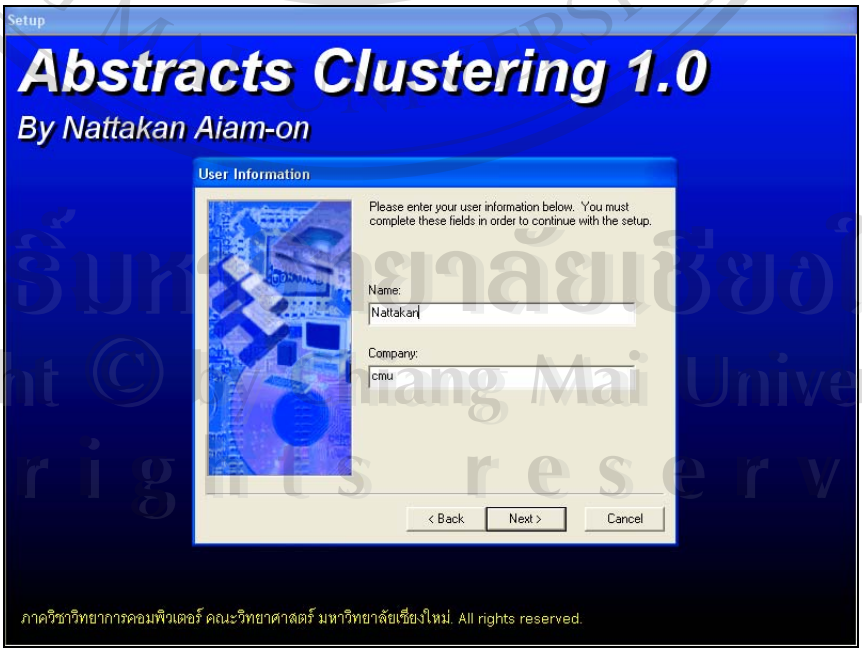
รูป ค.2 จอภาพ Welcome

3. จากนั้นจะปรากฏจอภาพ License Agreement แสดงข้อมูลเกี่ยวกับการค้นคว้าอิสระเชิงวิทยานิพนธ์ ดังรูปที่ ค.3



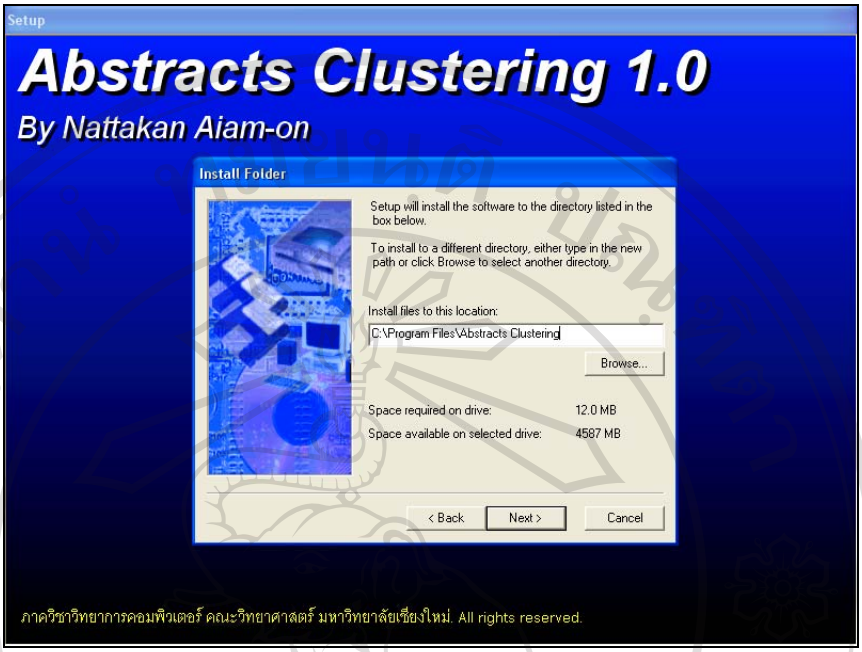
รูป ค.3 จอภาพ License Agreement

4. ทำการคลิก Accept จะปรากฏจอภาพให้ใส่ข้อมูล Name และ Company ซึ่งเมื่อทำการใส่ข้อมูลแล้วให้คลิก Next ดังรูปที่ ค.4



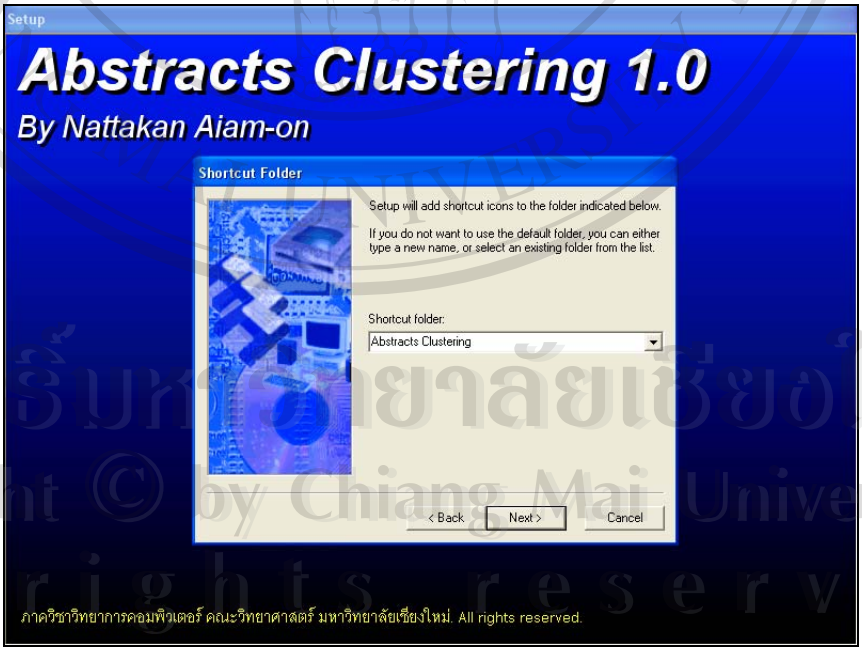
รูป ค.4 จอภาพให้ใส่ข้อมูลผู้ใช้

5. จากนั้นจะปรากฏจอภาพให้ระบุตำแหน่งที่ต้องการติดตั้ง เมื่อเลือกได้ให้คลิก Next ดังรูปที่ ค.5



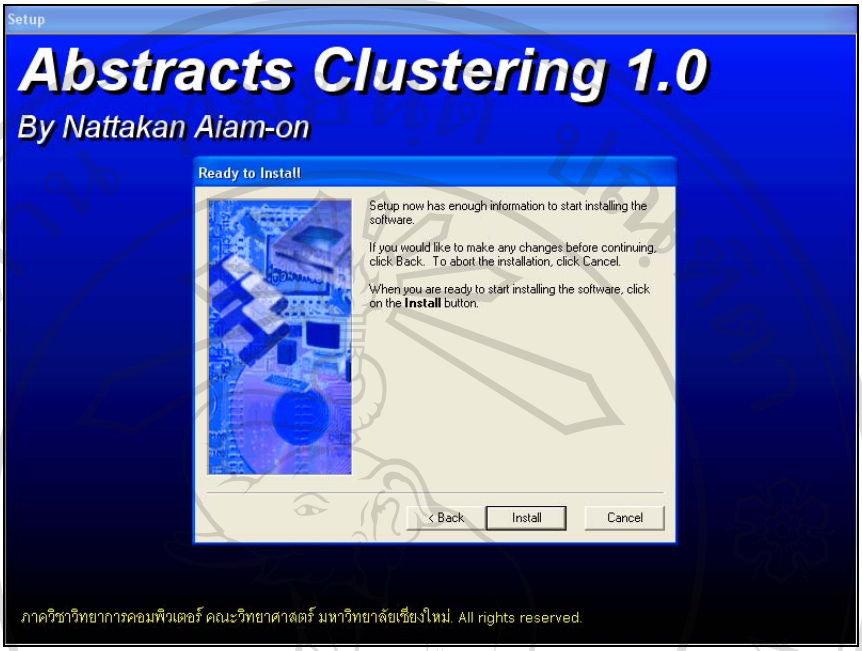
รูป ค.5 จอภาพเลือกตำแหน่งที่ต้องการติดตั้ง

6. จากนั้นจะปรากฏจอภาพทำการสร้าง Shortcut ดังรูปที่ ค.6



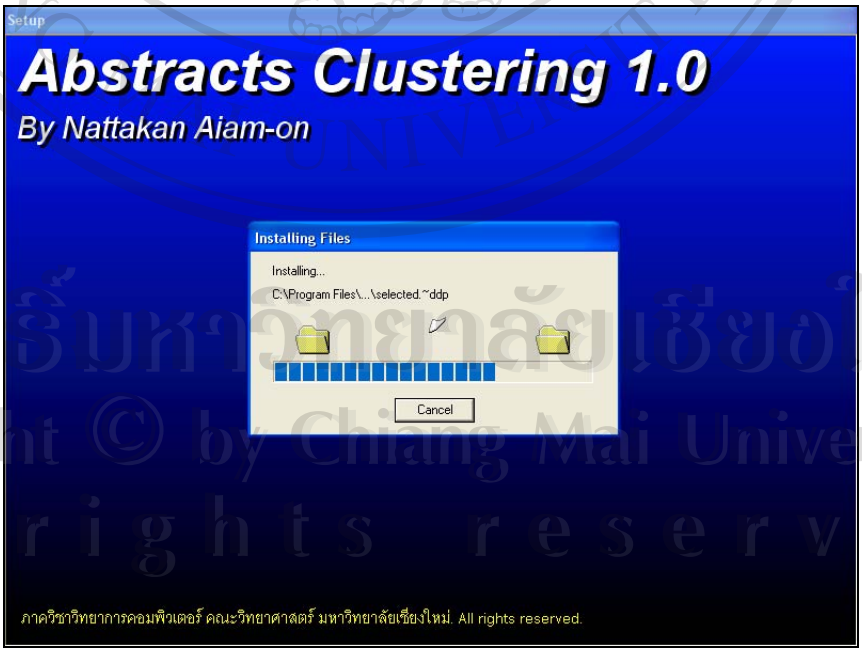
รูป ค.6 แสดงการจอภาพให้สร้าง Shortcut

7. จากนั้นจะปรากฏจอภาพให้ทำการติดตั้งโปรแกรม โดยคลิก Install เพื่อทำการติดตั้งโปรแกรม ดังรูปที่ ค.7



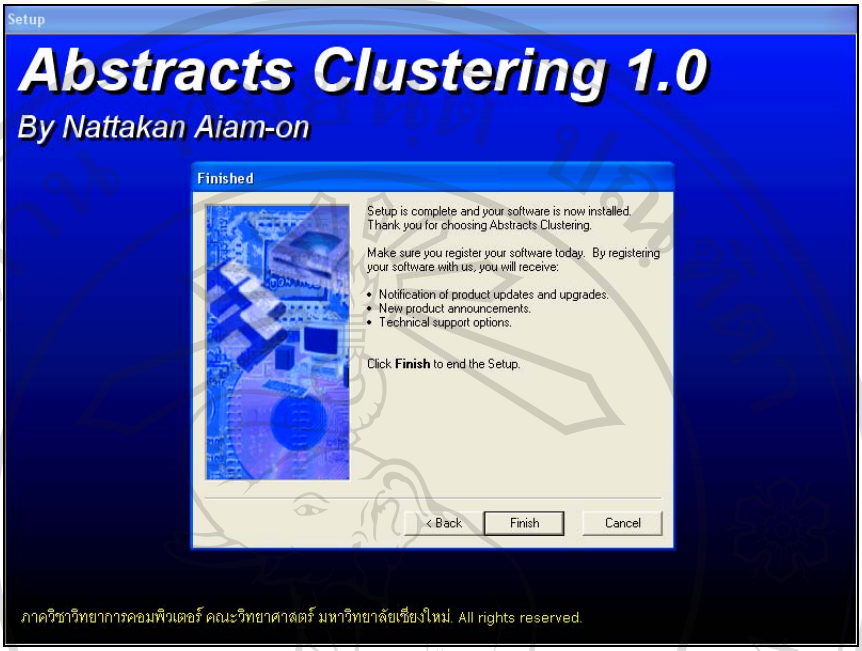
รูป ค.7 จอภาพให้ทำการติดตั้งโปรแกรม

8. โปรแกรมจะทำการติดตั้ง ดังรูปที่ ค.8



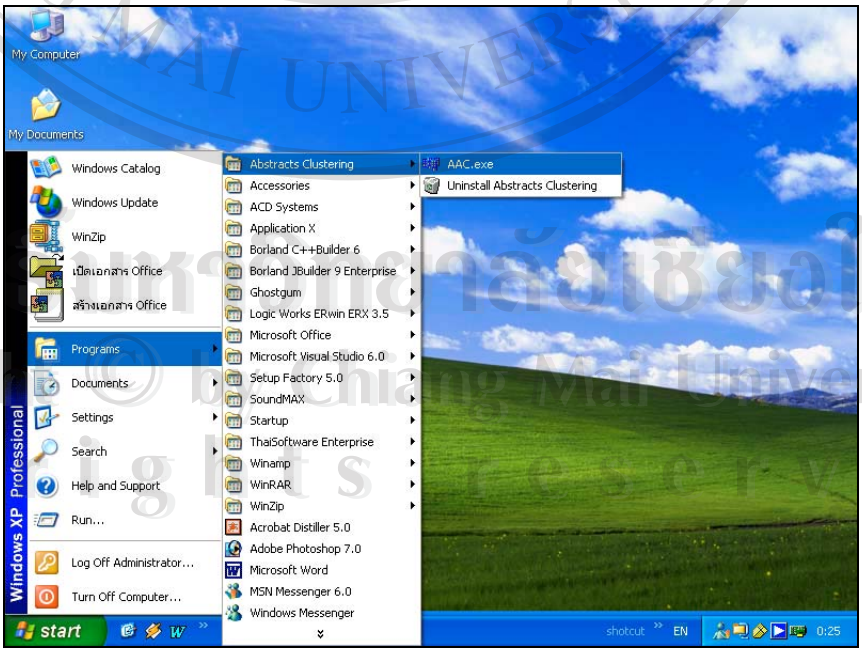
รูป ค.8 จอภาพแสดงการติดตั้งโปรแกรม

9. เมื่อทำการติดตั้งโปรแกรมเสร็จแล้ว จะปรากฏจอภาพเสร็จสิ้นการติดตั้งโปรแกรม ให้คลิก Finish ดังรูปที่ ค.9



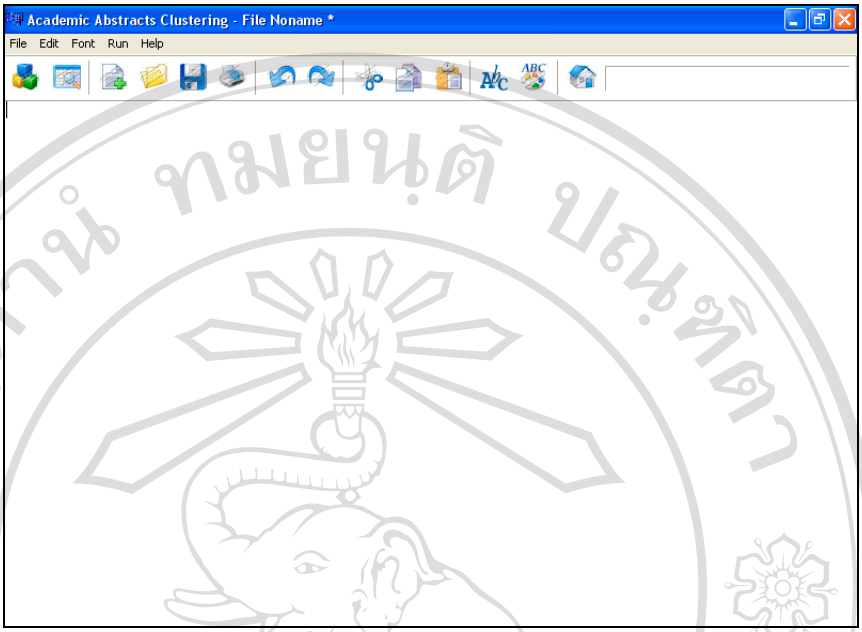
รูป ค.9 จอภาพเสร็จสิ้นการติดตั้งโปรแกรม

10. เมื่อทำการติดตั้งโปรแกรมเสร็จแล้ว สามารถทำการเรียกใช้โปรแกรมได้จาก Start menu จากนั้นเลือก Abstract Clustering ดังรูปที่ ค.10



รูป ค.10 จอภาพการเรียกใช้โปรแกรม

11. จากนั้นจะปรากฏจอภาพหลักของโปรแกรมเพื่อทำการเลือกใช้งานต่อไป ดังรูปที่ ค.11



รูป ค.11 จอภาพหลักของโปรแกรม

ค.2 การใช้งานโปรแกรม

โปรแกรมจัดกลุ่มบทคัดย่อทางวิชาการด้วยอัลกอริทึมเคมีเดียนส์ เป็นโปรแกรมที่ทำการจัดกลุ่มบทคัดย่อออกเป็นกลุ่ม ๆ โดยจะมีการทำงานกับแฟ้มข้อมูลเอกสาร (Text File) เท่านั้น นอกจากการจัดกลุ่มบทคัดย่อแล้วยังมีความสามารถในการจัดการแฟ้มข้อมูลเอกสารต่าง ๆ ได้อีกด้วย อาทิเช่น สร้างแฟ้มข้อมูลเอกสาร เปิดแฟ้มข้อมูลเอกสาร พิมพ์แฟ้มข้อมูลเอกสารออกทางเครื่องพิมพ์ เป็นต้น

ก่อนที่จะใช้ โปรแกรมจัดกลุ่มบทคัดย่อ ผู้ใช้ต้องทำการเตรียมข้อมูลต่าง ๆ ดังต่อไปนี้

(1) แฟ้มข้อมูลบทคัดย่อ คือบทคัดย่อที่ต้องการนำมาจัดกลุ่ม โดยจะต้องเป็นแฟ้มข้อมูลเอกสารที่มีนามสกุล .txt เท่านั้น และให้สร้างหนึ่งบทคัดย่อต่อหนึ่งแฟ้มข้อมูล แล้วนำแฟ้มข้อมูลบทคัดย่อทั้งหมดมารวมกันไว้ในกล่องแฟ้มข้อมูล (Folder) เดียวกัน

(2) แฟ้มข้อมูลคำสำคัญ คือแฟ้มข้อมูลเอกสารที่มีนามสกุล .txt เช่นกัน โดยจะเก็บคำสำคัญต่าง ๆ เพื่อใช้ในการจัดกลุ่มบทคัดย่อ ซึ่งในการจัดกลุ่มแต่ละครั้งจะใช้แฟ้มข้อมูลคำสำคัญเพียงหนึ่งแฟ้มเท่านั้น และในการสร้างแฟ้มข้อมูลคำสำคัญจะต้องสร้างให้มีหนึ่งคำสำคัญต่อหนึ่งบรรทัด ดังตัวอย่างต่อไปนี้

data mining

recovery

distributed database

data warehouse

query

เมื่อทำการเตรียมข้อมูลต่าง ๆ เรียบร้อยแล้ว ก็สามารถเรียกใช้โปรแกรมจัดกลุ่มบทคัดย่อได้ โดยเมื่อทำการเรียกใช้โปรแกรม จะปรากฏจอภาพหลักของโปรแกรมดังรูปที่ ค.11 ผู้ใช้สามารถเรียกการทำงานต่าง ๆ ของโปรแกรมได้จากจอภาพหลักโดยทำการคลิกเลือกจากเมนูบาร์ หรือไอคอนทูบาร์ ซึ่งการทำงานของโปรแกรมสามารถแบ่งได้เป็น 3 ส่วนด้วยกัน ดังนี้

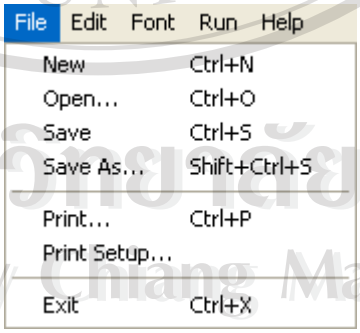
(1) การจัดการเกี่ยวกับแฟ้มข้อมูล

(2) การทำงานพื้นฐานต่าง ๆ เช่น การทำงานเกี่ยวกับการจัดการ (Edit) การจัดการข้อความ การจัดการตัวหนังสือ (Font) และส่วนอธิบายการใช้โปรแกรม

(3) การจัดกลุ่มบทคัดย่อทางวิชาการ

1. ส่วนการจัดการเกี่ยวกับแฟ้มข้อมูล

ส่วนนี้จะประกอบไปด้วยส่วนของการเปิด บันทึก การพิมพ์แฟ้มข้อมูล และการออกจากโปรแกรม โดยเมื่ออยู่ที่จอภาพหลัก สามารถเลือกการทำงานได้จากเมนูโดยคลิกเลือกที่ File จะปรากฏเมนูให้เลือก ดังรูปที่ ค.12 หรือเลือกการทำงานได้จากไอคอนทูบาร์  หากต้องการสร้างแฟ้มข้อมูลใหม่ (New) เปิดแฟ้มข้อมูล (Open) บันทึกแฟ้มข้อมูล (Save) และพิมพ์แฟ้มข้อมูล (Print) ตามลำดับ

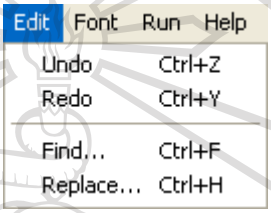


รูป ค.12 เมนูเกี่ยวกับการจัดการแฟ้มข้อมูล

2. ส่วนของการทำงานพื้นฐานต่างๆ

2.1 การทำงานเกี่ยวกับการจัดการ

ส่วนนี้จะประกอบไปด้วยส่วนของการทำย้อนกลับ (Undo) การทำซ้ำ (Redo) การค้นหา (Find) และการแทนที่ (Replace) โดยเมื่ออยู่ที่จอภาพหลักสามารถเลือกการทำงานได้จากเมนูโดยกดเลือกที่ Edit จะปรากฏเมนูให้เลือก ดังรูปที่ ค.13 หรือเลือกการทำงานได้จากไอคอนทูบาร์  หากต้องการทำย้อนกลับ และทำซ้ำตามลำดับ



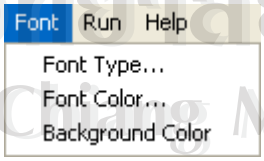
รูป ค.13 เมนูเกี่ยวกับการจัดการ

2.2 การทำงานเกี่ยวกับการจัดการข้อความ

ส่วนนี้จะประกอบไปด้วยส่วนของการตัดข้อความ (Cut) การคัดลอกข้อความ (Copy) และการวางข้อความ (Paste) โดยเมื่ออยู่ที่จอภาพหลักสามารถเลือกการทำงานได้จากไอคอนทูบาร์  ตามลำดับ

2.3 การทำงานเกี่ยวกับการจัดการตัวหนังสือ

ส่วนนี้จะประกอบไปด้วยส่วนของการเลือกชนิดของตัวอักษร การเลือกสีของตัวอักษร และการเลือกสีพื้นหลัง โดยเมื่ออยู่ที่จอภาพหลักสามารถเลือกการทำงานได้จากเมนู โดยกดเลือกที่ Font จะปรากฏเมนูให้เลือก ดังรูปที่ ค.14 หรือเลือกการทำงานได้จากไอคอนทูบาร์  หากต้องการเลือกชนิดของตัวอักษร และเลือกสีของตัวอักษร ตามลำดับ



รูป ค.14 เมนูเกี่ยวกับการจัดการตัวหนังสือ

2.4 การทำงานเกี่ยวกับส่วนอธิบายการใช้โปรแกรม

ส่วนนี้จะประกอบไปด้วยส่วนของการอธิบายโปรแกรม (Help) และรายละเอียดของโปรแกรม (About) โดยเมื่ออยู่ที่จอภาพหลักสามารถเลือกการทำงานได้จากเมนู โดยกดเลือกที่ Help จะปรากฏเมนูให้เลือก ดังรูปที่ ค.15 หรือเลือกการทำงานได้จากไอคอนทูบาร์  หากต้องการเรียกดูคู่มือการใช้โปรแกรม

รูป ค.15 เมนูเกี่ยวกับส่วนอธิบายการใช้โปรแกรม

3. ส่วนของการจัดกลุ่มบทคัดย่อทางวิชาการ

ส่วนนี้จะประกอบไปด้วยส่วนของการจัดกลุ่มบทคัดย่อ และการเรียกดูตารางคำสำคัญ โดยเมื่ออยู่ที่จอภาพหลักสามารถเลือกการทำงานได้จากเมนู โดยกดเลือกที่ Run จะปรากฏเมนูให้เลือก ดังรูปที่ ค.16 หรือเลือกการทำงานได้จากไอคอนทูบาร์   หากต้องการการจัดกลุ่มบทคัดย่อและการเรียกดูตารางคำสำคัญ ตามลำดับ

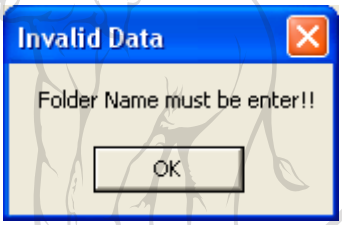
รูป ค.16 เมนูเกี่ยวกับการจัดกลุ่มบทคัดย่อทางวิชาการ

3.1 กรณีต้องการจัดกลุ่มบทคัดย่อ

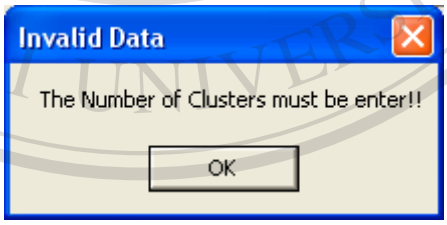
เมื่อคลิกเลือกการจัดกลุ่มบทคัดย่อ จะปรากฏจอภาพรับข้อมูล ผู้ใช้จะต้องกรอกชื่อแฟ้มข้อมูลคำสำคัญ (Keyword File Name) ชื่อกล่องแฟ้มข้อมูลบทคัดย่อ (Folder Name) และจำนวนกลุ่มที่ต้องการแบ่ง (The Number of Cluster) จากนั้นให้คลิกปุ่ม OK โปรแกรมจะทำการตรวจสอบข้อมูลนำเข้าที่ผู้ใช้กรอก หากไม่เป็นไปตามเงื่อนไขที่กำหนดไว้ จะปรากฏจอภาพแสดงข้อผิดพลาดที่แสดงข้อความตามข้อผิดพลาดที่เกิดขึ้น ดังรูปที่ ค.17 ถึงรูปที่ ค.22 แต่หากตรวจสอบข้อมูลนำเข้าแล้วเป็นไปตามเงื่อนไข โปรแกรมก็จะทำการจัดกลุ่มบทคัดย่อออกมาตามจำนวนกลุ่มที่ระบุ และแสดงผลการจัดกลุ่มออกมาทางจอภาพ แต่ถ้าต้องการยกเลิกการจัดกลุ่มบทคัดย่อก็สามารถคลิกปุ่ม Cancel ได้

โดยในจอภาพผลการจัดกลุ่มบทคัดย่อ โปรแกรมจะแสดงค่าผิดพลาดยกกำลังสอง และชื่อแฟ้มข้อมูลบทคัดย่อออกมาเป็นกลุ่ม ๆ โดยชื่อแฟ้มข้อมูลบทคัดย่อที่มีเครื่องหมาย *** จะเป็นบทคัดย่อศูนย์กลางของกลุ่มนั้น ๆ ผู้ใช้สามารถทำการจัดกลุ่มบทคัดย่อได้หลาย ๆ ครั้ง เนื่องจากในการจัดกลุ่มแต่ละครั้ง ผลลัพธ์ที่ได้อาจไม่เหมือนกัน จากนั้นก็เลือกเอาผลการจัดกลุ่มที่มีค่าผิดพลาดยกกำลังสองน้อยที่สุดเท่าที่พอใจมาใช้ นอกจากนี้ผู้ใช้ก็ยังสามารถบันทึก หรือพิมพ์ผลการจัดกลุ่มออกทางเครื่องพิมพ์ได้ ซึ่งขั้นตอนของการจัดกลุ่มบทคัดย่อ แสดงดังรูปที่ ค.23 และรูปที่ ค.24

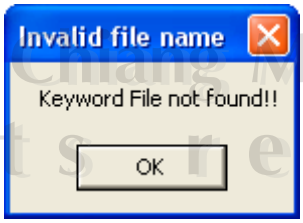
รูป ค.17 จอภาพแสดงข้อผิดพลาดกรณีไม่กรอกชื่อแฟ้มข้อมูลคำสำคัญ



รูป ค.18 จอภาพแสดงข้อผิดพลาดกรณีไม่กรอกชื่อกล่องแฟ้มข้อมูลบทคัดย่อ

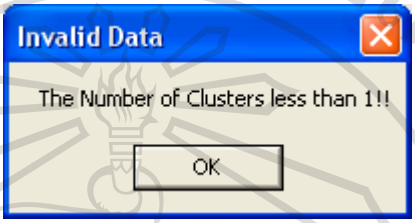


รูป ค.19 จอภาพแสดงข้อผิดพลาดกรณีไม่กรอกจำนวนกลุ่มที่ต้องการแบ่ง

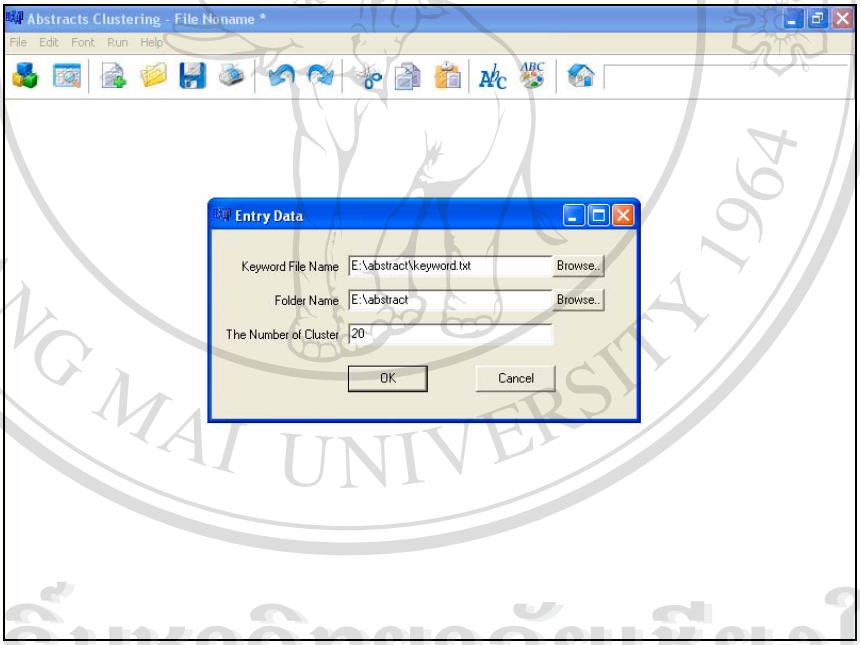


รูป ค.20 จอภาพแสดงข้อผิดพลาดกรณีที่กรอกชื่อแฟ้มข้อมูลคำสำคัญที่ไม่มีอยู่จริง

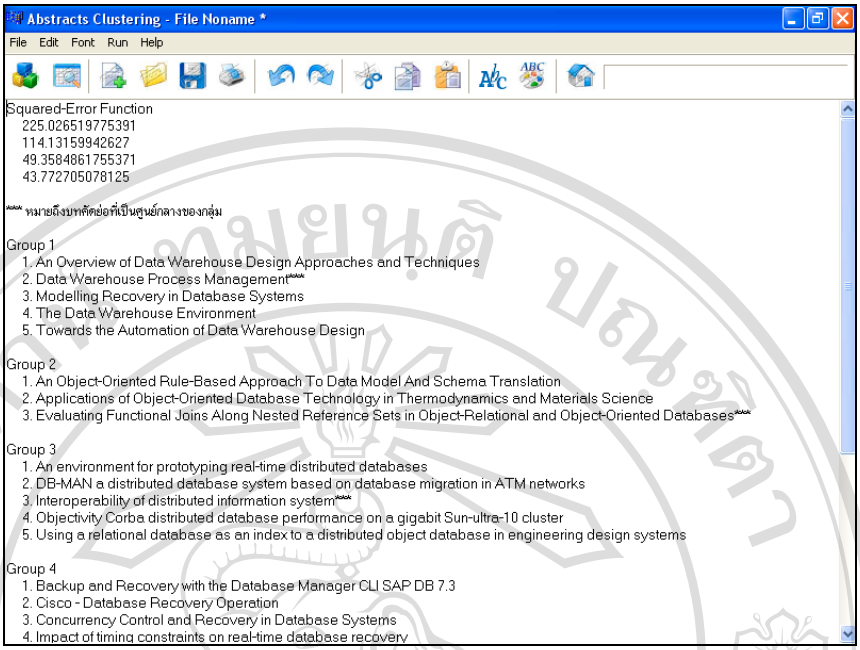
รูป ค.21 จอภาพแสดงข้อผิดพลาดกรณีที่กรอกชื่อกล่องแฟ้มข้อมูลบทคัดย่อที่ไม่มีอยู่จริง



รูป ค.22 จอภาพแสดงข้อผิดพลาดกรณีที่กรอกจำนวนกลุ่มที่ต้องการแบ่งน้อยกว่าหนึ่งกลุ่ม



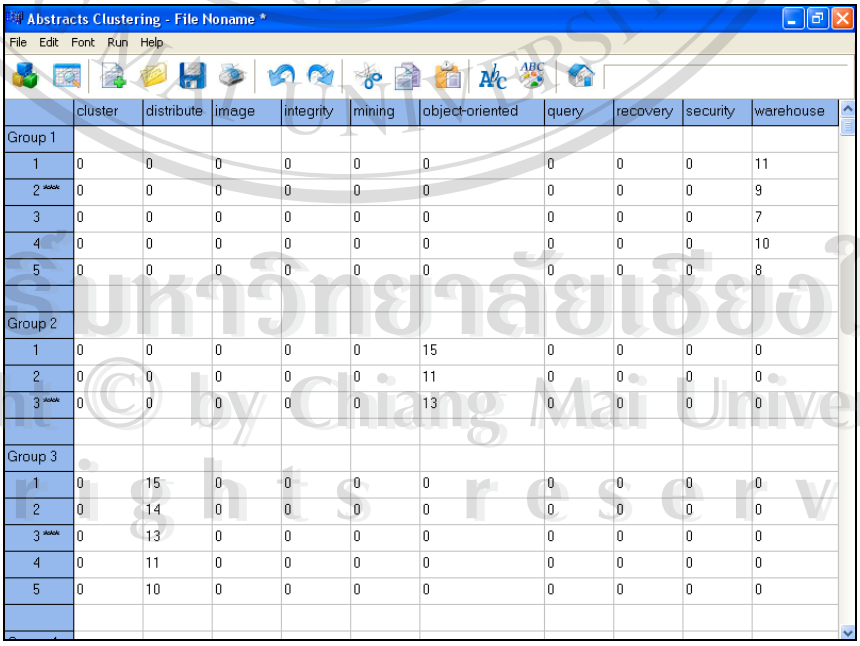
รูป ค.23 จอภาพรับข้อมูลจากผู้ใช้



รูป ค.24 จอภาพแสดงผลการจัดกลุ่มบทคัดย่อ

3.2 กรณีต้องการเรียกดูตารางคำสำคัญ

เมื่อคลิกเลือกการเรียกดูตารางคำสำคัญจะปรากฏจอภาพแสดงตารางคำสำคัญ ซึ่งเป็นตารางที่แสดงจำนวนคำสำคัญของแต่ละบทคัดย่อในแต่ละกลุ่มข้อมูลที่ทำการจัดกลุ่มได้ เพื่อให้เห็นถึงความคล้ายคลึงกันของบทคัดย่อภายในกลุ่มได้อย่างชัดเจนยิ่งขึ้น ดังรูปที่ ค.25



	cluster	distribute	image	integrity	mining	object-oriented	query	recovery	security	warehouse
Group 1										
1	0	0	0	0	0	0	0	0	0	11
2****	0	0	0	0	0	0	0	0	0	9
3	0	0	0	0	0	0	0	0	0	7
4	0	0	0	0	0	0	0	0	0	10
5	0	0	0	0	0	0	0	0	0	8
Group 2										
1	0	0	0	0	0	15	0	0	0	0
2	0	0	0	0	0	11	0	0	0	0
3****	0	0	0	0	0	13	0	0	0	0
Group 3										
1	0	15	0	0	0	0	0	0	0	0
2	0	14	0	0	0	0	0	0	0	0
3****	0	13	0	0	0	0	0	0	0	0
4	0	11	0	0	0	0	0	0	0	0
5	0	10	0	0	0	0	0	0	0	0

รูป ค.25 จอภาพตารางคำสำคัญ

ประวัติผู้เขียน

ชื่อ	นางสาวณัฐกานต์ เอี่ยมอ่อน
วัน เดือน ปี เกิด	8 พฤศจิกายน 2521
ประวัติการศึกษา	สำเร็จการศึกษามัธยมศึกษาตอนปลาย โรงเรียนสามัคคีวิทยาคม เชียงราย ปีการศึกษา 2539 สำเร็จการศึกษาปริญญาวิทยาศาสตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ มหาวิทยาลัยเชียงใหม่ ปีการศึกษา 2543