

การทำเหมืองข้อมูลเป็นเทคโนโลยีใหม่ของการวิเคราะห์ข้อมูลอัตโนมัติ เพื่อค้นหาความรู้ที่จะเป็นประโยชน์แก่การวางแผนดำเนินการของหน่วยงานเจ้าของข้อมูล ความรู้ที่ค้นพบนี้เป็นได้หลายรูปแบบ เช่น แพทเทิร์นที่เกิดขึ้นภายในกลุ่มข้อมูลที่สามารถใช้ทำนายลักษณะที่จะเกิดขึ้นในอนาคตหรือทำนายแนวโน้มการเปลี่ยนแปลงของข้อมูล ลักษณะที่สัมพันธ์เชื่อมโยงกันของข้อมูล ลักษณะที่เบี่ยงเบนไปจากข้อมูลกลุ่มใหญ่ และรูปแบบประเภทอื่นๆ อีกหลากหลาย ปัจจัยสำคัญของการทำเหมืองข้อมูลให้ได้ความรู้ที่มีประโยชน์ คือ อัลกอริทึมสังเคราะห์ความรู้ และ ตัวข้อมูล ถ้าอัลกอริทึมไม่มีประสิทธิภาพก็จะไม่สามารถค้นหาความรู้ที่แฝงอยู่ในข้อมูลได้ หรือถ้าข้อมูลไม่มีคุณภาพเพียงพอก็จะไม่ช่วยให้สังเคราะห์ความรู้ได้ออกมาได้ งานวิจัยนี้จึงเสนอขึ้นเพื่อพัฒนาซอฟต์แวร์ชื่อเอสยูทีไมเนอร์ที่มีส่วนประกอบทั้งส่วนฟิวเจอร์ ส่วนค้นหาแพทเทิร์น และส่วนประเมินแพทเทิร์น เอสยูทีไมเนอร์คือระบบเหมืองข้อมูลที่พัฒนาขึ้น เพื่อใช้เป็นเครื่องมือวิเคราะห์ข้อมูลอัจฉริยะสำหรับค้นหารูปแบบหรือสารสนเทศที่เป็นประโยชน์ จากข้อมูลที่เก็บอยู่ในฐานข้อมูล ผู้วิจัยได้นำเสนอวิธีการออกแบบและการพัฒนาเอสยูทีไมเนอร์ ที่จัดเป็นระบบเหมืองข้อมูลสมบูรณ์แบบ เนื่องจากได้ร่วมส่วนการประมวลผลก่อนและหลังการทำเหมืองข้อมูลเข้าไว้ในระบบเดียวกัน ในรายงานนี้ผู้วิจัยได้นำเสนอกรอบแนวคิดของระบบ และวิธีการพัฒนาระบบในส่วนของโปรแกรมการทำเหมือง โดยใช้วิธีการโปรแกรมเชิงตรรกะด้วยภาษาโปรล็อก วิธีการโปรแกรมเชิงประกาศของโปรล็อกช่วยให้การเขียนคำสั่งทำได้สั้นและชัดเจน นอกจากนี้การสนับสนุนการทำแพทเทิร์นแมตชิงของภาษาโปรล็อกยังเป็นข้อได้เปรียบอย่างมากสำหรับงานการค้นหารูปแบบ ผลสำเร็จของการพัฒนาซอฟต์แวร์นี้จะเป็นจุดเริ่มต้นของการสร้างระบบทำเหมืองข้อมูลสมรรถนะสูงที่ทำงานกับข้อมูลขนาดใหญ่มากและมีลักษณะการเกิดขึ้นอย่างต่อเนื่องได้ต่อไปในอนาคต

Data mining is a new technology in automatic data analysis. It is the search for valuable information, or knowledge, in large volumes of data. The discovered knowledge can greatly benefit the organization that owns the data in many ways, for example, to aid decision-making, to reduce the risk of business planning or to project the revenue in future investment. The common types of knowledge discovery are prediction, deviation detection, database segmentation, clustering, association rules and link analysis. Despite the many forms and functions of discovered knowledge, the most important factors in mining valuable knowledge are the learning algorithm and data. The practical data mining system needs an efficient algorithm and high quality data. This research project presents the design and implementation of a complete data mining system, called SUT Miner. The SUT Miner is a data mining system developed as an intelligent data analysis tool to discover patterns and extract useful information from facts stored in databases. In addition to the mining engine, our system incorporates the pre-mining and post-mining parts. In this report, we describe a framework of SUT Miner and present the implementation scheme on the mining engine part using a logic programming paradigm, a Prolog language in particular. A high-level declarative style of Prolog facilitates a clear and concise coding. The language also supports pattern matching, which is a big advantage for a task of pattern discovery. The success of this project will be the preliminary step toward the future development of a high-performance data mining system that can handle very large volume of continuously generated data.