

Jaruwan Kanjanasupawan 2011: Enterprise Expert Clustering of Public University by Two Stage Algorithms and Feature Weighting Techniques. Master of Science (Computer Science), Major Field: Computer Science, Department of Computer Science. Thesis Advisor: Associate Professor Anongnart Srivihok, Ph.D. 104 pages.

In a large university there are many experts in several domains such as agriculture, economics, science, and engineering. Therefore finding who is an expert in which domain is time consuming. The experts in different domains may have the same domain of expertise. This research proposed data mining techniques to segment expert's data by using keywords. Datasets were collected from 3 databases included (1) personal database (2) research projects and (3) advisors and thesis. Since, dataset included too many attributes (971 attributes), so hybrid attribute selection techniques using CFS (Correlation-based Feature subset Selection) and Genetic Search were conducted.

Clustering techniques used were two stage algorithms to determine the appropriate number of clusters by using SOM (Self-Organizing Maps) and to cluster data by using K-Means and Fuzzy C-Means. Clustering techniques were joined feature weighting techniques including (1) TF-IDF (Term Frequency-Inverse Document Frequency), (2) Logarithm Weight (Logarithm Term Frequency-Inverse Document Frequency) and (3) Augmented Weight.

Evaluators were F-Statistic, R-Squared and Silhouette. Results showed that the performance of dataset with attribute selection (from 971 to 258 attributes) were better than original dataset (non attribute selection). The appropriate number of clusters were 7 clusters by using SOM. The most efficient clustering was K-Means joined with Logarithm Weighting techniques. Next, expert extraction in each cluster was: (1) 50 experts, keyword: economic (2) 249 experts, keyword: agriculture (3) 63 experts, keyword: agriculture (4) 11 experts, keyword: agriculture (5) 43 experts, keyword: education (6) 2,671 experts and (7) 107 experts, keyword: engineer. Therefore, cluster 6 was too large. So this cluster was re-clustered. However, results showed that the re-clustered should be stopped in this process. The results of this study should bring about the efficient experiment clustering methods. In other ways, practitioners can apply this knowledge to develop information systems in the university such as expert searching system in the future.

---

Student's signature

---

Thesis Advisor's signature