

# *Abstract*

The amounts of data kept in computer files with database are growing at a phenomenal rate. The users of these data are expecting more sophisticated information from them. The DML (Data Management Language) is not adequate to support these needs. Data mining is concepts which finding unknown information in database. The Attribute-Oriented Induction (AOI) technique which is one of the data mining techniques used to analyzing database. It has its own unique features, which are designed to address unique AOI problems that cannot be solved by other data analysis tools.

A data mining algorithm is the mathematical and statistical algorithms that transform the cases in the original data source into the data mining model. An AOI is used to analysis the data to generalize the attributes according to two thresholds which determine the depth of rolling-up or rolling-down. The end result is some of attributes that succeeds in generalizing by passing the thresholds test, each one of them raises one value that represents the phenomenon in that database which is mined.

The general objective is to find the useful behavior in specific database to take right action in future work according to hidden information that is abstracted by AOI technique. The main obstacle in AOI algorithm is the input of algorithm includes (specified attributes that are relevant to the user query, also specified generalized attributes), as we have seen in this thesis, this is not practical. So we used another means to aid AOI technique through these obstacles by using Genetic Algorithm (GA), the main step in GA is Fitness Function, which is to determined it nontrivial mission, in our research we used another data mining technique as a fitness function and that is the Association Rules.

## الخلاصة

أن حجم البيانات المحفوظة في ملفات الحاسبة ضمت قواعد بيانات تتنامى بشكل كبير حسب درجة تعقيد النظام، مما يجعل المستخدمين لهذه البيانات يتوقعون الحصول على النتائج الخاصة بالظواهر الاجتماعية ذات زيادة في عدم دقة المعلومات. إضافة لذلك لغات التعامل مع البيانات (DML) المعتمدة في تلك الأنظمة غير كافية لدعم تلك المتطلبات..

هنا تظهر أهمية اعتماد مفاهيم تعدين البيانات (data mining concepts) كوسيلة لأستخلاص المعلومات الغير معرفة في القاعدة البيانات الموجودة سابقا". أن تقنية استقراء الصفة الموجهة (Attribute-Oriented Induction) هي إحدى تقنيات تعدين البيانات تتولى تحليل محتويات القاعدة البيانات وإيجاد حلول للمشاكل التي لا توفرها اللغات الخاصة بالتعامل مع البيانات المعتمدة في الأنظمة القواعدية التقليدية. إن خوارزمية تعدين البيانات هي خوارزمية تعتمد النظريات الرياضية والإحصائية لتحويل البيانات الأصلية إلى نموذج مناسب لتقنيات تعدين البيانات.

تتولى تقنية الاستقراء الموجه للصفة تعميم (Generalization) الصفات البياناتية استناداً إلى عتبتين (two thresholds) تحدد مدى عمق (deep) الالتفاف العلوي (rolling up) أو الالتفاف المنخفض (drilling down). و النتيجة النهائية عبارة عن عدد من الصفات التي نجحت في عبور فحص التعميم باعتماد العتبتين أعلاه، كل واحدة من هذه الصفات تكون مرفقة بقيمة معينة و تمثل ظاهرة في قاعدة البيانات. إن الهدف الرئيسي هو إيجاد سلوك مفيد في قاعدة بيانات معينة لأتحاذ الأفعال المناسبة في العمل المستقبلي، استناداً للمعلومات المخفية التي تم استخلاصها بواسطة تقنية استقراء الصفة الموجهة.

أن العقبة الرئيسية في الاستقراء الموجه للصفات هو عملية الإدخال للخوارزمية التابعة له (المتضمنة تحديد الصفات المستخدمة مع سؤال المستخدم وتحديد الصفات العامة)، و كما سنرى في هذه الأطروحة إن هذه الطريقة غير عملية. لذا استخدمنا في هذا البحث الخوارزمية الجينية (GA) لإلغاء إدخال الصفات. أن الخطوة الرئيسية في الخوارزمية الجينية هي خطوة دالة الهدف (fitness function) و تحديدها ليس بالعمل السهل ، في بحثنا استخدمنا تقنية تعدين بيانات أخرى كدالة هدف و هي تقنية القواعد التجميعية (association rule). تم بناء النظام باستخدام الخوارزمية الجينية لمعاملة الصفات ذات العلاقة بسؤال المستفيد باستخدام القواعد التجميعية كدالة هدف مع استخدام مخرجات الخوارزمية الجينية كمداخل لخوارزمية الاستقراء الموجه للصفة. و قد تم تنفيذ البحث بواسطة لغة (VB6) مع قاعدة بيانات (Microsoft access97) تحت نظام تشغيل (millennium) في حاسبة من نوع (Pentium IV).