

بسم الله الرحمن الرحيم
**جامعة السودان للعلوم و
التكنولوجيا**
كلية الدراسات العليا

**تطوير إطار عمل إرشادي للتنقيب في
بيانات السجل المدني**
**دراسة حالة: السجل المدني والخدمة
الوطنية**

**Development of a Guiding Framework for Mining in the
Civil Registry Data**

**A Case study: the Civil Registry and the National
Service**

**بحث مقدم لنيل درجة الماجستير في
علوم الحاسوب**

إشراف الأستاذ:

إعداد الدارس:

حاج حمد عبد الملك عبد الله
عوض الكريم
د. عوض محمد

الإستهلال

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ
(فَأَمَّا الزَّبَدُ فَيَذْهَبُ جُفَاءً وَأَمَّا مَا يَنْفَعُ النَّاسَ فَيَمْكُتُ فِي
الْأَرْضِ كَذَلِكَ يَضْرِبُ اللَّهُ الْأَمْثَالَ)

سورة الرعد 19

الشكر والعرفان

- جزيل الحمد والشكر والثناء الذي لا يذ قطع أبداً إلي الله تعالى الذي ألهمني المثابرة في طلب العلم .
- **الشكر أجزله وأوفره للدكتور: عوض محمد عوض الكريم** الذي كان بحراً من العلم و المعرفة، و الذي قادني من بداية الطريق إلي نهايته.
- **الشكر الجزيل لمدير هيئة الجوازات والسجل المدني ومدير الإدارة العامة للسجل المدني لدعمهم الكبير في جميع مراحل هذا البحث.**
- **أسمى ايات الشكر والعرفان أهديتها الى كل من أعاننى في إعداد هذا البحث و كل من أسهم في أن يري هذا الجهد النور وأخص بالشكر الاستاذ: عبد المطلب محمد والاستاذ: ابراهيم موسى .**

الإهداء

بسم الله الرحمن الرحيم

(قل إعملوا فسيرى الله عملكم ورسوله والمؤمنون)

صدق الله العظيم

إلهي لا يطيب الليل إلا بشكرك ولا يطيب النهار إلا بطاعتك .. ولا تطيب
اللحظات إلا بذكرك .. ولا تطيب الآخرة إلا بعفوك .. ولا تطيب الجنة إلا
برؤيتك الله جل جلاله
إلى من بلغ الرسالة وأدى الأمانة .. ونصح الأمة .. إلى نبي الرحمة ونور
العالمين ..

سيدنا محمد صلى الله عليه وسلم

إلى من كلله الله بالهبة والوقار .. إلى من علمني العطاء بدون
انتظار .. إلى من أحمل أسمه بكل افتخار .. وستبقى كلماتك نجوم
أهتدي بها اليوم وفي الغد وإلى الأبد ..
والذي العزيز له الرحمة والمغفرة
إلى ملاكي في الحياة .. إلى معنى الحب وإلى معنى الحنان والتفاني ..
إلى بسمة الحياة وسر الوجود
إلى من كان دعائها سر نجاحي وحنانها بلسم جراحي إلى أغلى الحبايب
إلى من بها أكبر وعليه أعتمد .. إلى شمعة متقدة تنير ظلمة حياتي ..

إلى من بوجودها أكتسب قوة ومحبة لا حدود لها ..
إلى من عرفت معها معنى الحياة

أمي الحبيبة

الى من سهرتى معى لتوفير المناخ الملائم لإنجاز هذا العمل
زوجتى العزيزة

المستخلص

تعد بيانات السجل المدنى ثروة وطنية يجب المحافظة عليها وتنميتها، فضلاً عن دورها الإستراتيجي فى التنمية الشاملة والمستدامة. التنقيب فى البيانات عبارة عن إستخدام خوارزميات لإستخلاص المعرفة والمعلومات من بين كمية كبيرة من البيانات. عن طريق التنقيب فى مستودعات البيانات المتكاملة بين السجل المدنى ومؤسسات الدولة الحيوية ذات الصلة يمكننا تحقيق الإستفادة المثلى لإنتاج المعرفة المطلوبة لصنع القرارات التنموية الراشدة فى شتى المجالات (مثل الصحة، التعليم، الزراعة، الخدمة الوطنية). يهدف هذا البحث لتطوير إطار عمل إرشادى سهل للتنقيب فى بيانات السجل المدنى مع بيانات مؤسسات الدولة الحيوية ذات الصلة. يقدم البحث تطبيق لإطار العمل المقترح (كدراسة حالة) على مستودع بيانات متكامل من قاعدة بيانات الجنسية وقاعدة بيانات الخدمة الوطنية لتؤكد جدواه وتطبيقه. ومما يعكس اهمية هذه الدراسة كونها المبادرة الاولى والنواة الاصلية لإنشاء وحدة للدراسات والبحث والتطوير بإدارة السجل المدنى كشريك استراتيجى للتنسيق مع الجهات المستفيدة وذات الصلة الأكاديمية منها والبحثية للتنقيب فى مستودعات البيانات المتكاملة لتحقيق الغايات المنشودة.

Abstract

The civil registry data is considered as a national wealth that should be attentively conserved and enlarged, rather its strategic role in the inclusive and sustainable development. The Data Mining is using algorithms to extract knowledge and information from a large amount of data. By mining in the integrated data warehouses of the civil registry and the relevant vital federation institutions , we can achieve the optimum utilization for the production of knowledge required for the decision-making of the mature development in various areas (such as health, education, agriculture, national service).

This research aims to develop a guiding simple framework for mining in the civil registry data with data of the relevant crucial country institutions. In addition, the research presents an application of the proposed framework as a case study on the integrated data warehouse of the civil registry and the National Service in order to assure it's the usefulness and applicability of the proposed scheme. The importance of this study emerges as it is the first initiative and e novel core for foundation of a unit in the civil registry management for research and development, as well as a strategic partner to coordinate with the stakeholders to achieve the desired goals.

جدول المحتويات :

رقم الصفحة	العنوان
أ	الغلاف
ب	الإستهلال
ج	الشكر والعرفان
د	الإهداء
هـ	مستخلص البحث
و	Abstract
ز	جدول المحتويات
ح	فهرس الجداول
م	فهرس الأشكال
1	الباب الأول: المقدمة Introduction
2	تمهيد
3	1.1 مشكلة البحث
4	2.1 أهداف الدراسة

5	3.1 أهمية البحث
6	4.1 حدود البحث
7	5.1 تنظيم البحث
8	الباب الثاني:إستعراض الخلفية Literature Review
9	مقدمة
9	1.2 السجل المدنى
11	2.2 التنقيب فى البيانات
12	3.2 نشأة التنقيب فى البيانات
13	4.2 تطبيقات التنقيب فى البيانات
13	5.2 خطوات إكتشاف المعرفة
14	6.2 طرق التنقيب عن البيانات
15	7.2 أنواع تنقيب البيانات (Type Of Data Mining)
15	1.7.2 التنقيب التنبؤي (Predictive Data Mining)
15	2.7.2 التنقيب الوصفي (Descriptive Data Mining)
15	8.2 مهام تنقيب فى البيانات (Data Mining Tasks)
17	9.2 مستودعات البيانات(Data Warehousing)
19	1.9.2 الخطوات التقنية لبناء مستودعات البيانات
20	10.2 خطوات التنقيب فى البيانات
20	1.10.2 تعريف وتحديد المشكلة
20	2.10.2 بناء قواعد البيانات
23	3.10.2 تكامل البيانات (Data Integration)
24	4.10.2 بناء النموذج
24	11.2 تقنيات التنقيب في البيانات
25	12.2 الاعمال ذات الصلة

26	الباب الثالث: منهجية البحث
27	مقدمة
27	1.3 تلخص لمنهجية البحث
28	الباب الرابع: إطار العمل المقترح
29	مقدمة
30	1.4 الهيكل العامة للإطار المقترح
31	2.1.4 مرحلة تعريف المشكلة وتحضير البيانات
34	3.1.4 مرحلة تحليل البيانات وبناء النموذج
35	4.1.4 مرحلة التطبيق والتثبيت للنموذج
36	الباب الخامس: التطبيق وإستخلاص النتائج
37	المقدمة
38	1.5 إختيار البيانات
38	2.1.5 عملية تحديد وإسترجاع البيانات
40	2.5 تهيئة البيانات
42	3.5 نقل البيانات
51	4.5 تنقيب البيانات
52	1.4.5 الأدوات المستخدمة فى التطبيق
55	5.4.2 إختيار الخوارزمية المناسبة للتصنيف
62	3.4.5 التنقيب عن البيانات والإحصاءات ضمن بيانات الاختبار
63	4.4.5 حفظ النموذج والنتائج
66	5.4.5 تدريب الشبكة العصبية
68	6.4.5 تدفق المعرفة
69	7.4.5 خطوات عملية بناء المعرفة
70	الباب السادس : الخلاصة والدراسات المستقبلية

71	1.6 الخلاصة
71	2.6 توصيات للباحثين
71	3.6 توصيات عامه
72	4.6 المراجع

فهرس الجداول :

الصفحة

عنوان الجداول

38	جدول (1.5) :يوضح جزء من قاعدة البيانات الموحدة (59 حقل)
----	---

39	جدول (2.5): يوضح الحقول التي تم إختيارها
40	جدول (3.5) يوضح البيانات بعد التهيئة
41	جدول (4.5) :يوضح الحقول التي تم إختيارها
41	جدول (5.5) يوضح السطر الثامن من الجدول

فهرس الأشكال :

الصفحة

عنوان الأشكال

- 3 الشكل رقم (1.1) : السجل المدنى وبعض أجهزة الدولة ذات الصلة
- 15 الشكل رقم (1.2) : يوضح انواع تنقيب البيانات
- 16 الشكل رقم (2.2) : يوضح نموذج للعنقدة
- 18 الشكل رقم (3.2) : يوضح مستودع للبيانات
- 30 الشكل رقم (1.4) : يوضح العمليات العامة للتنقيب فى البيانات
- 33 الشكل رقم (2.4): يوضح المرحلة الأولى تعريف المشكلة وإعداد البيانات
- 34 الشكل رقم (3.4) : يوضح إختيار التقنية المناسبة للتنقيب فى البيانات
- 35 الشكل رقم (4.4) : يوضح الإطار العام
- 50 الشكل رقم (1.5) : يبين تقسيم البيانات فى شكلها الجديد (. (Arff
- 52 الشكل رقم (2.5) : يوضح الواجهات المختلفة من الويكا
- 57 الشكل رقم (3.5) : يوضح الشاشة الرئيسية للويكا
- 58 الشكل رقم (4.5) : يوضح مسار الملف ونوع التجربة
- 59 الشكل رقم (5.5): يوضح إعدادات البيانات و ملف المخرجات بالإضافة إلى إختيار الخوارزميات
- 60 الشكل رقم (6.5) : يوضح إجراء التجربة و الزمن المستغرق بالإضافة إلى الأخطاء
- 61 الشكل رقم (7.5) : يوضح نتائج المقارنة بين الخوارزميات الأربع و يبدو جلياً أفضلية خوارزمية شجرة القرار(48J)
- 62 الشكل رقم (8.5) : يوضح فتح ملف بيانات الاختبار
- 64 الشكل رقم (9.5): يوضح النتائج والنسبة الصحة والخطأ فيها

- 65 الشكل رقم (10.5): يوضح النتائج فى شكل اخر(View The Tree)
- 66 الشكل رقم (11.5) : يوضح طريقة حفظ النموذج والنتائج
- 67 الشكل رقم (12.5) : يوضح النتائج بإستخدام دالة اخرى
- 68 الشكل رقم (13.5) : يوضح محتويات شاشة تدفق المعرفة
- 69 الشكل رقم (14.5) : يوضح خطوات عملية بناء المعرفة