تقييم أساليب الوصول المباشر لملفات نموذجية من واقع القطر مع التركيز على استعمال الاسماء العربية كمفاتيح

ندى عبد علي

معهد التدريب والبحوث للحاسبات والاجهزة الالكترونية

الملخص

نتائج بحث لتقييم اداء بعض أساليب تنظيم المعلومات للوصول المباشر باستعمال العنونة التحتية (Hach Addressing) من خلال تطبيقها على ملفات نموذجية للمفاتيح (Sample Keys Files) تم اختيارها من الواقع التطبيقي لأنظمة مهمة في قطرنا . الملفات متباينة الحجوم والمفاتيح مختلفة في أطوالها وكذلك في أنواعها ، أي عددية أو الفبائية او مختلطة . التركيز على استعمال الأسماء العربية كمفاتيح لأهميتها واعتماد ملف نموذجي لأسماء الأشخاص في إحدى مؤسسات القطر مصمم باستعمال روتين توليد الأعداد العشوائية (Random Number Generator) ويحتوي على 3000 مفتاح يمثل الاسم العربي (الثلاثي او الرباعي)

تجربة عدة قيم للعوامل والسعات المعينة

الخروج باستنتاجات عامة تؤيد البحوث العالمية السابقة في الموضوع مع توصيات ونتائج خاصة لتطبيقات القطر . البحث هو احد بحوث معهد البحوث والتدريب واستعملت فيه حاسبة (Honeywell) بصورة رئيسية .

المقدمة

غالبا ما تتم عملية تنظيم المعلومات في الأنظمة الممكننة على الحاسبات الالكترونية بشكل ملفات تتكون من قيود ، وهذه بدورها تتألف من بضعة حقول عادة ما يستعمل أحدها كمفتاح (Key) لتمييز القيود عن بعضها البعض . وفي ملفات الوصول المباشر نلاحظ أن العنونة النحتية (Hash Addressing) التي تقوم على أساس تخزين كل قيد في موقع (عنوان نحتي) يستخرج من اجراء بعض العمليات الرياضية او المنطقية على مفتاح ذلك القيد ، يعتبر من أسرع أساليب البحث حيث لا يعتمد وقت استرجاع أي قيد على عدد القيود المعطاة وبكلمات أخرى فإن الوصول الى أي مفتاح في جدول المفاتيح لا يعتمد على حجم الجدول وهذا عكس ما نلاحظه في أساليب البحث التقليدية مثل البحث التسلسلي (Sequential Search) والبحث الثنائي (Binary Search) فهنا كلما زاد حجم الجدول ، كلما زاد وقت الاسترجاع .

إن ميزة العنونة النحتية لا تأتي دون مقابل وهو مشكلة التضارب (Collision) وهي ظاهرة تولد عنوانين (أو أكثر) متساوية لقيود مختلفة مما يتطلب معالجة خاصة عند التخزين ومن الواضح أن الدالات النحتية (Hash Functions) المثلى هي التي تولد أفضل توزيع من العناوين النحتية من المفاتيح ، وهذا يعتمد طبعاً على طبيعة المفاتيح نظرا للاهتمام المتزايد بالجفرات العربية للرموز (Character Sets) ولغات البرمجة العربية ومجال التعريب في الحاسبات الالكترونية عموما في القطر ولكون

البحث يتناول معالجة الاسم العربي من خلال انتقاء ملفات نموذجية من واقع التطبيقات العملية في القطر فان نتائج هذا البحث تكتسب أهمية خاصة وفوائد مباشرة لكافة الأنظمة التي تعتمد الوصول المباشر من خلال المفاتيح الاسمية العربية سواء كانت تلك الأنظمة تطبيقية أو في مجال برمجة النظم (Systems Programming,

ومن الفوائد الأخرى لهذا البحث معالجة جداول الرموز (Symbol tables, وفي برامج التأليف والتجميع (Compilers and Assemblers, .

مصطلحات وتعاريف أساسية

فيما يلي تعاريف أبرز المصطلحات الشائعة في مجال العنونة التحتية (لقد تم اعتماد المعجم العربي الموحد لمصطلحات الحاسبات الالكترونية في تعريب هذه المصطلحات) :

1 ـ الدالة التحتية (Hashing Function, عبارة عن دالة رياضية تطبق على مفتاح معين لاشتقاق عنوانا تحتيا له في الجدول التحتي .

2 ـ العنوان التحتي (Hashing Address, وهو ذلك العنوان في الجدول التحتي لمفتاح معين الذي يستحصل من تطبيق دالة تحتية على المفتاح .

3 ـ الجدول التحتي (Hashing Table, وهو الجدول النهائي الذي تخزن فيه المفاتيح بالاعتماد على العناوين المستحصلة من تطبيق دالة تحتية عليها وقد تحتوي مؤشر ربط (Link, الذي يستعمل في حالة التضارب .

4 ـ التضارب (Collision, هي تلك الحالة التي ينتج فيها نفس القيمة للعنوان التحتي

لمفتاحين مختلفين أي يمكن تمثيلها رياضيا بالشكل التالي : ـ

$KI \neq KJ \quad H(KI) = H(KJ)$

5 ـ قيد فائض (Overflow Record) وهو ذلك القيد الذي يكون عنوانه النحتي مشغولا بقيد آخر فإذا وضع في عنوان آخر غير عنوانه أصبح القيد فائضا .

6 ـ قيد بائت (Home Record) وهو ذلك القيد الذي يكون في عنوانه النحتي الأصلي .

7 ـ الجس (Probing) وهي عملية مقارنة المفتاح مع محتويات عناصر الجدول النحتي حتى العثور عليه .

8 ـ عامل التحميل (Loading Factor) وهو نسبة عدد المفاتيح المراد استخراج عنوان نحتي لها على عدد المفاتيح الموجودة في الجدول النحتي (حجم الجدول) .

9 ـ الشقب (SLOT) موقع في جدول تخزين المفاتيح المنحوتة يكفي لتخزين قيد واحد .

10 ـ المحفن (Bucket) مجموعة من الشقوب تمثل مدخلا واحدا في الجدول النحتي وتقابل عنوانا نحتيا واحدا .

11 ـ سعة المحفن (Bucket Size) عدد شقوبه .

ظروف التجربة

هناك عدة عوامل تؤثر على أساليب النحت (Hashing Techniques) وجودة أدائها أهمها : ـ

- طبيعة المفاتيح المختارة .

- العوامل والمتغيرات (Factors and Variables) .

- طرق النحت (Hashing Methods) المختارة .

- طرق معالجة التضارب .

اولا - طبيعة المفاتيح المختارة :

يبين الجدول رقم (1) خصائص محتويات ملفات المفاتيح النموذجية المختارة حيث نجد أن حجومها مختلفة وإنها ذات مفاتيح مختلفة النوعية ولقد تمت دراسة طبيعتها بالنسبة لكل ملف قبل تطبيق أساليب النحت عليها اذ أن هذه الدراسة تمكننا من اختزال طول كل مفتاح وكذلك معرضة المراتب الأكثر عشوائية في المفاتيح والتي يجب التركيز عليها .

الملف النموذجي	عدد المفاتيح	النوع	طول المفتاح	الملاحظات
1 ـ ارقام الهواتف	500	عددي(N)	7	اختيار عشوائي من دليل الهاتف لمدينة بغداد
2 ـ اسماء مشاركي الهواتف	510	الفبائي(A)	≤ 24	
3 ـ ارقام تصنيف الكتب	200	الفعددي(AN)	13	من مكتبة المركز القومي للحاسبات
4 ـ ارقام الواردة	173	الفعددي(AN)	≤ 18	من المراسلات الرسمية في الادارة
5 ـ اسماء العمال	3000	الفبائي(A)	≤ 30	باستعمال مولد الاعداد العشوائية في المؤسسة العامة للتقاعد والضمان الاجتماعي

جدول (1) ملفات التجربة

ثانيا : العوامل والمتغيرات

إن أهم عاملين يؤثران على التجربة هما : ـ

ـ عامل التحميل (Load Factor)

ـ سعة المحفن (Bucket Size)

حيث كلما تقلل من قيمة عامل التحميل كلما يقل التضارب وكلما تزيد سعة المحفن يقل احتمال الـ Overflow .

بالنسبة للملفات الفيض (1-4) بالجدول رقم (1) تتراوح قيم عامل التحميل بين 0.70 -0.90 وبزيادة مقدارها 0.05 .

يتغير حجم المحفن المستعمل بالجدول النحتي وللملف أرقام الهواتف وبأخذ القيم 1 و2 و3 و5 .

أما حجم مساحة الفيض (Overflow Area) المستعملة في معالجة التضارب فهو متغير أيضا ويأخذ القيم 0.30 ,0.20 ,0.40 من مساحة الجدول النحتي .

وبالنسبة لملف أسماء العمال فإن قيم عامل التحميل تتغير بين 0.60 -0.95 وبزيادة مقداره 0.05 وسعة المحفن تساوي 1.0 وحجم مساحة الفيض بأخذ النسب 0.30, 0.22, 0.11 من مساحة الجدول النحتي .

وسنوضح فيما بعد كيفية معالجة المفاتيح الغير عددية بعد استعراض طرق النحت المجربة اضافة الى نماذج من المفاتيح نفسها .

ثالثا ـ طرق النحت (Hashing Method)

استعملت ستة أساليب مختلفة لدالات النحت (Hash Function) واحتساب العناوين النحتية وهي : ـ

1 ـ القسمة (Division)

وهي أكثر الطرق استعمالا نظرا لسهولة تطبيقها وتتلخص هذه الطريقة باعتماد قاسم (Divisor) ينتقى كأقرب عدد فردي صحيح الى حجم الجدول النحتي (ويفضل أن يكون عددا أوليا [3] (Prime) حيث تجري قسمة المفتاح المعطى على هذا القاسم ويعتمد الباقي (MOD) كعنوان نحتي مقابل لهذا المفتاح وحسب المعادلة .

HASH- ADDRESS = (KEY) MOD (PRIME)

2 ـ متوسط المربع (Mid- Squars)

وتتلخص بتربيع المفتاح واختيار (N) من الأرقام من وسط الناتج فتكون هي العنوان النحتي المطلوب ويقصد بـ N طول العنوان في الجدول النحتي .

وطبيعي عند تطبيق الطريقة على المفاتيح الأخرى فإن نفس المراتب يجب أن يتم اختيارها .

3 ـ التشذيب (Truncation)

هذه الطريقة تعتمد على التحليل الرقمي (Digit Analysis) للمفتاح أي على توزيع مراتب المفاتيح حيث يتم قطع أو اهمال المراتب التي تتوزع بصورة متحيزة (Biased) وتؤخذ الباقية أي الأكثر عشوائية كعنوان نحتي بحيث تكون بقدر طول العنوان النحتي في الجدول .

ويمكن أن يكون الحذف عشوائيا أي ليس بالاعتماد على التحليل الرقمي [1] ولكن يجب تشذيب نفس المراتب بالنسبة لكافة المفاتيح .

4 ـ **الطي** (Folding)

وفي هذه الطريقة تتم تجزئة المفاتيح الى عدة أجزاء ويفضل أن تكون الأجزاء كبيرة نوعا ما وبحيث لا يتعدى طول الجزء الواحد طول العنوان النحتي وعادة ما تكون هذه الأجزاء متساوية الطول عدا آخر جزء فهو إما يبقى على طوله او يكمل باصفـــار .

وهناك نوعان من الطي هما : ـ

أ ـ الطين الازاحي (Shift Folding)

وهذا النوع هو أبسط أنواع الطي حيث تتم تجزئة المفتاح كما أشرنا اعلاه فبمجرد جمع هذه الأجزاء نحصل على العنوان النحتي .

ب ـ الطي التخمي (Boundary Folding)

وهذا النوع من الطي يختلف عن سابقـه بطريقـة الجمع حيث يكون الجمع بطريقة تشابه تماما طي الورقة أي بجمع الجزء الثاني مع مقلوب الجزء الأول والجزء الرابع مع مقلوب الجزء الثالث وهكذا .

ويمكن تمثيل الطريقتين بالشكل التالي : ـ

ADDRESS	ADDRESS
LENGTH	LENGTH
KEY LENGTH	KEY LENGTH
BOUNDARY- FOLDING	SHIFT-FOLDING

5 ـ تحوير الأساس Radix Transformation or Lins Method

وتتلخص هذه الطريقة بالخطوات التالية :

أ ـ يعتبر المفتاح العددي المعطى مكتوبا بأساس معين

ب ـ يستخرج له أساس جديد

جـ ـ يحول الى الأساس العشري عن طريقة جمع حاصل ضرب الأساس الجديد بمراتب العدد بعد رفعها الى تسلسل المرتبة اي طبقا للقاعدة .

DEC=

حيث الاساس الجديد =R

تسلسل المراتب =I

اذا كانت المرتبة هي الاحاد فان I قيمتها (1) واذا كانت العشرات فقيمتها (2)

وهكذا الرقم من المرتبة KI = I

المقدار العشري الناتج =DEC

د ـ يقسم المقدار العشري الناتج على عدد صحيح موجب (قاسم) يكون قريبا الى حجم الجدول النحتي اما كيفية استخراج الأساس الجديد وتحديد العدد القاسم للمقدار العشري فيكون بالخطوات التالية :

بـ1 ـ يؤخذ حجم الجدول النحتي وتستخرج أقرب قيمة له شرط أن تكون من قوى (Power) عدد معين حيث يعتبر الأساس الأولي .

بـ2 ـ يستخرج الأساس الجديد عن طريق اضافة (1) الى الأساس الأولي وقد جاءت قيمة الأساس الجديد طبقا لهذه الفرضية من قبل التجارب المطبقة حيث وجد أن أفضل أساس هو الذي يزيد بمقدار (1) عن الأساس الأولي [7] .

وكما نلاحظ من الشكل رقم (2) فان أساليب معالجة ما يفيض من محافن المساحة الرئيسية تكون اما من خلال المساحة الرئيسية نفسها او من خلال مساحة الفيض وفي حالة استعمال السلسلة (Chaining) لمعالجة التضارب نلاحظ بأنه يجب اضافة مؤشر لكل محفن في الشكل المذكور .

ويمكن تلخيص طرق معالجة التضارب عموما كما يلي : ـ

1 ـ الجس الخطي Linear Probing

وتسمى هذه الطريقة أحيانا بالفيض المفتوح (Open Overflow) او العنونة المفتوحة (Open Addressing) ويمكن تلخيصها بانه كلما يمتلىء محفن فان الفائض يدخل في اول محفن شاغر (يحوى شقب شاغر) مجاور وقد يبدأ البحث في المحفن التالي الذي يليه وهكذا او في المحفن السابق ثم الذي يسبقه وعلى أية حال فان جدول النحت يعتبر دائريا كما في كافة الطرق أي ان الوصول الى نهايته يعيدنا الى بدايته والتوقف يكون عند العودة الى المحفن الذي انطلقنا منه .

2 ـ الجس العشوائي Random Probing

وتسمى أيضا باعادة النحت (Rehashing) وفي هذه الطريقة يتم توليد سلسلة من العناوين النحتية $h_0(K), h_1(K), h_2(K)$..... etc وتنتهي هذه السلسلة عند العودة الى نقطة البداية وتستعمل هذه العناوين تباعا لايجاد محفن شاغر .

وتوجد عدة طرق من الجس العشوائي تختلف فيما بينها وبصورة رئيسية في طرق توليد سلسلة العناوين النحتية وسوف نكتفي بالاشارة الى المراجع الغنية بتفاصيلها .

أ ـ طريقة النحت التربيعية Quadratic Hash Method [9,8,3]

ب ـ طريقة حاصل القسمة التربيعية Quadratic Quotient Method [6]

جـ ـ طريقة المتبقي التربيعية Quadratic Residue Method [10]

د ـ طريقة الاعداد شبه العشوائية Pseudo Random Numbers Method [4,1,3]

هـ ـ طريقة حاصل القسمة الخطية Linear Quotient Method [5]

3 ـ الجس الفيضي Overflow Probing

تعرف مساحة الفيض (Overflow Area) بأنها تلك المساحة الاضافية التي يوضع فيها المفتاح متى ما تضارب مع مفتاح آخر في الجدول النحتي الاصلي ويربط الموقع في الجدول النحتي مع الموقع الجديد في مساحة الفيض باستعمال رابط(Link)، لذلك نلاحظ في بداية تكوين الجدول النحتي تعتبر مساحة الفيض فارغة .

أما حجم مساحة الفيض فإن تحديده ليس من الأمور السهلة وفي أكثر الأحيان يعتمد على التجربة الفعلية .

وهناك بعض الاحصائيات العملية التي تظهر العلاقة بين عامل التحميل والنسبة المئوية لحجم مساحة الفيض الى المساحة الرئيسية [2]

تختلف طريقة السلسلة المتداخلة (Internal Chaining) عن الطريقة أعلاه بعدم استعمال مساحة فيض وربط المفاتيح الفائضة ضمن نفس مساحة الجدول النحتي الأصلي ولقد لوحظ [1] بأن هاتين الطريقتين تمثلان أكفأ الطرق ولذا تمت تجربتهما .

معالجة المفاتيح الغير عددية (Non- numeric Keys)

عادة ما تكون المفاتيح المعطاة في صيغ غير عددية ونظرا لكون معظم دالات النحت حسابية فان ذلك يتطلب تحويلها الى اعداد صحيحة .

ولقد لوحظ [7] بان أداء مختلف طرق تحويل المفاتيح الغير عددية الى عددية متشابه طالما نحافظ على انفرادية (Uniqueness) تجفير كل رمز . يبين الجدول رقم(2) نماذج من مفاتيح الملفات النموذجية ولقد تم استعمال الطرق التالية في التحويل الى الصيغ العددية .

ـ بالنسبة لملف اسماء مشاركي الهواتف وكذلك ملف أرقام الواردة تم استعمال طريقة اعتماد الجفرة الداخلية (Intern,al Code) لكل رمز ثم اجراء العملية المنطقية .

« اما الاستثنائية » (Execlusive Or) بين الكلمات المحتوية على الاسم واعتبار الناتج المفتاح العددي المقابل .

2 ـ بالنسبة لملف ارقام تصنيف الكتب فلقد تم تجفير الحروف(A,B,C, ----Z) كاعداد 36 و13 و12 و11 على التوالي وابقاء الأرقام كما هي مع اهمال النقط الفاصلة .

3 ـ بالنسبة لملف أسماء العمال وهو أضخم الملفات فلقد اعتمدت طريقة تجفير الحروف العربية رقميا فمثلا حرف (أ) اعتبر (01) والباء (02) وهكذا الياء(28) واعتبر الفراغ (00) وبما أن الحقل المخصص يبلغ طوله 30 حرف فلقد تم الحصول على سلسلة رقمية لكل اسم طولها (60) حرف استعملت معها طريقة (الطي الازاحي) لتقليصها الى سعة كلمة الحاسبة (36 Bits) .

لقد وجد بأن تحويل المفاتيح الغير عددية الى عددية بالطرق اعلاه وبالنسبة لهذه التجربة لم يؤ دي الى أي تضارب في النواتج . أما تأثير هذه الطرق احصائيا على توزيع المفاتيح المعطى فلم يدرس ضمن هذا البحث وفي الواقع فان هذا الجانب يشكل بحثا مهما بحد ذاته ويوصى باجراءه وخصوصا بعد التوصل الى جفرة رموز (Character Set) عربية موحدة في التطبيق .

1 ـ ملف ارقام الهواتف	7713948
2 ـ ملف اسماء مشاركي الهواتف	ابتسام خليل ابراهيم
3 ـ ملف أرقام تصنيف الكتب	019.330.A.639
4 ـ ملف ارقام الواردة	23457-1-2-106-4 G
5 ـ ملف اسماء العمال	حميد غفوري عبد الجبار المفرجي

جدول (2) نماذج من مفاتيح الملفات

مناقشة النتائج

تمثل المخططات البيانية المرفقة من (م ـ 1) الى (م ـ 4) نتائج البحث بالنسبة لتقييم طرق النحت وللملفات :

أرقام الهواتف واسماء مشاركي الهواتف وأرقام تصنيف الكتب وأرقام الواردة على التوالي

أما الجدول (جـ ـ 5) والمخطط البياني (م ـ 5) فيمثلان النتائج بالنسبة لملف اسماء العمال الجدولان (جـ ـ 6) و (جـ ـ 7) يوضحان نتائج طرق معالجة التضارب لملفي اسماء العمال وأسماء مشاركي الهواتف .

لقد تم اعتماد المؤشرات التالية في التقييم :

N : حجم أكبر تضارب (Maximum Collision Size) وهو عدد المفاتيح المتضاربة في أسوأ حالة تضارب

T: المجموع الكلي للتضارب (Total Number of Collision) وهو مجموع كافة حالات التضارب للملف المعين .

VA : معدل حالات التضارب للمفتاح الواحد (Average Number of Collision Per Key) وهو نسبة المجموع الكلي للتضارب الى عدد المفاتيح المعطى .

من خلال دراسة نتائج التجربة يمكن استخلاص نقاط المناقشة التالية :

1 ـ الملفات العددية والمختلطة

نلاحظ من المخططات البيانية (م ـ1) و (م ـ 3) و (م ـ 4) ما يلي :

أ ـ ان أداء طريقة متوسط المربع ضعيف عموما حيث أعطت قيما عالية للمجموع الكلي للتضارب مما يدل على عدم ملائمتها .

ب ـ يلاحظ تفوق طريقة القسمة وخصوصا لملف ارقام الهواتف ولعوامل التحميل المعتدلة اي 70% -85% كما تبرز طريقة التشذيب بالمرتبة الثانية حيث اعتمد فيها تشذيب (4) مراتب من يسار قيمة المفتاح وكانت المراتب الثلاثة الباقية اكثر عشوائية في التوزيع وهذا أمر متوقع في أرقام الهواتف .

جـ ـ عند مفاضلة طرق النحت التي تعتمد على الطي فان اداء طريقة الطي الازاحي افضل من الطي التخمي وبشكل بارز .

د ـ لقد أعطت طريقة تحوير الأساس أفضل نتائج لها باستعمال الأساس (S) ومع ذلك فان أدائها لا يصل الى حد التوصية خصوصا وان دالة تحوير الأساس تستغرق وقتا طويلا للاحتساب بالمقارنة مع الطرق الأخرى .

2 ـ الملفات الالفبائية (الاسماء)

من ملاحظة المخطط البياني رقم (م ـ2) والجداول (ج ـ 5) الى (جـ ـ7) وكذلك المخطط البياني (م ـ5) يمكننا الخروج بما يلي ولا بد من التأكيد هنا بأن كون الدالات التحتية تعمل أساسا على الصيغ العددية للمفاتيح يجعل لطريقة تحويل المفاتيح من غير عددية الى عددية تأثير جذري على توزيع المفاتيح المعطى وعليه فان النقاط أدناه مبنية على طرق التحويل المختارة والموضحة في معالجة المفاتيح الغير عددية سابقا .

أ ـ ان اداء طريقة تحوير الاساس وخصوصا لملف اسماء العمال يعتبر رديئاً مما يؤ يد استبعادها .

ب ـ المخطط البياني (م ـ5) يبين أن أفضل طريقة تحت للاسم العربي هي الطي الازاحي ولا يبدو بأن طريقة القسمة هي الطريقة المثلى . أما طريقة التشذيب فلقد أعطت قيما عالية نسبيا للمؤشر (M) ولم تدخل ضمن هذا المخطط البياني .

إن تفوق طريقة الطي الازاحي تتضح حتى من خلال المؤشرين (AV) و (M)

جـ ـ ان كافة الاحتمالات المجربة في النحت لقيم عوامل التحميل ولحجوم الملفات

المعطاة تشير بأن أفضل سعة للمحفن تستعمل في التصميم هي (4) او (5) مواقع ومن الطبيعي هناك عوامل أخرى تدخل في هذا الاختيار وذلك مثل حجم القيد الواحد وأفضليات النظام الفعلي ، هل هي السرعة ام الاقتصاد بالخزن وهكذا . ومن الجدول (جـ-6) نلاحظ أن سعة (4) تعتبر مناسبة حتى لعوامل التحميل العالية .

د - الجدول (جـ-7) يوضح أفضلية طريقة السلسلة بمساحة فيض على السلسلة المتداخلة لمعالجة التضارب ومن الواضح أن ذلك يكون على حساب الخزين

الاستنتاجات

1 - ان نتائج البحث هذا جاءت مؤيدة لاستنتاجات العلماء والبحوث العالمية السابقة مثل (Lum, Yuen, Dodd) [7] وهي أننا ولغرض معرفة طريقة التحت المثلى لتطبيق معين نحتاج غالبا لتجربة عدة طرق على الملف المعطى . وعلى ضوء تباين طبيعة الملفات المختارة في التجربة يمكن القول بأن هذا البحث يخرج بتوصية تجربة مجموعة معينة من الطرق وبتسلسل معين يستبعد الطرق التي أثبتت اداء ضعيفا لمختلف الملفات والتي تتميز أيضاً ببطء احتساب دالالتها التحتية .

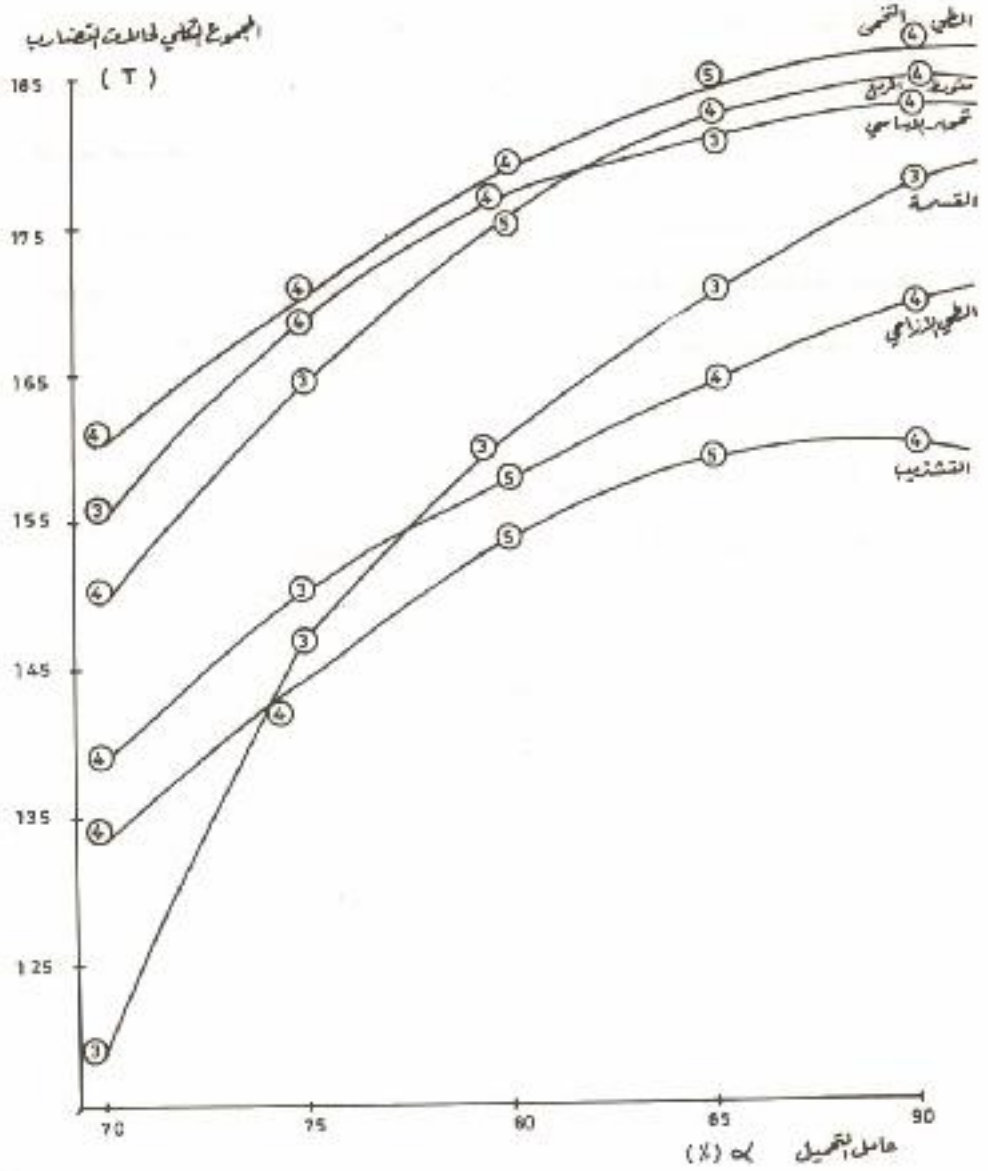
2 - لقد وجد من السهل تحويل المفاتيح الغير عددية الى الصيغة العددية بالطرق الاعتيادية كالتجفير الرقمي للرموز واستعمال الدالات المنطقية والطي وبدون حدوث مرادفات (Synonyms) الا ان هذا التحويل يتطلب دراسة مفصلة للتأثير على التوزيع الاحصائي ومدى المفاتيح وعليه فان استنتاجات هذا البحث مبنية على الطرق المختارة في التحويل البحث لا يشمل دراسة الملفات المتغيرة الحجم .

3 ـ بالنسبة للملفات العددية والمختلطة أثبتت طريقة القسمة جودة أدائها وكذلك طريقة التشذيب . ويفضل تجربة الطي الازاحي قبل التخمي ويستحسن استبعاد طريقة متوسط المربع وتحوير الاساس .

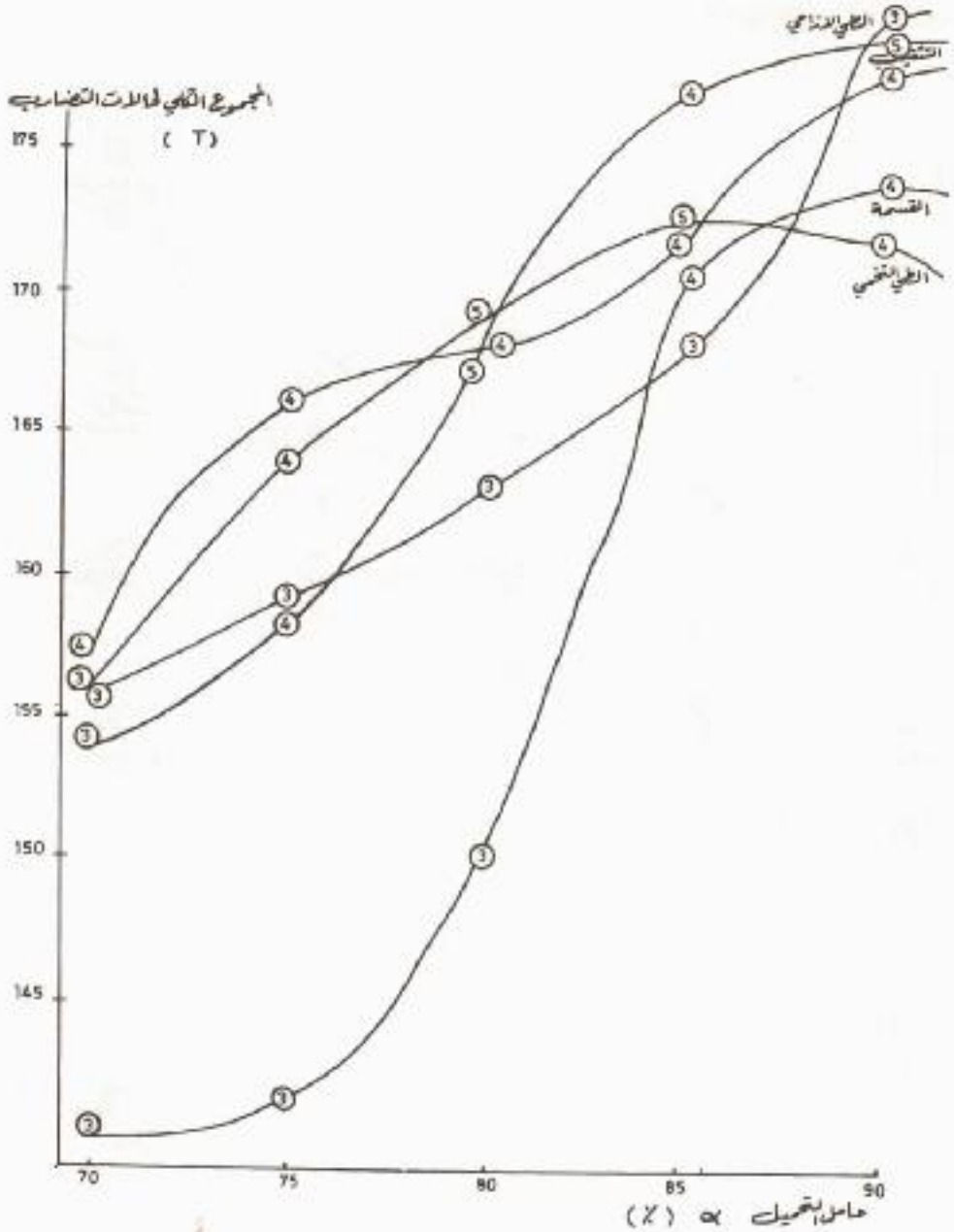
4 ـ بالنسبة للملفات الالفبائية ، فان البحث يوصي بالتسلسل التالي لتجربة الطرق : الطي الازاحي ثم التحتي ثم القسمة ثم متوسط المربع وينصح باستبعاد طريقة تحوير الاساس . أما التشذيب فلا يوصي بتجربتها خصوصا للتطبيقات التي تركز على سرعة استرجاع المعلومات بدلا من مساحة التخزين .

5 ـ بالنسبة لطرق معالجة التضارب فان البحث يؤيد استنتاجات (Knuth, [1]) بأن طريقة السلسلة بمساحة فيض تفضل على السلسلة المتداخلة عندما تكون السرعة في استرجاع القيود هي العامل الاول .

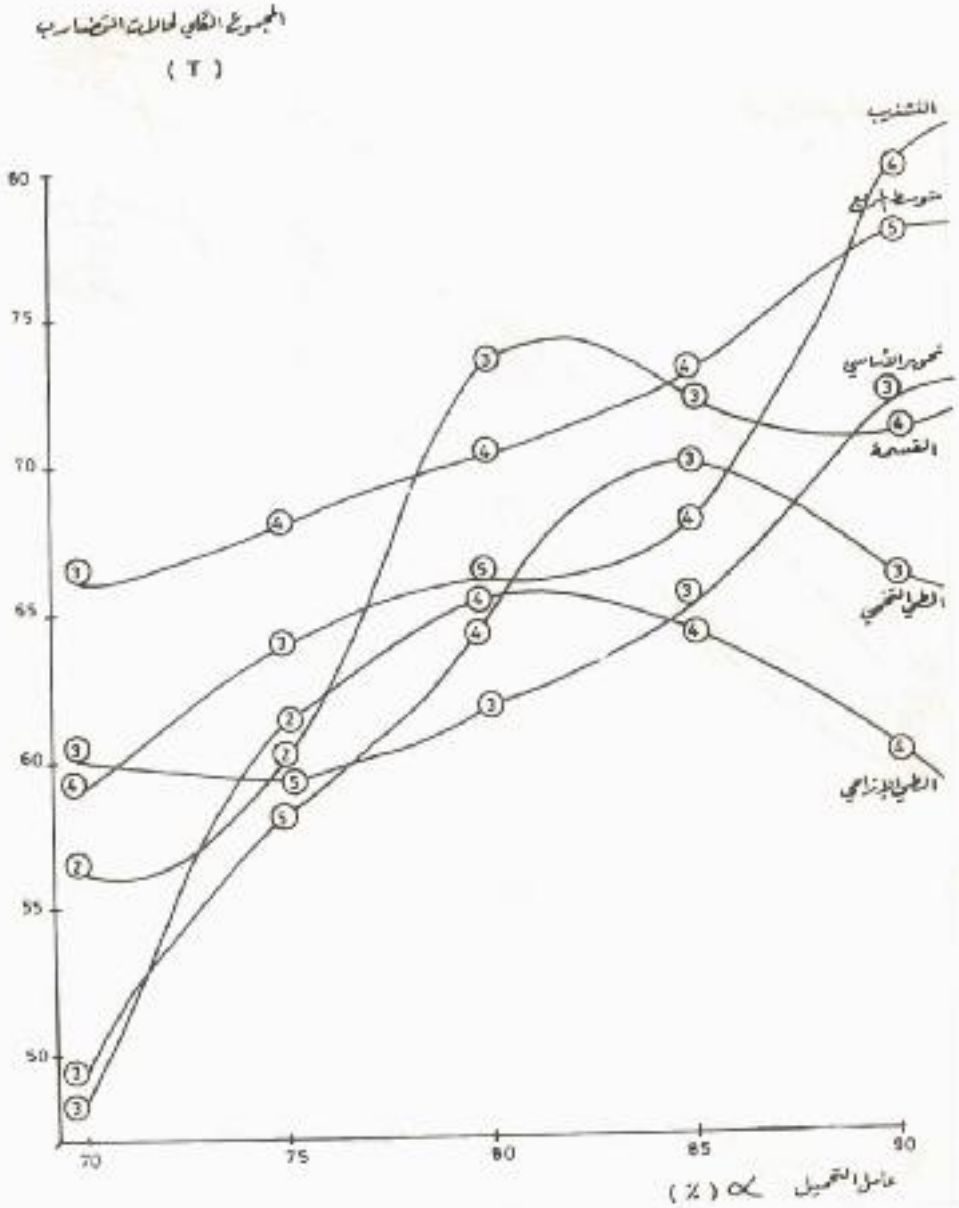
وعند التعامل مع الملفات المتوسطة الحجوم ، أي بضعة مئات من القيود فان اختيار سعة المحفن (4 أو 5) مواقع يعتبر مناسبا .



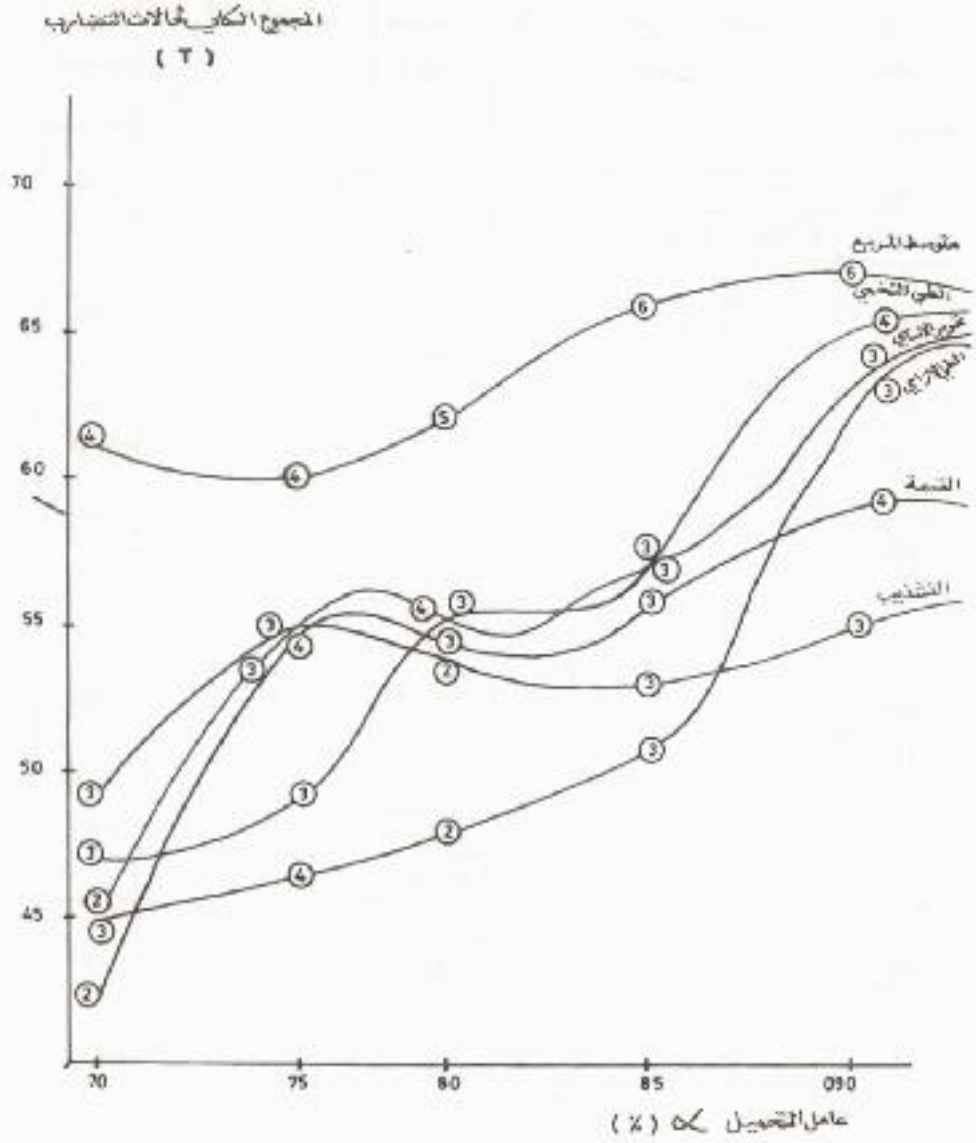
مخطط بياني رقم (م ـ 1) يبين المجاميع الكلية لحالات التضارب مقابل القيم المجربة لعامل التحميل بالنسبة لملف ارقام الهوائف (قيم M مبينة في الدوائر على المنحنيات اعلاه)



مخطط بياني رقم (م 2) يبين المجاميع الكلية لحالات التضارب مقابل القيم المجربة لعامل التحميل بالنسبة للف اسماء مشاركي الهواتف .



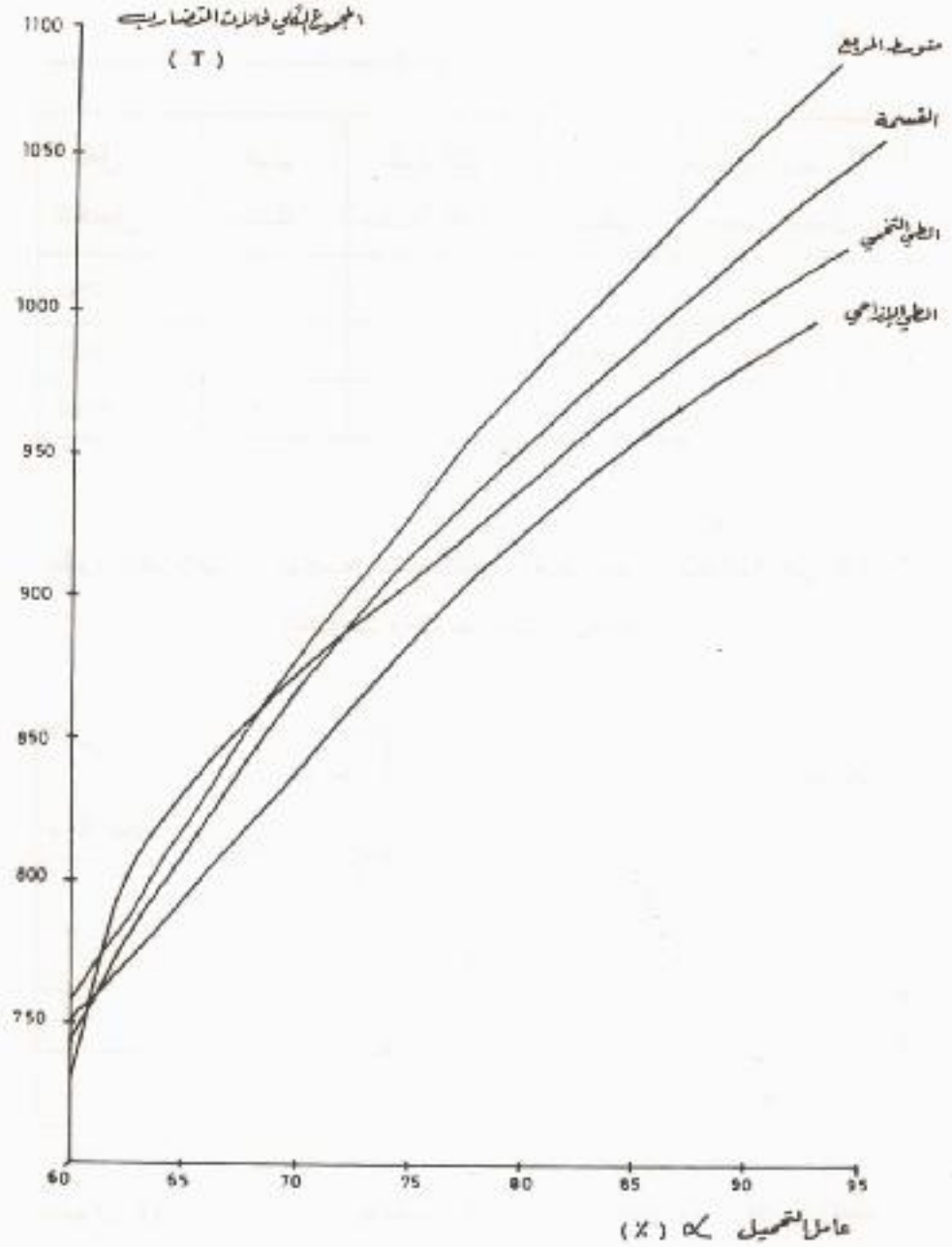
مخطط بياني رقم (م-3) يبين المجاميع الكلية لحالات التضارب مقابل القيم المجربة لعامل التحميل بالنسبة للصف أربعة تصنيف الكتب



مخطط بياني رقم (م ـ ٤) يبين المجاميع الكلية لحالات التضارب مقابل القيم المحربة لعامل التحميل بالنسبة لنفس أرقام الموازنة

عامل التحميل	طرق النحت / القياس	القسمة	متوسط المربع	الطي الازاحي	الطي التخمي	التشذيب	تحوير الأساس
%60	M	5	4	4	5	5	5
	T	741	756	749	728	772	1245
	AV	0.248	0.253	0.251	0.244	0.259	0.417
%65	M	5	4	4	4	6	9
	T	809	817	789	830	836	1696
	AV	0.271	0.274	0.264	0.278	0.280	0.569
%70	M	4	5	4	4	5	9
	T	866	878	850	838	897	1564
	AV	0.290	0.294	0.285	0.281	0.301	0.586
%75	M	4	5	4	5	5	10
	T	910	911	884	900	938	1812
	AV	0.305	0.305	0.296	0.302	0.314	
%80	M	5	5	4	5	6	10
	T	989	975	935	939	958	1857
	AV	0.331	0.327	0.313	0.315	0.321	0.623
%85	M	5	5	5	6	5	8
	T	989	1003	954	990	1002	1899
	AV	0.331	0.336	0.320	0.332	0.336	0.637
%90	M	5	5	4	6	5	10
	T	1044	1043	1010	1016	1036	1946
	AV	0.350	0.350	0.338	0.340	0.347	0.653
%95	M	5	5	5	6	5	9
	T	1064	1100	1100	1055	1070	2057
	AV	0.357	0.357	0.369	0.359	0.359	0.690

جدول رقم (جـ 5) يوضح النتائج بالنسبة للف أسماء العمال بمؤسسة التقاعد والضمان الاجتماعي



مخطط بياني رقم (م 5) يبين المجاميع الكلية لحالات اضطراب مقابل القيم المختلفة لعامل التحميل بالنسبة للقلب لسيناء العمان

عامل التحميل	اطول سلسلة	طول اكثر سلسلة تكرارا	مساحة الفيض	مساحة الفيض% حجم الجدول
%50	5	2	629	11
%75	5	3	849	22
%90	5	4	975	30

جدول رقم (جـ-6) يوضح نتائج التجربة بطريقة الربط الخارجي لمعالجة التضارب وبالنسبة لملف اسماء العمال

الطريقة / سعة المحفن بالشقوب	السلسلة المتداخلة	السلسلة بمساحة فيض
1	2.350	1.832
2	2.990	2.336
3	3.620	3.124
5	4.580	3.482

جدول رقم (جـ-7) يبين قيم معدل عدد عمليات الوصول (AVAC) لطرق معالجة التضارب ولعامل التحميل 90% وبالنسبة لملف اسماء مشاركي الهواتف

برنامج اشتقاق العناوين النحتية

يعتبر هذا البرنامج من أهم برامج البحث لكونه يتناول تطبيق الدالات النحتية المختلفة واشتقاق العناوين النحتية من المفاتيح المعطاة . والنسخة المرفقة من البرنامج تتعلق بملف اسماء العمال حيث يتم توليد العناوين النحتية للمفاتيح الاسمية بعد أن تم تحويلها الى الصيغة العددية . تستخرج اولا قيمة اقرب عدد اولي لحجم الجدول النحتي ومن ثم تستدعي الدوال الستة (دالة القسمة ، دالة متوسط المربع ، دالة الطي الازاحي ، دالة الطي التحمي ، دالة التشذيب ، دالة تحوير الاساس) لتكوين العنوان النحتي بهذه الطرق المختلفة وقد تم استخدام طريقة التقييس (Scaling Method.) لكل دالة من هذه الدوال كي يكون العنوان المستخرج ضمن حيز الجدول النحتي . لغرض احتساب تضارب المفاتيح يعتمد عداد تضارب لكل مفتاح تزداد قيمته بمقدار (1) عند كل تضارب وهكذا يستخرج عدد حالات التضارب لكل مفتاح بكل طريقة ويستخرج أيضا المجموع الكلي لحالات التضارب واكبر تضارب لكل طريقة .

أما عن المتغيرات المستعملة فهي : -

KEY(3000) : نسق المفاتيح التي تقرأ من الشريط المغناطيسي

DVC(HDV) : عداد خاص للعناوين المشتقة بطريقة القسمة

MQC(HMQ) : عداد خاص للعناوين المشتقة بطريقة متوسط المربع

SFC(HSF) : عداد خاص للعناوين المشتقة بطريقة الطي الازاحي

BFC(HBF) : عداد خاص للعناوين المشتقة بطريقة الطي التحمي

TRC(HTR) : عداد خاص للعناوين المشتقة بطريقة التشذيب

RDC(HRD) : عداد خاص للعناوين المشتقة بطريقة تحوير الاساس

IDVSR : اقرب عدد أولي لحجم الجدول التحتي

NKEYS : اسم لعدد المفاتيح الموجودة

NSIZE : اسم لحجم الجدول التحتي

RR : يمثل مدى المفاتيح المعطاة

ISUMDV : يمثل المجموع الكلي لحالات التضارب بطريقة القسمة

ISUMMO : يمثل المجموع الكلي لحالات التضارب بطريقة متوسط المربع

ISUMSF : يمثل المجموع الكلي لحالات التضارب بطريقة الطـي الازاحي

ISUMBF : يمثل المجموع الكلي لحالات التضارب بطريقة الطي التخمي

ISUMTR : يمثل المجموع الكلي لحالات التضارب بطريقة التشذيب

ISUMRD : يمثل المجموع الكلي لحالات التضارب بطريقة تحوير الاساس

أما عن الدوال المستخدمة لاشتقاق العناوين التحتية :

1 ـ طريقة القسمة : ـ

ويستدعي الدالة بالاسم IDVISN

يوضع العنوان التحتي لها في الخلية HDV

K : ويمثل المفتاح الذي قرأ في البرنامج الرئيسي

2 ـ طريقة متوسط المربع : ـ

ويستدعي دالة هذه الطريقة باسم MIDSQR

ويوضع العنوان التحتي لها في الخلية HMQ

K : ويمثل المفتاح

KL : ويمثل طول المفتاح

KHL : ويمثل طول العنوان النحتي المطلوب على أن لا يتعدى طول حجم الجدول

SQR : قيمة مربع المفتاح

IB1- IB2 : ويمثل حدود الجزء الوسطي لمربع المفتاح

: قيمة مربع المفتاح .

: ويمثل حدود الجزء الوسطي لمربع المفتاح

3 ـ طريقة الطي الازاحي : ـ

وتستدعي الدالة بالاسم ISFOLD

ويوضع العنوان النحتي لها في الخلية HSF

ND : طول الطيه الواحدة

NF : عدد الطيات المطلوبة

4 ـ طريقة الطي التخمي : ـ

وتستدعي الدالة بالاسم IBFOLD

ويوضع العنوان النحتي لها في الخلية HBF

FD (1) : ويمثل طيات الرقم

DGT (1) : وتمثل المرتبة من الطيه

IB1 : تمثل الطيه الاولى في الرقم

IB2 : تمثل الطيه الثانية في الرقم

IB3 : تمثل الطيه الأخيرة في الرقم

5 ـ طريقة التشذيب

وتستدعي الدالة باسم ITRUNC

ويوضع عنوانها النحتي في الخلية HTR

K1-K2 : وتمثل الحدود الوسطية للرقم

6 ـ طريقة تحوير الاساس

وتستدعي الدالة باسم RADIX

ويوضع عنوانها النحتي في الخلية HRD

QM : وتمثل 2M اي أقرب رقم لحجم الجدول النحتي ناتج من مربعات او مكعبات عدد معين .

أ ـ تمثل الاساس الجديد حيث أن أ تساوي قيمة2 مضافا اليها واحد وهناك دالة الى مجموع عدد حالات التضارب لكل مفتاح في كل طريقة تستدعي هذه الدالة بالاسم ISUM

ونرسل اليها في بادىء الأمر قيمة عداد مجموع التضارب الذي تكون قيمته للمفتاح الواحد (1) ولكن عند حصول تضارب مع هذا المفتاح تزداد قيمته .

تم تم احتساب اكبر تضارب بكل دالة من الدوال الستة حيث :

IDV : يمثل اكبر تضارب بدالة القسمة

MBG : يمثل أكبر تضارب بدالة متوسط المربع

ISF : يمثل اكبر تضارب بدالة الطي الازاحي

IBF : يمثل أكبر تضارب بدالة الطي التخمي

ITR : يمثل اكبر تضارب بدالة التشذيب

IRD : يمثل أكبر تضارب بدالة تحوير الاساسي

```
010$    IDENT     OTINGR,MAOK,***LOADING FACTOR IS 0.70***
020$    OPTION  FORTRAN
030$    FORTRAN NDECK
040C
050C          ************************************
060C          *  **  PROGRAM  HASH 1         **    *
070C          *    THIS PROGRAM WILL GENERATE THE *
080C          *    HASH ADDRESSES FOR ARABIC NAME *
090C          *    IN SIX DIFFERENT METHODS ****  *
100C          *                                   *
110C          ************************************
120C
130C          *******      MAIN  PROGRAM      ******
140C
150         DIMENSION KEY(3000),MQC(4500)
160         INTEGER DVC(4500),SFC(4500),BFC(4500),RDC(4500),TRC(4500)
170         INTEGER HDV,HMQ,HSF,HBF,HRD,HTR
180         NKEYS=2980
190         NSIZE=4257
200         IDVSR=4253
210C        LOOP TO CLEAR COUNTERS
220         DO 10 I=1,NSIZE
230         DVC(I)=0
240         MQC(I)=0
250         SFC(I)=0
260         BFC(I)=0
270         RDC(I)=0
280         TRC(I)=0
290   10    CONTINUE
300         WRITE(6,35)
310   35    FORMAT(4X,"KEY",8X,"DIVISION",3X,"MIDSQUARE",3X,"SHIFT-FOLD",2X
320        &,"BOUNDARY-FOLD",3X,"TRUNCATION",3X,"RADIX")
330         WRITE(6,40)
340   40    FORMAT(12("-"),3X,8("-"),3X,9("-"),3X,10("-"),2X,13("-"),3X,10(
350        &"-"),3X,5("-"))
360C          -----LOOP FOR GENERATING HASH ADDRESSES-------
370         DO 70 I=1,NKEYS
380         READ(08,50)KEY(I)
390   50    FORMAT(I10)
400         K=KEY(I)
410         KL=10
420         KHL=4
430         RR=9999.
440         P=3.
450         QM=4096.
460         HDV=IDVISN(K,IDVSR)
470         DVC(HDV)=DVC(HDV)+1
480         HMQ=MIDSQR(K,KL,KHL)
490         HMQ=HMQ+(FLOAT(NSIZE)/RR)
500         MQC(HMQ)=MQC(HMQ)+1
510         HSF=ISFOLD(K,KL,5)
520         HSF=HSF+(FLOAT(NSIZE)/RR)
530         SFC(HSF)=SFC(HSF)+1
540         HBF=IBFOLD(K,KL,5)
550         HBF=HBF+(FLOAT(NSIZE)/RR)
560         BFC(HBF)=BFC(HBF)+1
570         HTR=ITRUNC(K,KL,10,6)
580         HTR=HTR+(FLOAT(NSIZE)/RR)
590         TRC(HTR)=TRC(HTR)+1
```

```
      HRD=RADIX(K,KL,P,QM)
      HRD=HRD*(FLOAT(NSIZE)/RR)
      RDC(HRD)=RDC(HRD)+1
      WRITE(6,60)KEY(I),HDV,HMQ,HSF,HBF,HTR,HRD
   60 FORMAT(1X,110,6X,I4,7X,I4,9X,I4,9X,I4,11X,I4,6X,I4)
   70 CONTINUE
      WRITE(6,80)
   80 FORMAT(1H ,1X,"HASH ADDRESS",5X,"DV-COUNTER",2X,"MQ-COUNTER",2
     & X,"SF-COUNTER",2X,"BF-COUNTER",2X,"TR-COUNTER",2X,"RD-COUNTER")
      WRITE(6,90)
   90 FORMAT(1X,15("-"),5X,10("-"),2X,10("-"),2X,10("-"),2X,10("-"),2X
     &,10("-"),2X,10("-"))
      ISUMDV=0
      ISUMMQ=0
      ISUMSF=0
      ISUMBF=0
      ISUMTR=0
      ISUMRD=0
      DO 110 I=1,NSIZE
      DVC(I)=ISUM(DVC(I))
      MQC(I)=ISUM(MQC(I))
      SFC(I)=ISUM(SFC(I))
      BFC(I)=ISUM(BFC(I))
      TRC(I)=ISUM(TRC(I))
      RDC(I)=ISUM(RDC(I))
      WRITE(6,100)I,DVC(I),MQC(I),SFC(I),BFC(I),TRC(I),RDC(I)
  100 FORMAT(5X,I4,12X,I4,8X,I4,8X,I4,8X,I4,8X,I4,8X,I4)
      ISUMDV=ISUMDV+DVC(I)
      ISUMMQ=ISUMMQ+MQC(I)
      ISUMSF=ISUMSF+SFC(I)
      ISUMBF=ISUMBF+BFC(I)
      ISUMTR=ISUMTR+TRC(I)
      ISUMRD=ISUMRD+RDC(I)
  110 CONTINUE
      WRITE(6,120)
  120 FORMAT(18X,10("-"),2X,10("-"),2X,10("-"),2X,10("-"),2X,10("-"),2
     &X,10("-"))
      WRITE(6,130)ISUMDV,ISUMMQ,ISUMSF,ISUMBF,ISUMTR,ISUMRD
  130 FORMAT(2X,"NO OF COLLISION =",2X,I4,8X,I4,8X,I4,8X,I4,8X,I4,8X,I4)
      IDV=DVC(1)
      MBG=MQC(1)
      ISF=SFC(1)
      IBF=BFC(1)
      ITR=TRC(1)
      IRD=RDC(1)
      DO 20 J=2,NSIZE
      IF(DVC(J).GT.IDV) IDV=DVC(J)
      IF(MQC(J).GT.MBG) MBG=MQC(J)
      IF(SFC(J).GT.ISF) ISF=SFC(J)
      IF(BFC(J).GT.IBF) IBF=BFC(J)
      IF(TRC(J).GT.ITR) ITR=TRC(J)
      IF(RDC(J).GT.IRD) IRD=RDC(J)
   20 CONTINUE
      WRITE(6,25)IDV,MBG,ISF,IBF,ITR,IRD
   25 FORMAT(1X,"MAX NO OF COLLISION=",1X,I5,8X,I4,8X,I4,8X
     &,I4,8X,I4)
      STOP
      END
C------DIVISION METHOD---------
```

```
1275          FUNCTION IDVISN(K,IDVSR)
1280          IDVISN=MOD(K,IDVSR)+1
1285          RETURN
1290          END
1295C  ------MID SQUARE METHOD---------
1300          FUNCTION MIDSQR(K,KL,KHL)
1305          DOUBLE PRECISION SQR
1310          SQR=DABS(K**2)
1315          IF(MOD(KL,2).EQ.0) GO TO 10
1320          IB1=KL-(KHL/2)-KHL
1325          GO TO 20
1330    10    IB1=KL-(KHL/2)-KL/2
1335    20    IB2=KL+(KHL/2)-KL/2
1340          SQR=DMOD(SQR,10.0**IB2)
1345          MIDSQR=SQR/10**IB1
1350          RETURN
1355          END
1360C  ------SHIFT FOLD METHOD---------
1365          FUNCTION ISFOLD(K,KL,ND)
1370          INTEGER SFD(11)
1375          NF=KL/ND
1380          NKL=KL
1385          K1=K
1390          DO 10 I=1,NF
1395          SFD(I)=K1/10**(NKL-ND)
1400          K1=MOD(K1,10**(NKL-ND))
1405    10    NKL=NKL-ND
1410          ISFOLD=K1
1415          DO 20 I=1,NF
1420    20    ISFOLD=ISFOLD+SFD(I)
1425          IF(ISFOLD.GE.10**4) ISFOLD=MOD(ISFOLD,10**4)
1430          RETURN
1435          END
1440C  ------BOUNDARY FOLD METHOD---------
1445          FUNCTION IBFOLD(K,KL,ND)
1450          INTEGER FD(11),DGT(11)
1455          NF=KL/ND
1460          K1=K
1465          DO 10 J=1,KL
1470          DGT(J)=K1/10**(KL-J)
1475          K1=MOD(K1,10**(KL-J))
1480    10    CONTINUE
1485          IB1=0
1490          IB2=0
1495          IB3=0
1500          DO 20 J=1,ND
1505          IB1=IB1*10+DGT(ND+1-J)
1510          IB2=IB2*10+DGT(ND+J)
1515    20    CONTINUE
1520          N1=KL-(NF*ND)
1525          DO 30 I=1,N1
1530          IB3=IB3*10+DGT(KL+1-J)
1535    30    CONTINUE
1540          IBFOLD=IB1+IB2+IB3
1545          IF(IBFOLD.GE.10**4)IBFOLD=MOD(IBFOLD,10**4)
1550          RETURN
1555          END
1560C  ------TRUNCATION METHOD---------
1565          FUNCTION ITRUNC(K,KL,K1,K2)
```

```
1570          NEWK=MOD(K,10**(K1))
1575          NEWK=NEWK/10**K2
1580          ITRUNC=NEWK
1585          RETURN
1590          END
1595C  ------RADIX METHOD---------
1600          FUNCTION RADIX(K,KL,P,QM)
1605          RADIX=0.0
1610          K1=K
1615          DO 10 J=1,KL
1620          RADIX=RADIX+(K1/10**(KL-J)*P**(KL-J))
1625    10    K1=MOD(K1,10**(KL-J))
1630          RADIX=AMOD(RADIX,QM)
1635          RETURN
1640          END
1645C  ------ACCUMILATE THE COLLISION---------
1650          FUNCTION ISUM(K)
1655          ISUM=0
1660          IF(K.LE.1)GO TO 10
1665          ISUM=K-1
1670  10      RETURN
1675          END
1680$  EXECUTE
1685$  LIMIT    100,50K,,20000
1690$  TAPE     08,X1D,,AA11
1695$  ENDJOB
```

المصادر

REFERENCES

1- D.E. KNUTH
The art of Computer Programming. Sorting و Searching
VOL. 3
ADDISON WESLEY 1973

2- Severance, D. and R. Duhne: practitioners Guide to Addressing Algorithms Comunications of the ACM, VOL. 19, June 1976, PP. 314-326

3- DAVID GRIES
«Compiler Construction for Digital Computers».
John Wiley و Sons 1971

4- MORRIS R.
«Scatter Storage Techinques
C.A.C.M. Jan-1968

5- BELL KOMAN
«Linear quotient hash Code»
C.A.C.M. NOV. 1970

6- BELL
(The quadratic quotient Method)
C.A.C.M. Feb. 1970

7- Y.Y.LUM, P.S.T. YUEN, و M. DODD
(Key to address transformation Techniques)
C.A.C.M April 1971

8- Hopgood, F.R. and J.Davenport
(The Quadratic Hash Method When the Table Size is apower of 2) Compute Journal, 15, 4 November 1972 PP. 314-315.

9- Day, J.C.
(Full Table Quadratic Searching For Scatter Storage)
Communication of the ACM, 13, 8, Aug- 1970, PP. 481-482

10- Radke, C.E.
(The use of Quadratic Residue Research)
Communication of the ACM. 13, 2, Feb. 1970, PP. 103-107.