تنظيم الملفات باستخدام الأسلوب العشوائي

هيثم يوسف محمد
الجامعة التكنولوجية

التنظيم العشوائي (المباشر) Direct (Random) Organization

1 – التنظيم العام : – General Organization مع التنظيم المباشر توجد علاقة حتمية بين المفتاح (Key) والعنوان (Addres,) لقيد (Record) معين وهذه العلاقة تسمح بوصول (Access) سريع الى ذلك القيد اذا كان الملف منظم بعناية . ان القيود وعلى الاغلب تكون موزعه بشكل لا تسلسلي على الملف وفي هذه الحالة ولاجل الوصول الى القيود و بشكل متسلسل حسب المفاتيح نكون بحاجة الى تصنيف الملف او استخدام الملف الواجد (Finder File) والذي باستخدامه نستطيع ايجاد اي قيد .

2 – العنونة : – Addressing

مع التنظيم المباشر . هنا المستعمل (User) يعين مدى للعناوين بادئا بالصفر وحتى اعلى قيمة موجبة . مثال ذلك اذا كان لدى المستعمل 4000 قيد فيستطيع المستعمل تعيين العناوين من الصفر وحتى 5000 او 6000 او اي قيمة اخرى وبشرط ان هذا التعيين لا يؤدي الى مضيعة بالخزن .

وهناك عدة طرق لاستخدام او لتعيين العناوين وهي كما يلي : –

2 – 1 : الملف المعنون مباشرة Directly Addressed File باستخدام العنونة المباشرة كل مفتاح يتحول الى عنوان قيد واحد فقط (وحيد) فيجعل عملية ايجاد اي قيد من قيود الملف سهلة وممكنة بعملية تحريك (Seek) واحدة وعملية قراءة (Reed) واحدة ايضا ومن الطرق الشائعة في هذا المجال هي : –

2 – 1 – 1 : استخدام العنوان كمفتاح Using The Key as Address في هذه الحالة يستخدم المفتاح كعنوان الى قيد ولكن باستخدام هذا التنظيم هناك بعض الضوابط والتي يجب الاخذ بها وهي :

اولا : المفتاح يجب ان يكون عدديا . ولتنظيم الملف بهذه الطريقة فأن عملية واحدة فقط مطلوبة وهي قسمة المفتاح العددي على عدد القيود التي يمكن ان يحويها المسار (Track) الواحد فأن خارج القسمة يعطينا عنوان المسار الذي سوف يحل فيه القيد والباقي مضاف اليه (1) يعطي عنوان موقع القيد في المسار المعين .

ثانيا : يتبين مما سبق ان القيد يجب ان يكون ثابت الطول .

والان لنوضح هذه الطريقة بالمثال التالي : –

لنفرض ان جزء من مفاتيح قيود هي كما موضحة في الشكل (1) وكذلك يبين الشكل اسطوانة (Cylinder) بعد ان حلت المفاتيح المأخوذة بالمثال في مواقعها . وعلى فرض ان المسار يحتوي على 10 مواقع فقط .

المفاتيح	1	2	3	4	5	6	7	8	9	10	الموقع ← / المسار
4	0				4						0
11		11									1
22			22								2
34					34						3
43				43							4
56							56				5
68									68		6
75						75					7
87								87			8
99										99	9

شكل (1)

من أجل توضيح المثال نأخذ المفتاح 22 عند قسمة هذا المفتاح على 10 يكون الناتج 2 والباقي 2 .

∴ سيحل هذا القيد في المسار الثاني والذي عنوانه (1) وفي الموقع الثالث . لنأخذ مفتاح اخر وليكن 68 عند قسمته على 10 يكون الناتج (6) .

وهذا يعني ان القيد سيحل في المسار السادس الذي عنوانه (5) والباقي 8 وهذا يعني انه سيحل في الموقع التاسع . وهكذا بالنسبة للبقية .

ان هذه الطريقة في تنظيم الملفات ليس فقط تسمح لنا بوصول سريع الى موقع القيد في وسط الخزن بل انها تعتبر طريقة مثالية لمعالجة الملف بشكل متسلسل (Sequential) هذا بالرغم من ان الملف منظم بصورة عشوائية وذلك من خلال كتابة القيود حسب تسلسل مفاتيحها ولكن المضرة الوحيدة والتي قد تحصل هي احتمال وجود كمية كبيرة من الخزين غير مستعملة .

2 – 1 – 2 : – استخدام قائمة المرجع Using A Cross Refernce List بهذه الطريقة كل مفتاح يعطي له عنوان ضمن جدول خاص لذلك يسمى (قائمة المرجع) وهذا الجدول يقوم المبرمج باعداده لذا فأن استخدام هذه الطريقة تتطلب عمل اضافي من قبل المبرمج وهذا الجدول يجب ان يخزن ايضا حيث يشكل هذا الجدول بحد ذاته ملفا اخر ومن الممكن طبعه على الطابعة الخطية الخاصة بالحاسبة .

ولنوضح هذه الطريقة بالمثال التالي : –

لنفرض ان لدينا ملف مفاتيحه (اي مفاتيح قيوده) هي كما يلي :

(س$_1$ ، س$_2$ ، س$_4$ ، ص$_1$ ، ص$_4$ ، ص$_6$ ، ل$_1$ ، ل$_3$ ، ل$_5$ ، ل$_7$) لهذا الملف يعد المبرمج قائمة كالتي مبينة في شكل (2) حيث يعطي لكل مفتاح عنوان يختاره هو .

المفتاح	العنوان	المفتاح	العنوان	المفتاح	العنوان
س 1	0/2	ص 4	2/8	ل 5	2/7
س 2	1/3	ص 6	5/1	ل 7	1/8
س 4	2/5	ل 1	1/2		
ص 1	0/8	ل 3	3/4		

الشكل (2)

والان لنوضح معاني وتحليل العناوين . ان العنوان 0/2 يعني الموقع الثاني في المسار 0 . والعنوان 2/3 يعني الموقع الثالث في المسار الاول والعنوان 2/8 يعني الموقع الثامن في المسار الثاني . . . . . وهكذا .

ان السيطرة على هذه القائمة يجب ان تكون جيدة ومضبوطة وذلك لان اي خطأ فيها بسبب خطأ في البرامج المعدة لاسترجاع القيود .

وفي هذه الطريقة يكون الزمن اللازم لاسترجاع اي قيد اطول مما هو في الطريقة السابقة . حيث ان هذه الطريقة لها شبه كبير بطريقة تنظيم الفهرسة التسلسلية (Index Sequential Organization)

2 – 2^{1-} : العنونة الغير مباشرة للملف (Indirectly Addressed File)

من الطرق الاخرى لتنظيم الملف هذه الطريقة والتي يمكن استخدامها عندما يكون مدى المفاتيح كبير مع وجود نسبة كبيرة من المفاتيح غير مستعملة .

مثال ذلك اذا كان مدى مفاتيح ملف ما هو من الصفر الى 10000 وهذا الملف يحوي على 1000 قيد فقط .

هناك اسلوب يستخدم لتقليل مدى المفاتيح وتحويلها الى عناوين وهذا الاسلوب يسمى العشوائية (Randomization) ويتم استخدام هذا الاسلوب بتعيين دالة (معادلة) يتحول بموجبها المفتاح الى عنوان كما موضح في (1) .

العنوان = د (المفتاح) 1

ولنوضح ما نقصد بذلك نأخذ المثال البسيط التالي : –
ملف مفاتيحه (مفاتيح قيوده) هي 0 ، 2 ، 4 ، 6 ، 8 ، 10 ، 12 ، 14 ، 16 ، 18 ، 20) هذا يعني ان مدى المفاتيح من الصفر وحتى 20 ومن اجل تقليل هذا المدى لنعرف داله بالشكل التالي .

$$\text{العنوان} = \frac{\text{المفتاح}}{2}$$

وبذلك ستكون العناوين هي (0 ، 1 ، 2 ، 3 ، 4 ، 5 ، 6 ، 7 ، 8 ، 9 ، 10) و بذلك سيكون المدى من الصفر وحتى 10 بدلا من 20 .

عند اتباع هذا الاسلوب او اي اسلوب اخر في تنظيم الملفات يجب ان يأخذ بنظر الاعتبار ما يلي :

اولا : كل مفتاح قيد موجود بالملف يجب ان يتحول الى عنوان موجود فعليا بالمدى المخصص .

ثانيا : المفاتيح يجب ان تكون موزعة بشكل جيد ومرتبة على مواقع الخزن التابعة لها بطريقة تمنع او تقلل عمليات التماثل (Synonyms) وهي وجود عنوان واحد فقط لمفتاحين مختلفين او اكثر اي وجود موقع خزن واحد لقيدين او اكثر واذا ما حدث ذلك فان القيد الاول يحل في موقعه الصحيح اما القيد المماثل فيحمل الى موقع اخر يسمى موقع خزن القيود الفائضة (Overflow Area) ومن الطرق التي تستخدم لتقليل عمليات التماثل هي حجز مواقع خزن اكثر اي زيادة حجم الخزين المطلوب لخزن الملف عن ما هو مطلوب وتكون النسبة المئوية للمواقع المستخدمة فعلا الى عدد المواقع المحجوزة لخزن الملف (Packing Factor) بحدود 80٪ الى 85٪ نسبة جيدة ومقبولة .

ومن الطرق المتبعة لتقليل الفائض من القيود هي الاستعاضة عن عملية العنونة العشوائية الى قيد (Randomizing to Record Address) بعملية العنونة العشوائية الى مسار (Randomizing to Track Address) وهذا يعني ان مجموعة قيود يصبح عنوانها هو نفس المسار وان عملية العنونة الى مسار هي اكثر اقتصادية من عملية العنونة العشوائية الى قيد . والقيد الذي يعنون عشوائيا الى مسار (الذي يكون عنوانه مسار) لا يعتبر قيد فائض الا متى ما امتلأ ذلك المسار ولم يعد فيه متسع للخزن ولكن من مضار عملية العنونة العشوائية الى مسار هي زيادة عدد التماثلات ولكن يقل الفيض باستخدام عملية العنونة العشوائية الى مسار .

المثال التالي يوضح عملية العنونة العشوائية الى قيد وعملية العنونة العشوائية الى مسار والفرق بينهما .

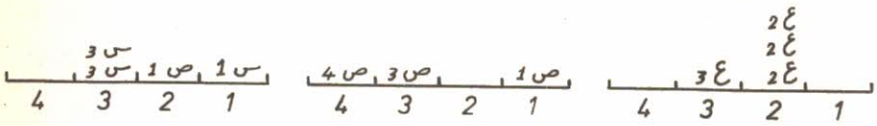
لو أخذنا مجموعة من القيود واصبحت عناوينها بعد اجراء العملية العشوائية هي كما يلي :

س$_1$ ، س$_3$ ، س$_3$ ، ص$_1$ ، ص$_2$ ، ص$_4$ ، ع$_2$ ، ع$_2$ ، ع$_2$ ، ع$_3$

يوضح لنا الشكل (3) عدد التماثلات وعدد القيود الفائضة . هذا اذا افترضنا ان

عملية العنونة العشوائية الى قيد وان كل مسار يتسع الى اربعة قيود فقط من الشكل نستطيع ان نتبين ان النسبة المئوية لعدد التماثلات هي 30% وان النسبة المئوية لعدد القيود الفائضة هي 30% ايضا .

المسار ع المسار ص المسار س

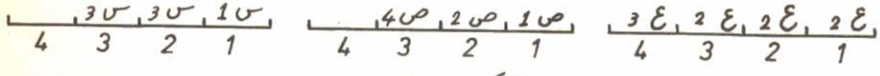


الشكل – 3 –

أما لو تأملنا الشكل (4) والذي يوضح عدد التماثلات وعدد القيود الفائضة والنسبة المئوية لكل منهما . هذا اذا اخذنا بعملية العنونة العشوائية الى مسار .

نرى من الشكل (4) ايضا ان النسبة المئوية لعدد التماثلات هي 70% بينما النسبة المئوية للقيود الفائضة هي صفر .

المسار ع المسار ص المسار س



الشكل – 4 –

ومن مضار زيادة عدد التماثلات زيادة الوقت اللازم للوصول الى مواقع القيود واسترجاعها .

3 – اسلوب العشوائية (Randomization Technique)

هناك عدة طرق لاجراء او اتباع هذا الاسلوب في تنظيم الملفات ولاختيار احسن هذه الطرق يقتضي المحاولة والخطأ ويجب ان نؤكد في اختيارنا لطريقة ما ان تكون نسبة التماثل اقل من 20% ومن الطرق هذه ما يلي : –

3 – 1 : طريقة القسمة والباقي (Division Remainder Method)

عند القيام بعملية تنظيم ملف ما فأننا نقترح استخدام هذه الطريقة اولا لانها اسهل طريقة وغالبا ما تعطينا نتائج جيدة وتتم هذه الطريقة كما يلي : –

يقسم المفتاح على عدد اولى وهذا العدد يختار على اساس اقرب عدد اولى الى عدد العناوين او عدد مواقع الخزن المحجوزة للملف حيث ان باقي القسمة يعطينا العنوان .

لنفرض ان لدينا 8000 قيد يراد خزنها وقد حجزنا لهذا الملف 10000 موقع خزن اي ان النسبة المئوية 80% فأن اقرب عدد اولى للـ 10000 هو 9973 .

يقسم المفتاح على 9973 ويأخذ باقي القسمة ويقسم على عدد القيود التي يمكن ان يحويها المسار وليكن 13 قيد ان ناتج القسمة الاخيرة يشير الى رقم المسار الذي سوف يخزن فيه القيد والباقي مضاف اليه (1) يعطي رقم الموقع الذي سوف يخزن فيه القيد .

هذا اذا كانت العنونة العشوائية الى قيد اما اذا كانت العنونة العشوائية الى مسار فأن العملية تصبح كما يلي وبنفس الافتراضات السابقة (8000 قيد ، 10000 موقع خزن والمسار يتسع الى 13 قيد) : —

يقسم 10000 على 13 فيكون الناتج 23 . 769 .

∴ نحن نحتاج 770 مسار لخزن هذا الملف .

ان اقرب عدد اولى للعدد 770 هو 769 .

يقسم المفتاح على 769 الباقي يمثل العنوان المباشر الى المسار الذي يخزن فيه القيد .

ان هذه الطريقة قابلة للاستخدام في حالة كون المفاتيح لا عددية وذلك باستخدام النظام الثنائي اي ان المفتاح يتحول الى اعداد بالنظام الثنائي حيث من المحتمل ان تكون النتائج افضل وذلك نظرا لوحدانية الرموز والحروف بالمفاتيح .

ان هذه الطريقة تحقق الهدف الاول بصورة تلقائية (اى ان كل مفتاح يتحول الى عنوان ضمن المدى المعرف) اما بخصوص عدد التماثلات فلا نستطيع تحديدها الا بعد التجربة والتطبيق الفعلي .

3 — 2 طريقة الطي (Folding)

في هذه الطريقة يجزء المفتاح الى جزئين او اكثر وتجرى عملية جمع هذه الاجزاء سوية حيث ان المجموع او جزء منه يستعمل كعنوان مباشر مثال ذلك : —

ليكن مفتاح قيد ما 746298 يمكن ان يجزء اي عددين هما (298) و (746) نجمع هذين الجزئين ليكون الناتج 1044 والذي يمثل العنوان .

ويمكن ان يجزء الى ثلاثة اجزاء هي (74) و (62) و (98) بجمع هذه الاجزاء فالناتج 234 يمثل العنوان المباشر .

3 – 3 : – طريقة تحويل الاساس (Radix Transformation)

في هذه الطريقة يعتبر المفتاح العددي عدد في نظام عددي اخر غير النظام العشرى وليكن الحادي عشر مثلا (بالرغم من انه من الناحية الصحيحة في النظام العشرى) ثم يحول الى النظام العشري .

مثال ذلك

المطلوب ايجاد عنوان المفتاح 42356 بالاساس العشري ، نستطيع اعتبار هذا المفتاح وكانه مكتوب في النظام الحادي عشر ثم نحوله الى النظام العشري اي ان

$$4\times^{4}11+2\times^{3}11+3\times^{2}11+5\times^{1}11+6\times^{6}11 = 61650$$

ان العدد 61650 يمثل عنوان مباشر الى القيد ولو اردنا ان يكون العنوان معين بـ (4) مراتب فقط فيمكن قطع العدد (6) ويكون العنوان 1650 هو العنوان المباشر للقيد .

4 – كلمة اخيرة :

عند تقييم الاسلوب العشوائي واي طريقة من طرقه افضل ليس فقط علينا حساب النسبة المئوية لعدد التماثلات بل يجب حساب العدد المتوقع لعمليات القراءة لكل قيد وهذا شيء مهم في معالجة الملف . اما كيفية حساب هذا العدد فأننا سنعطي المثال التالي :

لتكن لدينا 10قيود مفاتيحها 1 ، 2 ، 3 ، 4 ، 5 ، 6 ، 7 ، 1 ، 2 ، 3 ان القيود السبعة الاولى تحتاج الى عملية قراءة واحدة لكل منها . اما القيود الثلاثة الاخيرة فتحتاج الى عمليتي قراءة لكل منها .

∴ عدد القراءات الكلية قراءة $7+3\times2=13$

$$\text{معدل قراءة كل قيد} = \frac{\text{عدد القراءات}}{\text{عدد القيود}}$$

$$= \frac{12}{10} = 1.3 \text{ قراءة لكل قيد}$$

وان هذه النسبة كلما كانت قريبة من الواحد الصحيح فأن ذلك افضل .

المواد المترجمة