في التقنية الهشية

نبيـل خليـل عمر وسحر عبدالعزيز
جامعـــة الموصـــل

موجـــز :

يقارن هذا البحث طريقة السلسلة المباشره مع بعض الطرق الاخـرى في التقنية الهشية ، ويرصد المساوىء والمحاسن من خلال تنفيذ طريقـــة السلسلة المباشرة على حاسبة الكترونية صغيرة IBM 1130 بلغة فورتران .

مفـــردات :

Binary search method :	طريقة البحث الثنائي .
Collision :	تضـارب .
Direct chaining method :	طريقة السلسلة المباشـرة .
Entry address :	عنـوان ادخـال .
Hashing address :	عنـوان هشي .
Hashing function :	دالـة هشـية .
Hashing technique :	تقنيـــة هشـــية .
Linear search method :	طريقـــة البحث الخطـي .
Link address :	عنـوان ربـط .
Overflow area :	مساحة فيض .
Radix :	اساس
Record :	قيـــد .
Retrieving :	استرداد .

1 ـ مقدمـــة :

تأتي اهمية التقنية الهشية(*) في انها اسلوب يختصر زمن الخـــزن والاسترداد للبيانات المنشودة ، متجاوزا الاساليب التقليدية المعروفة ·

وتعرف التقنية الهشية بانها الاسلوب المستخدم للخزن والاســترداد في جدول ما ذى عدد من القيود ، وبزمن لا يعتمد على عدد القيود هـــذه بالمقارنة مع التقنيات الاخرى ، كطريقة التحث الخطى ، حيث يتزايد الزمن طرديا مع ازدياد عدد القيود ، أو طريقة البحث الثنائي حيث يتناســب الزمن مع اللوغاريتم الثنائي لعدد القيود · الا انه ينبغي القول هنا بان زمن البحث في التقنية الهشيه يعتمد من جهة اخرى على حجم الجدول ، فليقل بصورة ملحوظة كلما ازداد الحجم (ملحق رقم 1) ·

وينبغي ابتداء التفريق بين طرق التقنية الهشية ، وطرق حساب الدالة الهشية · فالاولى تعني اختيار الاسلوب الذي يتم بواسطته تنظيم الجدول وتحديد شكل عمليتي الخزن والاسترداد.بينما تعني الثانية ، اختيـــار الطريقة التي يتم بواسطتها حساب العنوان الهشي من عنـــوان الادخال · وللتقنية الهشية كما للدالة الهشية طرق متعددة ، لكل منها مســـاوىء ومحاسن وتحديدات ·

وقد ذكر مورو ولويس (1) أهم الطرق المستخدمة ، وسنوجز هنـــا أشهرها لنشرح فيما بعد شرحا مقارنا السلسلة المباشرة والطرق الاخرى ، بعد ان مثلنا طريقة السلسلة المباشرة بلغة فورتران وعلى حاسبة الكترونية ذات ثمانية آلاف كلمـــة ·

2ـ1 طرق حساب الدالـــة الهشـــية :

أ ـ طريقة القسمة : وتتم بتقسيم عنوان الادخال على حجم الجدول وايجاد المتبقي الذي يمثل العنوان الهشي حيث يتم بموجبه خزن القيد ·

فاذا كـان :

K : عنـــوان الادخـــال

N : حجـــم الجـــدول

h(K) : العنـوان الهشـــي

فـــان :

$$h(K) = K - \frac{K}{N} * N$$

(*) : عربها صاحب قاموس المورد كفعل: يمزج، يفرم، يشوش· ويأتي المعنى قريبا من كلمه (هش) حيث تنفرط جزيئات الشيء فلا تتماسك · ولعل في هذا توافق جميل !

ملحق رقم (1)

عدد البيانات = 3000

عدد القيود في الطف (ع) (حجم الطف)	زمن الخزن (دقيقة)	عدد حالات التضارب
3000	95,76	305,984
4500	38,40	57,415
6000	35,46	45,406
8000	31,50	29,990
9500	25,44	7,104

وعلى الرغم من ان هذه الطريقة تعتبر من اكثر الطرق شهرة في حساب العنوان الهشي ، وقد استعملناها هنا ، الا ان مشكلتها تكمن في الوقوع بحالة (التضارب) ، والتي نعني بها حصول عدة عناوين لادخال على نفس العنوان الهشي (3) .

ب ـ الطريقة العشوائية : يتم اختيار العنوان الهشي في هذه الطريقة عشوائيا من عنوان الادخال ، باستعمال برنامج ما أو صيغة رياضية معينة . هنا يستفاد من عنوان الادخال كأساس لتوليد العنوان العشوائي ، وعند ظهور حالة تضارب فان العنوان هذا يستعمل ثانية لتوليد عنوان هشي آخر ، وهكذا .

ج ـ طريقة الاساس : بأخذ عنوان الادخال واعتباره سلسلة من اعداد ذات أساس ثماني ، ثم تغيير الاساس الى أساس آخر كالاساس 11 مثلا واعادة تحويله الى الاساس العشري للحصول على عنوان هشي

د ـ وهناك طرق اخرى كطريقة تحليل الاعداد ، والترميز الجبري ، والضرب ... الخ ، تستخدم تبعا لنوع الحاسبة وقابليتها البرمجية .

2 ـ طـرق التقنيـة الهشـية :

تعتبر حالة (التضارب) العقبة الكأداء التي عندها تتوزع وتختـلف طرق التقنية الهشية بحثا عن زمن أقل في الخزن والاسترداد • فالطريقـة الفضلى هي تلك التي تقل فيها حالات التضارب ، أو تلك الطريقة التـي يتم فيها معالجة حالات التضارب بشكل اذكى • وما دامت التقنية الهشـية تعتمد التوزيع العشوائي للقيود ، فأن لحساب الدالة الهشية وللطريقـة نفسها أهميتها في التقليل من خطر (التضارب) المذكور •

ولكي تتشكل الخلفية المناسبة ، يجب وضع خوارزمية عامة للخـزن واخرى للاسترداد باستعمال التقنية الهشية ، وبعيدا عن طريقة حسـاب الدالة والاسـلوب •

2ـ1 خوارزمية الخـزن :

بدايـة

احسب الدالة الهشية للقيد من عنوان الادخال ،
اذا توفر الموقع اخزن والا احسب الدالة ثانية للحصـول على موقع خـال ، اخـزن ،

نهايـة

2ـ2 خوارزميـة الاسـترداد

بدايـة

احسب الدالة الهشية للقيـد المبحوث عنه في الجدول ،
اذا تطابق قيد العنوان الهشي مع القيد المخزون اذن اقصد س
والا ابحث عن القيـد المطلوب بتكرار حساب الدالـة ،
اذا لم يعثر على القيـد اقصـد ص ،
س : القيـد موجـود ،
ص : نهايـة

2ـ3 أهم طرق التقنيـة الهشـية هي ما يلي :

أ ـ طريقة البحث الخطي : بعد حساب الدالة الهشية لعنوان الادخـال يتم اختيار الموقع في الجدول ، فان كان خاليا تم خزن القيد الجديد فيه ، وان كان مشغولا بحثنا في الموقع المجاور ، وهكذا حتى يتـم العثور على موقع شاغر • وهناك اقتراح يقول : انه بدلا من زيادة

منتظمة فانه بالامكان ان تكون الزيادة عند حدوث التضارب غير منتظمة ولكنها تخضع لتسلسل مسبق .

ب ـ طريقة مساحة الفيض : عند حدوث التضارب ، فان القيد الجديد لا يخزن في نفس الجدول ، وانما في مساحة اخرى خارج الجدول تخزن فيها كل حالات التضارب ، وعند البحث عن قيد ما ، فان الاختبار يبدأ بالجدول الرئيس اولا ، ثم ينتقل الى مساحة الفيض (الاضافية) فيما بعد .

جـ ـ طريقة السلسلة المباشرة : في هذه الطريقة يربط لكل قيد عنوان جانبي يصل هذا القيد بالقيد الاخر الذي تضارب معه ، وتتكون من جراء ذلك سلسلة أو عدة سلاسل داخل الجدول الواحد ، يرتبط اعضاء السلسلة الواحدة فيما بينها بعناوين ربط ، ويحتوي عنوان الربط للقيد الموجود في نهاية السلسلة على صفر للدلالة على نهاية السلسلة .

2ـ4 ملاحظات عن طريقة السلسلة المباشرة :

تتعقد كتابة برنامج هذه الطريقة عن كل الطرق السابقة التي ذكرناها ولكن التعقيد هذا يقابله اختصار زمني ملحوظ ، وشطب على بحث لا نهائي ممقوت ، وبالتالي تسهل عملية الاسترداد لنجد القيد المطلوب (أو لا نجده) بعد عناء قليل . وتزداد سرعة الخزن كما تزداد سرعة الاسترداد كلما قل عدد القيود المخزونة أو زاد حجم الجدول ، ويعود ذلك الى قلة حالات التضارب التي هي مشكلة المشاكل في التقنية الهشية .

وفي الملحقين (2 و 3) مخططان انسيابيان لكيفية الخزن والاسترداد وقد استعملنا طريقة القسمة لحساب العنوان الهشي لانها الطريقة السرعى والفضلى نسبيا (3) . وتتم معالجة حالات التضارب هنا بالبحث الخطي من موقع التضارب عن موقع خال ، وعند الحصول عليه تخزن معلومات القيد كما يخزن في الوقت نفسه عنوان الموقع مجاور موقع التضارب لتسهيل عملية الاسترداد فيما بعد . وبذلك تتشكل سلسلة أو مجموعة من السلاسل داخل هذا الجدول ترتبط فيه قيود السلسلة الواحدة مع بعضها ـ كما نوهنا ـ بعناوين ربط . لذلك يصعب تماما وضع قاعدة رياضية لحساب زمن الخزن أو زمن الاسترداد للسببين التاليين :

أ ـ لصعوبة تحديد عدد السلاسل داخل الجدول الواحد .

ب ـ لصعوبة تحديد طول السلسلة الواحدة .

وكل ذلك يتحدد بالبيانات وبتسلسل البيانات نفسها ايضا ، اذ ان اختلاف التسلسل يقود الى اختلاف في تكوين السلاسل داخل الجدول . لذلك اضطررنا عندحساب زمن الخزن لمجموعة من البيانات لحجوم مختلفة

للجدول ان نستخدم نفس المجموعة وبنفس التسلسل من البيانات (ملحق رقم 1) .

ولتوضيح شكل السلسلة والسلاسل وعناوين الربط داخل الجدول، نرسم هذه الصورة :

عنوان الربط	القيد	العنوان الهشي	
	1	349500	0
	2	000099	3
	3	013499	11
	4	041203	10
	5	000304	6
	6	000004	7
	7	051204	9
	8	000007	12
	9	099906	0
	10	034803	0
نهاية سلسلة	11	012499	0
	12	64707	94
	13	004512	55
	14	003413	62
راس سلسلة	100	023799	2

البيانات حسب تسلسل الخـــزن :

.1	4512
.2	304
.3	7
.4	4
.5	349500
.6	23799
.7	99
.8	41203
.9	13499
.10	51204
.11	99906
.12	34803
.13	12499
.14	64707

وقد اقترح موريس (2) اقتراحا نظريا يلخص بما يلي :

- احسب العنوان الهشي للقيـــد .
- اذا كان الموقع محجوزا من قبل قيد له نفس العنوان ويشكل رأس السلسلة ابحث في موقع مجاور .
- اما اذا كان الموقع محجوزا من قبل قيد دخيل ، يخزن القيد الجديد في الموقع ، ويسحب القيـد الدخيل الى موقع آخر .

بمعنى اخر ، فان هذه الطريقة تسعى الى ان يحتل كل قيد موقعـــه الحقيقي مع أبعاد القيود الدخيلة الى مواقع اخرى تكون قريبة من الموقـع الراس . الا ان من مساوىء هذه الطريقة كما يذكر موريس نفسه هو تغيير المواقع الدائم للقيود الذي ينعكس على زمن الخزن . بينما لجأنا في هذا البحث الى ان يتسلم كل قيد موقعه حسب قدومه دون الحاجة الى تغيير المواقع . صحيح ان طريقة موريس لها مردودها الزمني في الاسترداد ، الا ان الخزن سيزيد كما ستزيد درجة تعقيد البرنامج نفسه لاحتياجه الى عنواني ربط لكل قيد ، الاول يشير الى القيد السابق والثاني للقيـــد اللاحـق ، وبالتالي فان فرق المعدل الزمني للخزن والاسترداد سوف لا يكون شاسعا بين ما اقترحه موريس وما نفذنـــاه .

3 ـ مقارنـــة :

الملاحظة التي يجب ان لا تغيب عن الذهن هي ان التقنية الهشية تزداد سوء كلما اقترب الجدول من الامتلاء ، والسبب يعود الى ازدياد التضارب وبالتالي ازدياد زمن البحث عن موقع خال . لذلك يفضل أن لا يزيد عدد القيود عن 80 بالمائة من حجم الجدول باية حال من الاحوال .

في طريقة البحث الخطي ، يتم الخزن ـ كما عرفنا ـ بحساب العنـوان الهشي ، واختبار الموقع ، وزيادة العنوان الهشي بواحد عند عدم خلو الموقع ، وهكذا الا ان الخطورة الحقيقية في هذا الاسلوب تظهر عند استرداد قيـد مفقود ، أي قيد لا وجود له في الجدول ، عند ذلك سيجري البحث عنه في كل الجدول لكي يتم بعدئذ الاعلان عن عدم وجوده ، فاذا افترضنا ان الجدول بحجم عدة آلاف من القيود سيتبين كم ستكون الخسارة الزمنية من جـراء البحث عن قيد مفقود . ان الامر نفسه يتكرر في اسلوب (مساحة الفيض) ، اذ سيجري التفتيش عن القيد المفقود في كل مساحة الفيض لكي يعرف القرار بعدئذ . كما ان تحديد حجم مساحة الفيض يعتمد أساسا على عدد حالات التضارب التي تعتمد بدورها على نوع البيانات وتسلسلها ، ومن ثم يصعب تحديد حجم هذه المساحة بشكل دقيق . أما اسلوب السلسلة المباشـــرة فانه يتخلص تماما من هذه المأساة . فالبحث عن القيد المطلوب يتم مـــن خلال مسح السلسلة التي يتوقع تواجد القيد فيها ، وعند انتهاء السلسلة دون العثور على القيد المطلوب فان ذلك يعني عدم وجوده . هذه نقطـة ، والنقطة الاخرى فان عملية استرداد القيد المطلوب لا تتم بشكل خطي كما يتم الخزن ، اذ ان دور عناوين الربط في السلسلة الواحدة هو في نقل التنفيذ عشوائيا حسب تركيبه السلسلة نفسها ، لذا فالوصول الى نهاية السلسلة سوف لا يكون عسيرا بأي شـكل .

ان المشكلة التي تواجهنا في طريقة السلسلة المباشرة تأتي من تعقـد البرنامج نفسه مما ينعكس في صعوبة صيانته ما لم يشفع بتوثيق جيـد وواضح ، اضافة الى ان عناوين الربط ستزيد من حجم الجدول . وما عدا ذلك تبقى كفاءة هذه الطريقة معتمدة على جهاز الخزن نفسه ، الذي يجب ان يكون من اجهزة البحث العشوائي كالقرص المغناطيسي حيث ان زمـن الوصول بالنسبة لرأس الجهاز الى الموقع المطلوب له تأثيره الملحوظ على التقنية الهشية اذ كلما قل زمن الوصول زاد ذلك في كفاءة التنفيذ .

4 ـ تجربـــة :

يجري اختبار هذه التقنية في احدى النظم التجارية المنفذة على حاسبة جامعة الموصل فالنظام التجاري المذكور يستخدم طريقة تقليدية في الخزن

والاسترداد لمجموعة كبيرة من القيود • ان الاشارات الاولية توحي بـــان الاختصار الزمني يشكل عقبة كبيرة باستخدام اسلوب السلسلة المباشـــرة عما هو عليه في الوقت الحاضر • وتجدر الاشارة الى ان هذا الاســـلوب لا يقف عند حد قبول قيد واسترداده لمرة واحدة وانما يمتد الى عدد غير محدود من الخزن والاسترداد ، لذلك فان الموقع الواحد في الجدول قـــد يستغل لاكثر من مرة •

وقد اضطررنا ايضا عند كتابة البرنامج لاستخدام برامج جاهزة ، اذ ان عنوين الادخال للنظام التجاري المذكور يزيد حجمها عن حجم اكبر رقم صحيح تقبله حاسبتنا وهو 32767 ، لذلك جرت معاملة عنوان الادخال على اساس المراتب مما اعطانا حرية استقبال أي عنوان مهما بلغ حجمه • وفي الملحق رقم (4) اسم ووظيفة البرنامج الجاهز الذي استخدمناه ، أما الملحق رقم (5) فيعطينا جدولا بازمان الخزن والاسترداد لمجموعة من البيانات •

5 ـ خلاصـــــة :

تبين من خلال هذا البحث ، كيف ان لاســـلوب السلسلة المباشـــرة نتائجه الايجابية في الاسراع بعمليتي الخزن والاسترداد ، وليس من شك في ان حاسبة جامعة الموصل تعاني من صغر حجم ذاكرتها ، ومن بطء زمن دورتها والبالغ 3.6 مايكروثانية ، ومع ذلك تبقى لهذه المحاولة مردوداتها الطيبة • اضافة الى ان هذه التقنية لهـــا القابلية على تقديم خدماتها في شتى المجالات ، ما دام الامر يعتمد على خزن واسترداد ، ومن ثم بامكان كل مراكز حاسبات القطر الاستفادة منه اذا اقتضى الامـــر •

6 ـ المصـــــادر :

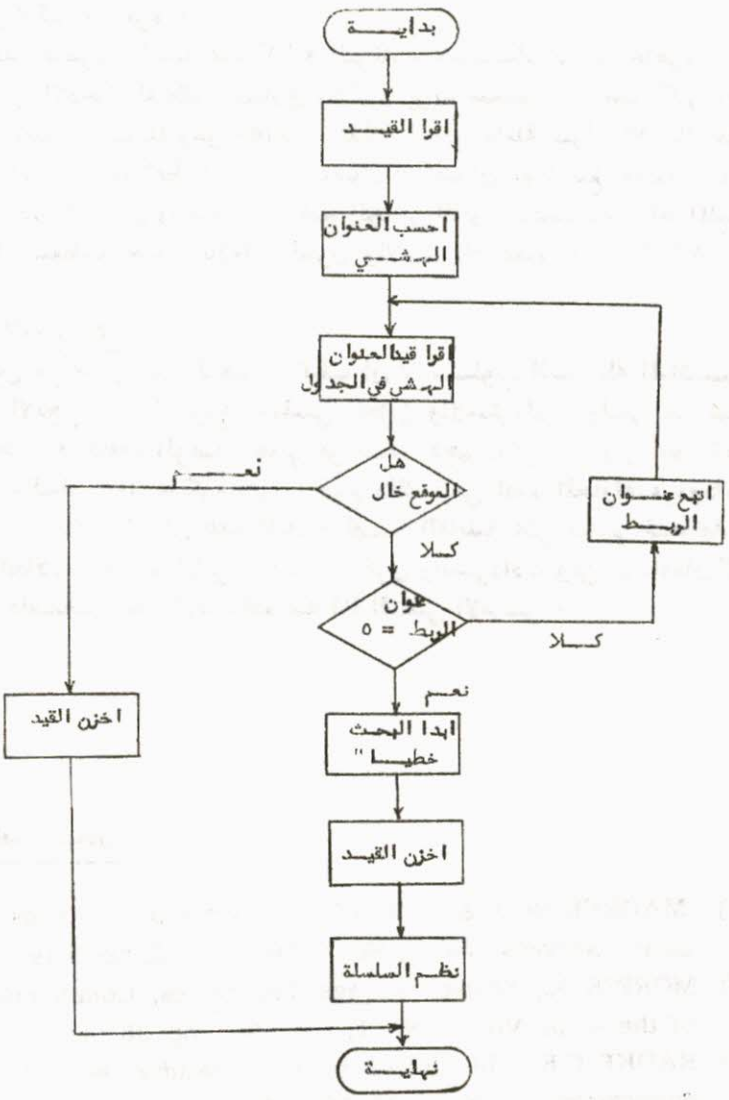
(1) MAURER W.D. & LEWIS T.G., Hash table methods, Computing Surveys, Vol. 7, No. 1, March 1975, pp: 5-19.

(2) MORRIS R., Scatter storage Techniques, Communications of the ACM, Vol. II, No. 1, Jan. 1968, pp: 38-44.

(3) RADKE C.E., The Use of quadratic residue research, Communications of the ACM, Vol. 13, No. 2, Feb. 1970, pp: 103-105.

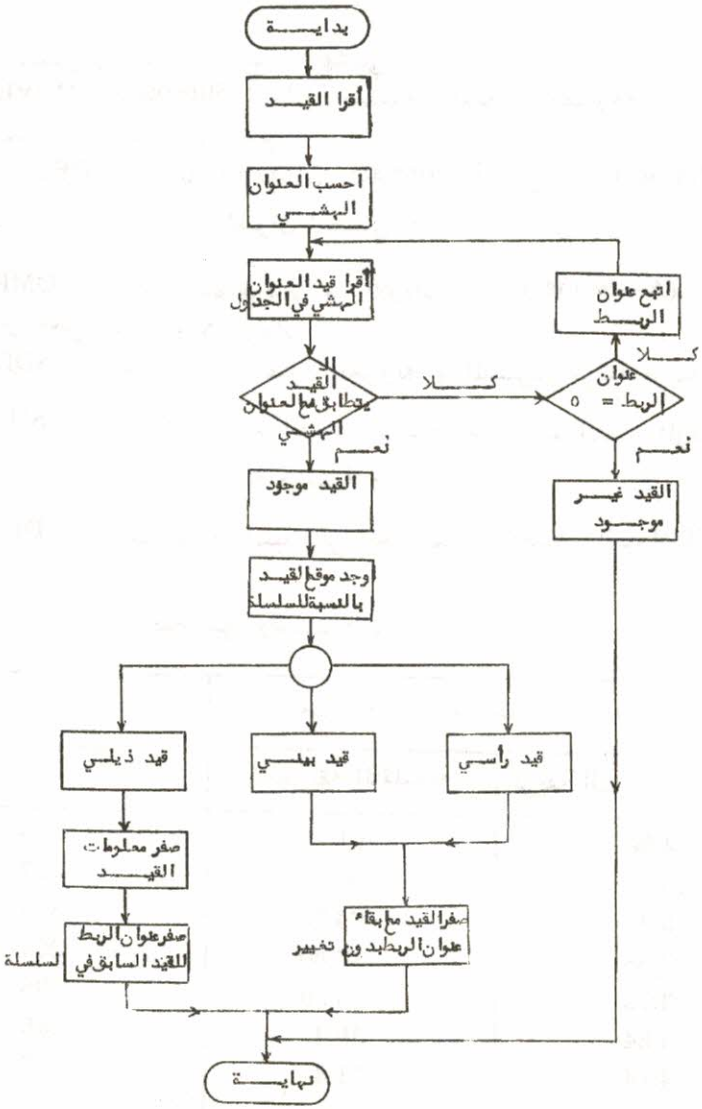
ملحق رقم (2)

مخطط انسيابي للخـــــــــزن



ملحق رقم (3)

مخطط انسيابي للاسترداد



ملحـــق رقـــم (4)

البرامج التجارية التي استعملت لتقليل زمن التشغيل :

1. Function GET : لتحويل الارقام من (A—format) الى (I—format)

تستعمل في حساب العنوان الهشـي ·

2. Subroutine MOVE : لنقل مصفوفة رقمية الى مصفوفة اخرى لتلافي تغير قيمها خلال البرنامج ·

3. DECAI : لتحويل الارقام من (I—format) الى ارقام (A—format) وتستعمل ايضا لحساب العنوان الهشـي ·

4. ICOMP : لمقارنة مصفوفتين من الارقام بـ (A—format) في برنامجي الخزن والاسترداد ·

5. ADD : لجمع مبالغ القيود المخزونة او المستردة مرتبة مرتبة ·

6. SUB : لطرح مبالغ القيود المستردة من مبالغ القيود الموقوفة في برنامج الاسترداد مرتبـة مرتبـة ·

7. DIV : لتقسيم رقم القيد على حجم الجدول لحساب العنوان الهشي ·

ملحـــق رقـــم (5)

عدد البيانـــات	زمن الخـــزن والاسترداد بالدقيقـــة	
	الطريقة التقليدية	طريقة السلسلة المباشرة
395	21	4.80
757	99	15.60
373	64.80	5.40
254	37.80	2.88
296	49.20	3.72
285	38.4	3.84
317	71.4	4.14

المراجـــع

تحتوي المراجع التالية على معلومات اكثر تفصيلا حول الموضوع ٠

1. BAHRANI, K, "Compiling Techniques and Table organization", AL-MUHANDIS — Vol. 17 — No. 1—Serial 47—1972.
2. BAHRANI, K, "Introduction to computor science — Part II" Printed by Baghdad Bureau — 1972 — 1973.