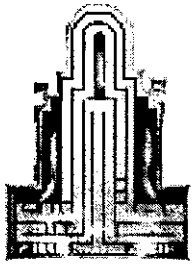
العنوان:	خوارزميات بحث سريعة للسلاسل الحرفية
المؤلف الرئيسي:	الحسنات، أحمد بشير عبدالله
مؤلفين آخرين:	عبابنة، إسماعيل محمد(مشرف)
التاريخ الميلادي:	2004
موقع:	المفرق
الصفحات:	1 - 67
رقم MD:	571559
نوع المحتوى:	رسائل جامعية
اللغة:	Arabic
الدرجة العلمية:	رسالة ماجستير
الجامعة:	جامعة آل البيت
الكلية:	كلية الأمير الحسين بن عبد الله لتكنولوجيا المعلومات
الدولة:	الاردن
قواعد المعلومات:	Dissertations
مواضيع:	الخوارزميات ، السلاسل الحرفية ، الحاسبات الالكترونية ، البرمجة
رابط:	http://search.mandumah.com/Record/571559

بسم الله الرحمن الرحيم



جامعة آل البيت

كلية الأمير الحسين بن عبدالله لتكنولوجيا المعلومات

قسم الحاسوب

عمادة الدراسات العليا والبحث العلمي

رسالة ماجستير بعنوان:

خوارزميات بحث سريعة للسلاسل الحرفية

Fast String Searching Algorithms

إعـــــداد

احمد بشير عبدالله الحسنات

٠٢٢٠٩٠١٠٠٩

المشرف

د. إسماعيل عبابنه

٢٠٠٤م

خوارزميات بحث سريعة للسلاسل الحرفية

Fast String Searching Algorithms

إعـــــداد

احمد بشير عبدالله الحسنات

٠٢٢٠٩٠١٠٠٩

المشرف

د. إسماعيل عبابنه

أعضاء لجنة المناقشة	التوقيـــــــع
د. إسماعيل عبابنه	.....
د. مأمون الربابعه	.....
د. "محمدعلي" العكور	.....
د. محمد علي العبادي	.....

قدمت هذه الرسالة استكمالا لمتطلبات الحصول على درجة الماجستير في علم الحاسوب في كلية سمو الأمير الحسين بن عبدالله لتكنولوجيا المعلومات في جامعة آل البيت.

نوقشت وأوصى بإجازتها بتاريـــــخ : ٢٠٠٤/٧/٢٩ م

الإهـــــــداء

اهدي هذا العمل المتواضع إلى والدي رمز العطاء،

وإلى والدتي رمز الحنان،

وإلى زوجتي رمز المحبة والوفاء،

وإلى كافة أفراد عائلتي لدعمهم وتشجيعهم،

وإلى عمي الشيخ عوض بن خليف لدعمه وتشجيعه.

الشـــكر والتقدير

الحمد لله حمداً كثيراً طيباً مباركـــاً فيـــه، واشـــكره شـــكراً يليـــق بجـــلال وجهه وعظيم قدرته، على ما يسر وسهل وهدى. أما بعد:

فلا يسعني إلا أن أتقدم بالشكر الجزيل إلى كـــل مـــن ســـاهم فـــي إنجـــاح هذا العمل وإخراجه إلـــى حيـــز الوجـــود، وأخـــص بالـــذكر الـــدكتور إســـماعيل عبابنه، والدكتور محمد العبـــادي لـــدعمهما المتواصـــل وتعبيـــد طريـــق البحـــث العلمي أمامي. واشكر الدكتور مأمون الربابعـــه، والـــدكتور محمـــدعلي العكـــور على توجيهاتهم.

كما واشكر السيد منذر الســـلامين، والســـيد حســـن الحســـنات للمســـاعدة في التدقيق اللغوي.

احمد الحسنات

المحتويات

الفصل الرابع: النتائج والمقارنة

فهرس الجداول

فهرس الأشكال

قائمة المصطلحات

Σ	المجموعة التي تحتوي على الأحرف الأبجدية
Advanced server	خادم متقدم
Algorithms	الخوارزميات
ASCII	مجموعة الشيفرة الأمريكية المعيارية لتبادل المعلومات
Bad-character shift	إزاحة الحرف السيئ
Begin	ابدأ
BF	الخوارزمية الساذجة
Binary Search	طريقة البحث الثنائي
Blocks	مجموعة من السلاسل الحرفية الفرعية
BM	خوارزمية بوير مور
C++	لغة برمجة
Charfreq	المصفوفة التي تحتوي على أعداد التكرارات لكل حرف
Complexity theory	نظرية التعقيد للاقترانات
Computes	يحتسب
Conclusion	الاستنتاجات
Constant	الثابت
CPU	وحدة المعالجة المركزية
Data Structure	تركيب بيانات
End	انهي
Exponential	الأسي
Factorial	المضروب
Filtering	تصفية
Geometric	الهندسي
Good-suffix shift	إزاحة اللاحقة الجيدة
HAB1	الخوارزمية المقترحة الأولى

HAB2	الخوارزمية المقترحة الثانية
HAB3	الخوارزمية المقترحة الثالثة
HAB4	الخوارزمية المقترحة الرابعة
Hashing function	اقتران النحت
Hashing Table	جدول اقتران النحت
Hp Compaq	ماركة حواسيب مسجلة عالميا
Index	فهرسة
Integer	عدد صحيح
Intel	اسم شركة صانعة للمعالجات
KR	خوارزمية كارب رابن
Landau's symbol	رمز لاندوز
Left-shift	ازاحة لليسار
Length	الطول
Linear	الخطي
linearithmic	الخطي اللوغاريتمي
List	قائمة
Load	الحمل (عدد النماذج)
Logarithmic	اللوغاريتمي
Lookup Table	جدول بحث
Loop	حلقة تنفيذية (تنفيذ نفس الأمر بشكل متكرر)
M	طول السلسلة الحرفية
Matrix	مصفوفة ثنائية
MAX( )	اقتران القيمة القصوى
MB. *mega byte*	وحدة قياس الحجم (مليون حرف)
Memcmpr()	اقتران المطابقة بين سلسلتين حرفيتين
N	طول النص
Newarray	حجز مصفوفة جديدة

Nodes count	عدد العناصر
non-over-lapped	سلاسل حرفية غير متداخلة الأحرف
Nox	عدد النماذج
Noy	عدد الأحرف في النص
O( )	درجة التعقيد للخوارزمية
Operator	عامل
Order	مرتبة الاقتران
OUTPUT	المخرجات
Overlap	متداخل
Pattern	النموذج (السلسلة الحرفية)
Pentium	اسم ماركة تجارية للمعالجات
Polylogarithmic	اللوغاريتمي المتعدد
Preprocessing	المعالجة الأولية
Proposed	المقترح
q-gram	سلسلة حرفية بطول ك تستخدم في فهرسة النصوص
QS	خوارزمية البحث السريع
Quadratic	تربيعي
RAM	ذاكرة القراءة العشوائية
REHASH	إعادة احتساب اقتران النحت
Repetition-Based - Indexing	الفهرسة المبنية على التكرار
Results	النتائج
Searching	عملية البحث
Speed	السرعة
String	سلسلة حرفية
String Matching	مطابقة السلاسل الحرفية
Structure	تركيب يحتوي على بيانات وإجراءات لمعالجتها
Sub-String	سلسلة حرفية فرعية

Suffix Array	مصفوفة اللاحقة
Suffix Indexing	فهرسة اللاحقة
Suffix Tries	شجيرات اللاحقة
T	الوقت المستهلك
Text indexing	فهرسة النصوص
Visual C++	لغة برمجة C++ المرئية
Windows 2000	نظام تشغيل
X	السلسلة الحرفية (النموذج)
Y	النص

ملخـــــص

بلغ حجم المعلومات وعدد الوثائق المحفوظة على شكل نصوص مكتوبة وبكافة لغات العالم حدا لا يمكن إحصاؤه، ولا يمكن التنبؤ بالحد الذي قد تصل إليه مستقبلا، وأصبحت عملية البحث عن معلومة معينة وسط هذا الكم الهائل تحتاج إلى وقت كبير. من هنا فان الضرورة تقتضي إيجاد وسائل بحث سريعة وغير تقليدية لتخدم هذا الغرض، حيث أن أسرع خوارزمية تقوم بهذا العمل هي خوارزمية بوير مور (Boyer Moore) والتي تعتمد على عدد قليل من المقارنات للأحرف، والقفزات الكبيرة التي تجريها على النص أثناء عملية البحث.

تقترح هذه الدراسة عدة خوارزميات للبحث في النصوص عن السلاسل الحرفية بشكل سريع وفعال، حيث تعتمد على معالجة أولية تجريها على النص وليس على النماذج المبحوثة كما تفعل الخوارزميات التقليدية. تعتمد المعالجة الأولية على عمل مصفوفة متغيرة الحجم تحتوي جميع مواقع حرف ما، وعمل مصفوفة ثابتة بطول (٢٥٦) عنصرا وهو طول الأبجدية، المفترضة في هذه الدراسة على أنها مجموعة الشيفرة الأمريكية المعيارية لتبادل المعلومات (ASCII)، وهذه المصفوفة مكونة من (٢٥٦) مصفوفة متغيرة الحجم (أي مصفوفة لكل حرف في الأبجدية (ASCII))، وكل مصفوفة متغيرة الحجم لحرف ما تحتوي جميع مواقع ذلك الحرف في النص. عندما يراد البحث عن سلسلة حرفية ما فان هذه الخوارزميات سوف تستخدم إحدى هذه المصفوفات في عملية البحث بدلا من عملية مسح النص كاملا، حيث أن طول المصفوفة المستخدمة في المعدل حتما سيكون اقل من حجم النص، وهذا بالتأكيد سيقلل من الوقت اللازم لإنهاء عملية البحث. بالإضافة إلى المعالجة الأولية، تم اعتماد بعض التحسينات الإضافية على هذه الخوارزميات، مثل مقارنة الحرف الأخير قبل عملية المقارنة الكلية للأحرف، واستخدام مصفوفة الحرف الأقل تكرارا الذي يقلل من الوقت بشكل كبير، وكذلك مقارنة الحرف الاول والأخير قبل إجراء عملية المقارنة مما يزيد من الانتقائية للخوارزمية، وينعكس ذلك على أداءها وفعاليتها حيث يقلل من الوقت المستهلك في عملية البحث.

تمت في هذه الدراسة برمجة جميع الخوارزميات المقترحة وكذلك برمجة عدة خوارزميات تقليدية ومقارنة نتائجها. وتبين النتائج بان أداء هذه الخوارزميات المقترحة أفضل بكثير من أداء الخوارزميات التقليدية وبنسب تحسين كبيرة، وبأداء أفضل عند زيادة الأحمال.

الفصل الأول

مقدمة الدراسة
Introduction

١-١ تمهيد.

إن مشكلة البحث في النصوص المكتوبة عن السلاسل الحرفية مشكلة كلاسيكية قديمة، وتستخدم عملية البحث هذه في كثير من مجالات علم الحاسوب، وقد برزت هذه المشكلة بشكل واضح في السبعينات من القرن الماضي، عندما بدأت شركات البرمجة بإنتاج برمجيات معالجة النصوص الكتابية ولغات البرمجة، وازدادت الحاجة لإيجاد حلول ناجعة وسريعة لهذه المعضلة. من الواضح أيضا أن البحث في النصوص عن السلاسل الحرفية من أساسيات عملية استرجاع المعلومة في كثير من الحقول مثل قواعد البيانات، خاصة عندما تكون قاعدة البيانات غير منتظمة، أو إذا كانت تكلفة فهرسة قاعدة البيانات عالية سواء في استهلاك الوقت أو المساحة من القرص الصلب أو استهلاك الذاكرة (Yamashita, et al, 1996; Hoffman and McCullough, 1971).

ويعرّف البحث في النصوص عن السلاسل الحرفية بأنه إيجاد حرف أو كلمة أو مجموعة كلمات (أي سلسة حرفية أو نموذج) في نص غير منتظم(أي لا شكل قياسي له). وقد زاد الاهتمام في هذا الموضوع حديثا بسبب كثرة استخدام البحث عن المواقع ومحتوياتها في شبكة الانترنت، بالإضافة إلى كثرة استخدام معالجات النصوص وقواعد البيانات (Liddell,1997).

١-٢ تعاريف أساسية.

نفترض هنا أن الأبجدية التي تُكون عناصرها (أو حروفها) كلا من النص والسلسة الحرفية هي مجموعة الشيفرة الأمريكية المعيارية لتبادل المعلومات (ASCII) أو أية مجموعة فرعية منها. سوف يتم تمثيل أية سلسلة حرفية (أو نموذج) X بطول m بالمصفوفة

$$(X[0],.....,X[m-1])= X[0...m-1]$$

ويحتوي الموقع الذي يلي الموقع الأخير (X[m]) حرفاً خاصاً لا يظهر إلا في آخر كل سلسلة حرفية ليكون إشارة لمعرفة انتهاء السلسلة الحرفية، كما سنفترض أن كلا من السلسلة الحرفية المراد البحث عنها والنص المبحوث فيه موجودان في الذاكرة الرئيسة للحاسوب (RAM)، وأن عملية البحث تتضمن إيجاد مطابقة واحدة أو أكثر للسلسلة الحرفية في النص

(Makinen, 1999)، وسوف نرمز للنص بالرمز Y = Y [0..n-1] أي أن طوله يساوي n، وسنفترض أن كل من النص والسلسلة الحرفية مبنيان على مجموعة منتهية من الحروف الأبجدية، ونرمز لها بالرمز Σ ولطولها بالرمزσ، وبناءً على ذلك يمكن اعتماد التعاريف التالية:

- السلسلة الحرفية: مجموعة مرتبة مكونة من حرف واحد أو أكثر مثل (abc)، وسنعتبرها النموذج المراد البحث عنه في هذا البحث (Makinen, 2003) .
- النص: مجموعة مرتبة مكونة من سلسلة حرفية أو أكثر مثل: (abcabadfdsdsd).
- الكلمة U هي بادئة (prefix) للكلمة W إذا وفقط إذا كانت هناك كلمة V حيث أن W=UV.
- الكلمة V هي لاحقة (suffix) للكلمة W إذا وفقط إذا كانت هناك كلمة U حيث أن W=UV.
- Σ : مجموعة الأحرف الأبجدية وهي هنا (ASCII).
- X : السلسلة الحرفية أو النموذج وهو بطول m.
- Y : النص وهو بطول n.
- النافذة : وهي نص فرعي (أي جزء من النص) وتساوي $Yi...Yi+m$ حيث $0 \le i < n$
- $Xi \in \Sigma$ لكل i .
- $Yi \in \Sigma$ لكل i.

إن جميع الخوارزميات في هذه الدراسة هدفها البحث عن X في Y لبيان فيما إذا كان هناك تطابق (match) أو اكثر، والتطابق يحصل إذا وفقط إذا كان هناك نص فرعي (نافذة) $Y:Yi...Yi+m$ حيث $0 \le i < n-m$ و $X1....Xm$ تساوي $Yi...Yi+m$ لكل i (Charras and Lecroq, 2004).

١-٣ هدف الدراسة والمشكلة التي تعالجها.

تهدف هذه الدراسة إلى وضع عدة خوارزميات فعالة وسريعة للبحث في النصوص عن سلاسل حرفية معينة، وذلك لما لهذا الموضوع من أهمية كبيرة تزداد مع مرور الزمن، نظراً لازدياد كمية المعلومات ونوعيتها، واختلاف لغاتها وتدفقها بشكل هائل، وخاصة من خلال شبكة الانترنت. سيتم في هذا البحث دراسة الخوارزميات المقترحة والخوارزميات الموضوعة سابقا لحل هذه المعضلة، ومقارنتها من خلال برمجة جميع الخوارزميات وتنفيذها في نفس البيئة

وتحت نفس الظروف ودراسة أفضل وأسوأ الحالات، وكذلك دراستها في الوضع الاعتيادي عندما يكون النص حقيقياً، وعندما تكون النماذج متغيرة الطول والعدد، مع الأخذ بعين الاعتبار كلاً من كلفتي الوقت المستهلك والمساحة المحجوزة من الذاكرة لجميع الخوارزميات.

١-٤ ترتيب محتوى الدراسة.

تم في الفصل الاول تعريف مشكلة البحث في النصوص المكتوبة عن السلاسل الحرفية وذكر بعض التعاريف الأساسية المستخدمة في هذه الدراسة، وحدد هدف الدراسة.

ويحتوي الفصل الثاني مقتضبا درجة التعقيد للخوارزمية، وكذلك عرضاً لأربعة من اشهر خوارزميات البحث في النصوص عن السلاسل الحرفية هي الخوارزمية الساذجة (Brute Force)، وخوارزمية كارب رابن (Karp Rabin)، وخوارزمية بوير مور (Boyer Moore)، وخوارزمية البحث السريع (Quick Search)، وذلك من خلال التعريف بكل خوارزمية، وبرمجتها بلغة C++، وذكر أهم خصائصها، وعرض مثال عملي يبين كيف تعمل.

أما الفصل الثالث فيتحدث عن الخوارزميات المقترحة من خلال توضيح فكرة المعالجة الأولية التي تعتمد عليها هذه الخوارزميات، وبرمجتها بلغة C++ وعرض مثال عملي. وكذلك توضيح جميع الخوارزميات المقترحة من خلال تعريف الخوارزمية، وذكر خصائصها الرئيسة، وبرمجتها بلغة C++، وعرض مثال عملي عليها. كما ويتحدث هذا الفصل عن آليات فهرسة النصوص ومقارنتها مع الخوارزميات المقترحة.

أما موضوع الفصل الرابع فهو النتائج والمقارنة، حيث يتناول الفصل الافتراضات التي تقوم عليها الدراسة، ومقارنة نتائج الفحوصات التي تمت على العينة، وهي الفحص والمقارنة باستخدام النصوص الاعتيادية الطبيعية، والفحص والمقارنة عندما يكون النموذج متغير الطول، والفحص والمقارنة بزيادة الحمل. بالإضافة إلى ذكر الحالات الأسوأ للخوارزميات المقترحة وكذلك عرض الاستنتاجات والعمل المستقبلي.

الفصل الثاني

خوارزميات بحث السلاسل الحرفية

String Searching Algorithms

٢-١ تمهيد.

سوف نتحدث الآن باختصار عن درجة التعقيد النظرية للخوارزميات، وكذلك عرض لبعض من اشهر خوارزميات البحث في النصوص، ونبدأ بالخوارزمية الساذجة (Brute Force)، وسنرمز لها بالرمز (BF) في هذه الدراسة، ثم خوارزمية كارب رابن، وسنرمز لها بالرمز (KR)، وتليها خوارزمية بوير مور، وسنرمز لها بالرمز (BM)، وأخيرا خوارزمية البحث السريع، وسنرمز لها بالرمز (QS).

إن جميع هذه الخوارزميات تعمل كالتالي: تقوم الخوارزمية بمسح النص بمساعدة نافذة بطول النموذج المراد البحث عنه ويساوي m، وتكون بداية البحث بمحاذاة يسار النافذة مع يسار النص، ثم مقارنة أحرف النافذة مع أحرف النموذج، وهذه العملية تدعى محاولة (attempt) ثم بعد أن تتم مطابقة كاملة للنموذج، تقوم هذه الخوارزميات بإزاحة النافذة إلى اليمين للبحث عن مطابقة أخرى، وفي حالة عدم المطابقة أيضا تتم إزاحة النافذة إلى اليمين للبحث عن مطابقة قد تكون موجودة في مكان آخر على طول النص، وتتم إعادة جميع هذه الخطوات حتى تصل إلى نهاية النص، وذلك لان هدف هذه الخوارزميات هو إيجاد جميع التطابقات في النص إن وجدت، أو إعلان عدم وجود للنموذج في النص (Charras and Lecroq, 2004).

في حالة مقارنة النموذج مع النافذة للتأكد من التطابق أو عدمه (أثناء عملية البحث) فان جميع الخوارزميات السابقة تستند إلى الحقيقة التي تقول: "إن تطابق أي حرف للنموذج في أي موقع مع أي حرف للنافذة في نفس الموقع أمر غير محبب، لذلك على الخوارزمية أن تكون مبنية على أساس الرفض السريع وإعلان أن التطابق غير موجود بأسرع وقت ممكن، أي دون اللجوء إلى مقارنة جميع الأحرف" (Baker, 1991). وبعبارة أخرى عند مقارنة أحرف النموذج مع أحرف النافذة أثناء عملية البحث يجب أن تتم هذه العملية بسرعة، أي يجب أن يكون البحث هنا عن الأحرف غير المتشابهة وذلك لان حرفا واحد غير مشابه يعني عدم وجود تطابق، لأننا معنيون بالتطابق الكامل في هذه الدراسة (Exact Matching)، وليس التطابق التقريبي (Approximate Matching).

٢-٢ درجة التعقيد النظرية للخوارزمية.

يستخدم عادة اقتران يسمى باقتران الواو الكبيرة ( )O للتعبير عن درجة التعقيد للخوارزمية أثناء التنفيذ، وتسمى أيضا برمز لاندو (Landau's symbol) وهذا الرمز يستخدم في نظرية التعقيد (complexity theory) وعلم الحاسوب والرياضيات لوصف سلوك الاقترانات بشكل تقريبي، وبشكل أساسي يبين سرعة نمو أو انحدار الاقتران. وقد جاء الاسم لاندو من اسم العالم الألماني ادموند لاندو (Edmund Landau) الذي أوجد هذا الرمز، يستخدم الحرف (O) للتعبير عن نمو أو انحدار الاقتران لان نسبة نمو الاقتران تسمى مرتبة الاقتران (order)، وهذا النمو أو الانحدار يستخدم للتعبير عن المصادر التي يستخدمها الاقتران كالوقت أو المساحة المحجوزة من ذاكرة الحاسوب (Naps and Pothering,1992) .

على سبيل المثال عند تحليل خوارزمية معينة يمكن أن يكون الوقت (أو عدد الخطوات) اللازم لحل معضلة بحجم n يعطى على شكل: $T(n) = 4\,n^2 - 2\,n + 2$. وإذا تم تجاهل الثوابت (وهذا منطقي لأنها تعتمد على الحاسوب الذي يتم استخدامه لتنفيذ البرنامج وتختلف من حاسوب لآخر) وتم إبطاء فترة النمو، عندها يمكن القول أن $T(n)$ تنمو بمرتبة n^2 ويكتب:

$T(n) = O(n^2)$. ٥٩٤٣٤٠

الجدول ٢-١ التالي يمثل أهم الاقترانات التي تواجهنا أثناء تحليل عمل الخوارزميات (Naps and Pothering,1992).

الاسم	الاقتران
الثابت (constant)	$O(1)$
اللوغاريتمي (logarithmic)	$O(\log(n))$
الخطي اللوغاريتمي (linearithmic)	$O(n \log(n))$
اللوغاريتمي المتعدد (polylogarithmic)	$O((\log(n))^c)$
الخطي (linear)	$O(n)$
التربيعي (quadratic)	$O(n^2)$
الهندسي (geometric)	$O(n^c)$
الآسي (exponential)	$O(c^n)$
المضروب (factorial)	$O(n!)$

الجدول ٢-١: أهم الاقترانات التي يمكن أن تسلكها معظم الخوارزميات.

٢-٣ الخوارزمية الساذجة (BF).

تسمى هذه الطريقة بالخوارزمية الساذجة، وذلك لأنها بسيطة وسهلة التطبيق، لكنها تأخذ وقتا طويلا أثناء التنفيذ.

٢-٣-١ الخصائص الرئيسة.

تتميز هذه الخوارزمية بالخصائص التالية (Charras and Lecroq, 2004):

- لا توجد مرحلة معالجة أولية.
- بحاجة إلى مساحة ثابتة إضافية.
- الإزاحة إلى اليمين تكون فقط بمقدار واحد.
- يمكن أن تتم عملية المقارنة في أي ترتيب.
- درجة التعقيد لوقت عملية البحث O(mn) .
- عدد المقارنات المتوقع للأحرف 2n في المعدل.

٢-٣-٢ الوصف.

تتم مقارنة جميع الأحرف حرفا تلو الآخر حتى تجد الخوارزمية عدم تطابق لأي حرف أو تطابق لجميع الحروف، ثم تعمل إزاحة بمقدار حرف واحد لليمين، ومن جديد تقوم الخوارزمية بإعادة المقارنات جميعها مرة أخرى حتى تصل إلى نهاية النص (Charras and Lecroq, 2004).

تقوم هذه الخوارزمية بإجراء الإزاحة لليمين بمقدار حرفا واحدا فقط في كل محاولة بحث وفي جميع الحالات، في حالة التطابق فإنها تنفذ الإزاحة للبحث عن تطابقات أخرى، وفي حالة عدم التطابق فإنها تنفذ الإزاحة للبحث عن أية مطابقة قد تكون موجودة على طول النص. أما مقدار الإزاحة فهو ثابت على الدوام ويساوي واحد، وذلك لان هذه الخوارزمية لا تملك آليات تساعدها في تحديد طول القفزة (الإزاحة)، وتنفذ إزاحة واحدة عند المطابقة للتأكد من أن النماذج غير متداخلة.

٢-٣-٣ برمجة الخوارزمية (BF) بلغة C++.

وتتم برمجتها بلغة C++ كالتالي (Charras and Lecroq, 2004) :

```
void BF(char *x, int m, char *y, int n) {
  int i, j;

  /* Searching */
  for (j = 0; j <= n - m; ++j) {
    for (i = 0; i < m && x[i] == y[i + j]; ++i);
    if (i >= m)
      OUTPUT(j);
  }
}
```

٢-٣-٤ مثال على الخوارزمية (BF).

على فرض أن الصف الأول من الأحرف يمثل النص Yوالثاني يمثل النموذج X.

المحاولة الأولى

ج	س	أ	ت	س	ج	س	أ	ج	أ	ج	أ	ج	ت	أ	ت	أ	س	أ	ج	ت	أ	س	ج
1	2	3	4																				
ج	س	أ	ج	أ	ج	أ	ج																

المحاولة الثانية

ج	س	أ	ت	س	ج	س	أ	ج	أ	ج	أ	ج	ت	أ	ت	أ	س	أ	ج	ت	أ	س	ج
	1																						
	ج	س	أ	ج	أ	ج	أ	ج															

المحاولة الثالثة

ج	س	أ	ت	س	ج	س	أ	ج	أ	ج	أ	ج	ت	أ	ت	أ	س	أ	ج	ت	أ	س	ج
		1																					

	ج	أ	ج	أ	ج	أ	س	ج

المحاولة الرابعة

ج	س	أ	ت	ج	أ	س	أ	ت	أ	ج	ت	أ	ج	أ	ج	أ	ج	س	ج	أ	ت	س	ج	س	أ	ج
																					1					
																					ج	س	ج	أ	ج	أ	ج

المحاولة الخامسة

ج	س	أ	ت	ج	أ	س	أ	ت	أ	ج	ت	أ	ج	أ	ج	أ	ج	س	ج	أ	ت	س	ج	س	أ	ج
																					1					
																				ج	أ	ج	أ	ج	أ	س	ج

المحاولة السادسة

ج	س	أ	ت	ج	أ	س	أ	ت	أ	ج	ت	أ	ج	أ	ج	أ	ج	س	ج	أ	ت	س	ج	س	أ	ج
													8	7	6	5	4	3	2	1						
													ج	أ	ج	أ	ج	أ	س	ج						

المحاولة السابعة

ج	س	أ	ت	ج	أ	س	أ	ت	أ	ج	ت	أ	ج	أ	ج	أ	ج	س	ج	أ	ت	س	ج	س	أ	ج
																		1								
													ج	أ	ج	أ	ج	أ	س	ج						

المحاولة الثامنة

ج	س	أ	ت	ج	أ	س	أ	ت	أ	ج	ت	أ	ج	أ	ج	أ	ج	س	أ	ج	ت	س	ج	س	أ	ج
																			1							
												ج	أ	ج	أ	ج	أ	س	ج							

المحاولة التاسعة

ج	س	أ	ت	ج	أ	س	أ	ت	أ	ج	ت	أ	ج	أ	ج	أ	ج	س	أ	ج	ت	س	ج	س	أ	ج

						2	1
ج	أ	ج	أ	ج	أ	**س**	**ج**

المحاولة العاشرة

ج	س	أ	ت	ج	أ	س	أ	ت	أ	ت	ج	أ	ج	أ	ج	أ	س	ج	س	ت	أ	س	ج
														1									
							ج	أ	ج	أ	ج	أ	س	**ج**									

المحاولة الحادية عشرة

ج	س	أ	ت	ج	أ	س	أ	ت	أ	ت	ج	أ	**ج**	أ	ج	أ	س	ج	س	ت	أ	س	ج
												2	1										
						ج	أ	ج	أ	ج	أ	**س**	**ج**										

المحاولة الثانية عشرة

ج	س	أ	ت	ج	أ	س	أ	ت	أ	ت	ج	أ	ج	أ	ج	أ	س	ج	س	ت	أ	س	ج
												1											
					ج	أ	ج	أ	ج	أ	س	**ج**											

المحاولة الثالثة عشرة

ج	س	أ	ت	ج	أ	س	أ	ت	أ	**ت**	**ج**	أ	ج	أ	ج	أ	س	ج	س	ت	أ	س	ج
										2	1												
				ج	أ	ج	أ	ج	أ	**س**	**ج**												

المحاولة الرابعة عشرة

ج	س	أ	ت	ج	أ	س	أ	ت	أ	**ت**	ج	أ	ج	أ	ج	أ	س	ج	س	ت	أ	س	ج
										1													
			ج	أ	ج	أ	ج	أ	س	**ج**													

المحاولة الخامسة عشرة

ج	س	أ	ت	ج	أ	س	أ	ت	أ	ت	ج	أ	ج	أ	ج	أ	س	أ	ت	س	ج	س	أ	ت	ج	س
									1																	
	ج	أ	ج	أ	ج	س	ج																			

المحاولة السادسة عشرة

ج	س	أ	ت	ج	أ	س	أ	ت	أ	ت	ج	أ	ج	أ	ج	أ	س	أ	ت	س	ج	س	أ	ت	ج	س
									1																	
	ج	أ	ج	أ	ج	س	ج																			

المحاولة السابعة عشرة

ج	س	أ	ت	ج	أ	س	أ	ت	أ	ت	ج	أ	ج	أ	ج	أ	س	أ	ت	س	ج	س	أ	ت	ج	س
									1																	
ج	أ	ج	أ	ج	أ	س	ج																			

يلاحظ في المثال السابق أن الخوارزمية الساذجة نفذت المطلوب بـــ (٣٠) ثلاثين عملية مقارنة للأحرف.

٢-٤ خوارزمية كارب رابن (KR).

"إن تصنيف النماذج يعتمد على طريقتين هما: ١) اختيار مقياس رقمي لتمثيل النماذج، و ٢) خوارزمية لتعريف وتحديد النماذج" (Greenberg and Konheim,1964). وعلى المبدأ نفسه تعتمد هذه الطريقة حيث تقوم بعمل بصمة (مقياس رقمي) للنموذج ، وعمل بصـــمة بالطريقـــة نفسها للنافذة، ومقارنة البصمتين بدلا من مقارنة جميع الحروف (المكلف نسبيا) وفي حال تطابق البصمتين (الرقمين) تكون هناك احتمالية أكبر لتطابق النموذج مع النافذة، وعندها تتم المقارنـــة حرفا بحرف(Karp and Rabin, 1987) .

إن بصمات النماذج لا تمثل النماذج بشكل فريد يميز كل نموذج عن الآخر بواســـطتها، ولكنها تقدم الخصائص التالية:

- أي نموذجين متطابقين فان لديهما نفس البصمة.