مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (٧) العدد(٥) ٢٠٢٣
Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (7) No. (5) 2023

تحسين خوارزمية ترتيب الصفحات الموزونة في محركات البحث

د. راغب طعمه*

جبران قشعور**

(تاريخ الإيداع ٢٠٢٣/٣/٢٣ . قُبِل للنشر في ٢٠٢٣/٧/١١)

□ ملخّص□

تعيد محركات البحث عدد كبير من الصفحات المحتملة التي تحتوي على المعلومات ذات الصلة بناءً على استعلامات المستخدم، ومن أجل ذلك يوجد العديد من خوارزميات التصنيف التي ترتب النتائج المعادة حسب أهميتها. إن أغلب خوارزميات التصنيف تركز على التنقيب في بنية الويب web structure mining وبعضها الأخر يضيف مفهومي التنقيب في استخدام الويب web usage mining والتنقيب في محتوى الويب web content mining. إن خوارزميات التصنيف بالأعتماد على استخدام الويب تركز على عدد زيارات الروابط بين الصفحات بشكل عام، أو الوقت الذي قضاه المستخدم في الصفحة، أو ما هي التفاعلات التي قام بها المستخدم في الصفحة المفتوحة، ولكن لم تأخذ هذه الدراسات عامل الزمن عند حساب هذه التفاعلات، اي متى حدثت هذه التفاعلات و هل هي حديثة أم قديمة.

في هذا البحث سنطرق لموضوع الزمن حيث سندخل عامل الزمن عند حساب التصنيف المتمثل في تاريخ حدوث زيارات الروابط بالإضافة إلى الإعتماد على العدد الكلي لهذه الزيارات ودراسة التأثير الإيجابي الناجم عن ذلك على عملية التصنيف وخصوصاً بالنسبة للصفحات الحديثة.

كلمات مفتاحية : تصنيف، روابط دخل، روابط خرج، عدد الزيارات، التنقيب

* مدرس في قسم هندسة تكنولوجيا المعلومات – كلية هندسة تكنولوجيا المعلومات والاتصالات –جامعة طرطوس – سوريا

** طالب دراسات عليا (ماجستير) – قسم المعلومات – كلية هندسة تكنولوجيا المعلومات و الاتصالات – جامعة طرطوس – سوريا

مجلة جامعة طرطوس للبحوث والدراسات العلمية _ سلسلة العلوم الهندسية المجلد (٧) العدد(٥) ٢٠٢٣
Tartous University Journal for Research and Scientific Studies - engineering Sciences Series Vol. (7) No. (5) 2023

Enhancement of the algorithm of weighted page rank in search engines

Dr. Ragheb Toumah*
Gibran Kashour**

(Received 23/3/2023 . Accepted 11/7/2023)

□ **ABSTRACT**

Search engines retrieve a large number of possible pages containing relevant information based on user queries, and for this purpose, there are various classification algorithms that rank the returned results according to their importance. Most classification algorithms focus on web structure mining, while others incorporate the concepts of web usage mining and web content mining. Web-based classification algorithms focus on factors such as the number of visits to links between pages, the time spent by the user on the page, and the interactions the user had on the open page, but these studies do not take into account the factor of time when calculating these interactions, i.e., when they occurred and whether they are recent or old.

In this study, we will address the issue of time by incorporating it into the classification process, specifically by considering the date of link visits in addition to the total number of these visits, and studying the positive impact this has on the classification process, especially for newer pages.

Keywords: classification, incoming links, outgoing links, number of visits, mining

*** Lecture In Department of Information Technology Engineering - Faculty of Information and Communication Technology Engineering - University of Tartous - Syria**
**** Graduate Student (Master) - Department of Information - Faculty of Information and Communication Technology Engineering - University of Tartous - Syria**

مقدمة

إن شبكة الأنترنت تحتوي على ملايين من الصفحات والتي يتضاعف أعدادها بشكل كبير مع مرور الوقت، الأمر الذي خلق تحدي كبير لمحركات البحث من أجل إعادة أفضل النتائج المطابقة لاستعلام المستخدم. من أجل ذلك تستخدم هذه المحركات العديد من خوارزميات التصنيف لترتيب الصفحات المعادة حسب أهميتها ومدى صلتها باستعلام المستخدم. يوجد مجموعة واسعة من الخوارزميات التي تم تطويرها مثل PageRank, HITS, SALSA, RANDOMZE HITS...الخ.

إن الخوارزمية المقترحة في هذا البحث تأخذ بعين الأعتبار أهمية الصفحات خلال الفترة الزمنية السابقة وذلك لإعطاء الصفحات الحديثة والتي لها عدد طلبات متنامي فرصة أكبر لتظهر في التصنيف دون التقليل من أهمية الصفحات القديمة والتي ما زالت أهميتها عالية.

وأسمينا هذه الخوارزمية المقترحة Weighted PageRank based on number of visits of links with time factor

الدراسات السابقة

Brin و **Page** [1] قاموا باستخدام بنية الروابط في صفحات الويب من أجل تقييم الصفحات. خوارزمية PageRank أستخدمت في google لترتيب النتائج المعادة نتيجة عمليات البحث . هذه الخوارزمية تعمل على مبدأ : إذا احتوت الصفحة روابط دخل مهمة إذاً روابط الخروج والتي تشير إلى صفحات أخرى تُعتبر مهمة أيضًا ولهذا فهي تعتمد على روابط الدخل من أجل حساب التقييم كما في المعادلة (١)

$$PR(u) = (1-d) + d \sum_{v \in B(u)} \left(\frac{PR(v)}{Nv}\right) \qquad (1)$$

حيث u تمثل صفحة الأنترنت، PR(u) و PR(v) تمثلان تقييم الصفحتين u و v على التوالي، B(u) تمثل مجموعة الصفحات التي تشير الى الصفحة u، N_v تمثل مجموع الروابط الخارجة من الصفحة v .

d هي معامل التخميد damping factor والتي تمثل أحتمالية استخدام المستخدم لروابط مباشرة وهي بين ٠ و ١.

Wenpu Xing و **Ali Ghorbani** [2] قدما خوارزمية اسمها Weighted PageRank وهي تطوير للخوارزمية السابقة. حيث كلما كانت الصفحات أكثر طلباً فإن الصفحات الأخرى تميل لتشكيل روابط معها، ففي هذه الخوارزمية يتم إسناد تصنيف أعلى إلى الصفحات الأكثر أهمية (طلباً) بدل من تقسيم تصنيف الصفحة بالتساوي على روابطها الخارجية. فكل صفحة يُشار لها تحصل على جزء من التصنيف مساوي لأهميتها (عدد الروابط الداخلة لها و الخارجة منها).

$W^{in}(v,u)$ وهي تمثل أهمية روابط الدخل

$$W^{in}(v,u) = \frac{I_u}{\sum_{p \in R(v)} I_p} \qquad (2)$$

I_u و I_v تمثل عدد الروابط الداخلة إلى الصفحتين u و v على التوالي، R(v) تمثل مجموعة الصفحات التي يشار لها من الصفحة v

$$W^{out}(v,u) = \frac{O_u}{\sum_{p \in R(v)} O_p} \qquad (3)$$

O_u و O_v تمثل عدد الروابط الخارجة من الصفحتين u و v على التوالي، R(v) تمثل مجموعة الصفحات التي يشار لها من الصفحة v

فيكون الشكل العام للمعادلة كالتالي

$$PR(u) = (1-d) + d \sum_{v \in B(u)} PR(v) W^{in}(v,u) W^{out}(v,u) \qquad (4)$$

Gyanendra Kumar et. al. [3] استخدموا سلوك تصفح المستخدم في حساب تقييم الصفحة. في هذه الخوارزمية The Page Ranking based on Visits of Links (VOL) يتم إعطاء تصنيف أعلى إلى الروابط الخارجة والتي تكون أكثر زيارة من قبل المستخدم أي ان تصنيف الصفحة يتم حسابه بالأعتماد على عدد زيارات الروابط الداخلة كما في المعادلة التالية

$$PR(u) = (1-d) + d \sum_{v \in B(u)} L_u \frac{PR(v)}{TL(v)} \qquad (5)$$

L_u تشير الى عدد زيارات الرابط الذي يشير من الصفحة v الى الصفحة u، TL(v) هي العدد الكلي للزيارات لكل الروابط من الصفحة v.

Neelam Tyagi and Simple Sharma [4] قاموا بتضمين سلوك تصفح المستخدم في خوارزمية Weighted PageRank لتطوير خوارزمية جديدة سميت ب Weighted PageRank based on number of visits of links (VOL). في هذه الخوارزمية يتم إعطاء تصنيف أعلى للروابط التي لها قيمة VOL عالية. هي تأخذ فقط في عين الأعتبار شعبية روابط الدخل وتتجاهل شعبية روابط الخرج. فيكون الشكل العام للمعادلة كالتالي

$$WPR_{vol}(u) = (1-d) + d \sum_{v \in B(u)} L_u \frac{WPR_{vol}(v) W^{in}(v,u)}{TL(v)} \qquad (6)$$

حيث $WPR_{vol}(u)$،$WPR_{vol}(u)$ تمثلان تقييم الصفحتين u و v على التوالي محسوبين باستخدام الخوارزمية Weighted PageRank based on number of visits of links (VOL).

L_u تشير الى عدد زيارات الرابط الذي يشير من الصفحة v الى الصفحة u، TL(v) هي العدد الكلي للزيارات لكل الروابط من الصفحة v.

$W^{in}(v,u)$ وهي تمثل أهمية روابط الدخل

$$W^{in}(v,u) = \frac{I_u}{\sum_{p \in R(v)} I_p} \qquad (2)$$

I_u و I_v تمثل عدد الروابط الداخلة إلى الصفحتين u و v على التوالي، R(v) تمثل مجموعة الصفحات التي يشار لها من الصفحة v .

Sonal Tuteia [5] قام بتحديث خوارزمية Weighted PageRank بحيث تتضمن سلوك تصفح المستخدم عند حساب تقييم الصفحة وقام بتسميتها ب Enhancement in Weighted PageRank Algorithm Using VOL وذلك من خلال استخدام المفهومين التاليين:

- $W^{in}_{VOL}(v,u)$ وهي تمثل الأهمية المحسوبة من خلال عدد زيارات الروابط الداخلة وتعطى بالعلاقة

$$W_{VOL}^{in}(v,u) = \frac{I_{u(VOL)}}{\sum_{p\in R(v)} I_{p(VOL)}} \qquad (7)$$

$I_{u(VOL)}$ و $I_{p(VOL)}$ تمثلان عدد زيارات الروابط الداخلة الى الصفحتين u و p على التوالي

- $W_{VOL}^{out}(v,u)$ وهي تمثل الأهمية المحسوبة من خلال عدد زيارات الروابط الخارجة وتعطى بالعلاقة

$$W_{VOL}^{out}(v,u) = \frac{O_{u(VOL)}}{\sum_{p\in R(v)} O_{p(VOL)}} \qquad (8)$$

$O_{u(VOL)}$ و $O_{p(VOL)}$ تمثلان عدد زيارات الروابط الخارجة من الصفحتين u و p على التوالي

فيكون الشكل العام للمعادلة كالتالي

$$EWPR_{VOL}(u) = (1-d) + d \sum_{v\in B(u)} EWPR_{VOL}(u) W_{VOL}^{in}(v,u) W_{VOL}^{out}(v,u) \qquad (9)$$

حيث $EWPR_{vol}(u)$،$EWPR_{vol}(u)$ تمثلان تقييم الصفحتين u و v على التوالي محسوبين باستخدام الخوارزمية Enhancement in Weighted PageRank Algorithm Using VOL.

أهمية البحث وأهدافه

أن خوارزميات التصنيف بالأعتماد على استخدام الويب تركز في حساباتها على عدد زيارات الروابط بين الصفحات بشكل عام حيث يتم حساب مجموع عدد الزيارات منذ تاريخ بداية إنشاء هذه الروابط ودون الأهتمام بالتاريخ الدقيق لحدوث هذه الزيارات وهنا ستظهر مشكلة وهي أن كل الصفحات القديمة الحاوية على روابط داخلة إليها وخارجة منها فأن هذه الروابط بحكم عامل الزمن (أي لأنه تم إنشائها منذ فترة زمنية طويلة) فأنها ستحوي عدد كبير من الزيارات.

وعند مقارنة الصفحات القديمة مع الصفحات الجديدة وذلك حتى لو كان مجموع عدد الزيارات الحديثة (اي زيارات حدثت في فترة زمنية سابقة ولكنها حديثة نسبياً وليس منذ تاريخ إنشاء الصفحة) لهذه الصفحات الجديدة أكبر من عدد الزيارات الحديثة للصفحات القديمة ولكن بجمع عدد الزيارات الكلي للصفحات الجديدة ولكون هذه الصفحات الجديدة اقل عمرا من الصفحات القديمة سيكون عدد زيارات روابطها أقل من الصفحات القديمة وهذا سيعطيها تصنيف أقل.

وبهدف حل المشكلة السابقة فإننا في هذا البحث سنقوم بإدخال عامل الزمن في عملية حساب التصنيف بالإضافة إلى الأعتماد على بنية الروابط بين الصفحات .

منهجية البحث

لقد رأينا انه في الخوارزميات السابقة مثل Weighted PageRank فقد تم الأعتماد على بنية الويب فقط او كما في الخوارزمية Enhancement in Weighted PageRank Algorithm Using VOL تم إدخال سلوك تصفح المستخدم مع بنية الويب وذلك من خلال الاعتماد على عدد زيارات الروابط عند حساب تقييم الصفحة ولكن دون الأهتمام بعامل الزمن والصفحات الرائجة حالياً.

في الخوارزمية المقترحة تم تقديم مفهوم جديد $W_{VOL}^{in-recent}(v,u)$ وهو يمثل أهمية الصفحات محسوبة من خلال حساب عدد زيارات الروابط الداخلة خلال فترة زمنية محددة (مثلاً خلال أخر ثلاث سنوات حيث كلما قلت الفترة الزمنية كلما ارتفع تأثير أزدياد الطلب على الصفحة في حساب التقييم) فتصبح المعادلة بالشكل التالي:

$$PR(u) = (1-d) + d \sum_{v \in B(u)} \frac{PR(u) W_{VOL}^{in}(v,u) W_{VOL}^{out}(v,u) W_{VOL}^{in-recent}(v,u)}{CT(u)}$$

حيث أن :

- $W_{VOL}^{in}(v,u)$ وهي تمثل الأهمية المحسوبة من خلال عدد زيارات الروابط الداخلة
- $W_{VOL}^{out}(v,u)$ وهي تمثل الأهمية المحسوبة من خلال عدد زيارات الروابط الخارجة
- $W_{VOL}^{in-recent}(v,u)$ وهي تمثل الأهمية المحسوبة من خلال عدد زيارات الروابط الداخلة

وتحسب بنفس طريقة حساب $W_{VOL}^{in}(v,u)$ مع فرق إن عدد زيارات الروابط تكون محسوبة خلال فترة زمنية محددة

- $CT(u)$ وهي عدد سنوات إنشاء الصفحة

خطوات حساب تصنيف الصفحات في الخوارزمية المقترحة

١. إيجاد موقع الويب : موقع الويب ذو العدد الكبير من الروابط التشعبية يتم أختياره لأن هذه الخوارزمية تعتمد بشكل كبير على بنية الويب

٢. توليد مخطط لهذا الموقع حيث أن العقد تمثل الصفحات والأسهم تمثل الروابط بين الصفحات

٣. حساب عدد زيارات الروابط وتاريخ إنشاء الصفحات

٤. حساب تصنيف كل صفحة وذلك من خلال حساب قيم $W_{VOL}^{in}(v,u)$, $W_{VOL}^{out}(v,u)$, $W_{VOL}^{in-recent}(v,u), CT(u)$

٥. إعادة الخطوة ٤ حتى نصل إلى قيمتي تكرارين متساويين.

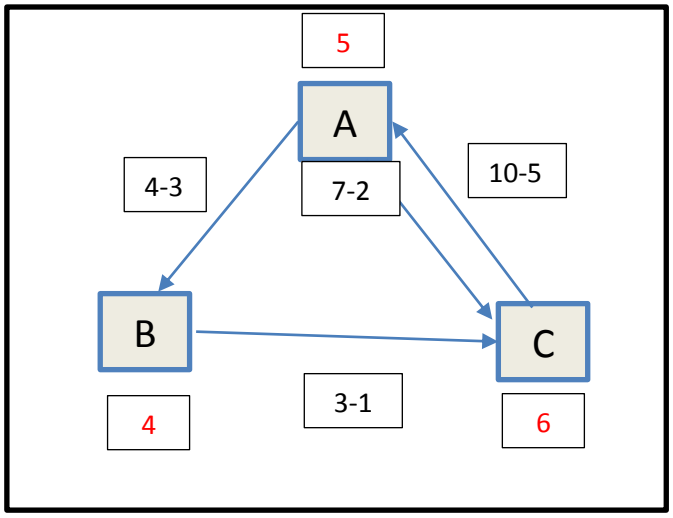
طريقة حساب عدد زيارات الروابط وتاريخ إنشاء الصفحة بالنسبة للخوارزمية المقترحة

لحساب عدد زيارات الروابط التي تدل إلى صفحة الأنترنت نستخدم client side script (وهو كود برمجي يتم تشغيله على جهاز المستخدم). في كل مرة يتم فيها الدخول الى الصفحة فأن هذا الكود سيتم تحميله على جهاز المستخدم من السيرفر ثم سيقوم هذا الكود بمراقبة نشاط المستخدم والروابط التي زارها وتسلسل الأنتقال بين هذه الروابط مع تاريخ كل زيارة ثم يقوم بإرسال المعلومات الى السيرفر وبعدها يتم قراءة هذه المعلومات بواسطة ال الزواحف أو المحركات الزاحفة Crawlers (و هي برامج تعمل على تصفح الإنترنت بشكل آلي وجمع المعلومات من المواقع والصفحات الإلكترونية، وتخزينها في قواعد بيانات خاصة بمحركات البحث لتسهيل عملية البحث عن المعلومات في المستقبل)

كذلك يجب في كل صفحة أنترنت كتابة تاريخ النشر بشكل واضح ضمن ال meta tags لكي تقوم ال crawlers بجمع هذه المعلومات أيضا (meta tags هي عبارة عن عنصر من العناصر الخاصة بلغة HTML وهو يحتوي على معلومات عن الصفحة ولا تظهر هذه المعلومات على الشاشة عند عرض الصفحة في المتصفح بل هي معلومات تفيد المتصفح عند تحليله لمحتويات الصفحة).

مثال يشرح ألية عمل الخوارزمية المقترحة

لشرح ألية عمل الخوارزمية المقترحة سنأخذ في هذا المثال – المخطط (١) – بنية عدة صفحات ويب مترابطة مع بعضها عن طريق مجموعة من الروابط التشعبية حيث أن المربعات الزرقاء تمثل صفحات الأنترنت والأسهم هي الروابط بين الصفحات (الصفحة A تشير إلى الصفحتين B و C والصفحة B تشير إلى الصفحة C والصفحة C تشير إلى الصفحة A) أما المربعات التي تحوي أرقام حمراء تمثل عدد السنين منذ تاريخ إنشاء الصفحات (الصفحة A أنشئت منذ ٥ سنوات والصفحة B أنشئت منذ ٤ سنوات والصفحة C أنشئت منذ ٦ سنوات) وعلى كل رابط يوجد رقمين حيث أن الرقم على اليسار هو عدد زيارات الرابط منذ تاريخ إنشائه و الرقم على اليمين هو عدد زيارات الرابط خلال فترة زمنية محددة (مثلا أخر ثلاث سنوات)



المخطط (١) – مخطط مبسط لشرح ألية عمل الخوارزمية المقترحة

$$PR(A) = (1-d) + d\left(\frac{PR(C)\, W_{VOL}^{in}(C,A) W_{VOL}^{out}(C,A) W_{VOL}^{in-recent}(C,A)}{CT(C)}\right)$$

$$PR(B) = (1-d) + d\left(\frac{PR(A)\, W_{VOL}^{in}(A,B) W_{VOL}^{out}(A,B) W_{VOL}^{in-recent}(A,B)}{CT(A)}\right)$$

$$PR(C) = (1-d) + d\left(\frac{PR(A)\, W_{VOL}^{in}(A,C) W_{VOL}^{out}(A,C) W_{VOL}^{in-recent}(A,C)}{CT(A)} + \frac{PR(B)\, W_{VOL}^{in}(B,C) W_{VOL}^{out}(B,C) W_{VOL}^{in-recent}(B,C)}{CT(B)}\right)$$

$$W_{VOL}^{in}(C,A) = \frac{I_A}{I_A} = \frac{10}{10} = 1$$

$$W_{VOL}^{out}(C,A) = \frac{O_A}{O_A} = \frac{4}{4} = 1$$

$$W_{VOL}^{in-recent}(C,A) = \frac{I_A}{I_A} = \frac{5}{5} = 1$$

$$W_{VOL}^{in}(A,B) = \frac{I_B}{I_B + I_C} = \frac{4}{4+10} = \frac{4}{14}$$

$$W_{VOL}^{out}(A,B) = \frac{O_B}{O_B + O_C} = \frac{3}{3+10} = \frac{3}{13}$$
$$W_{VOL}^{in-recent}(A,B) = \frac{I_B}{I_B + I_C} = \frac{3}{3+3} = \frac{3}{6}$$
$$W_{VOL}^{in}(A,C) = \frac{I_C}{I_B + I_C} = \frac{10}{4+10} = \frac{10}{14}$$
$$W_{VOL}^{out}(A,C) = \frac{O_C}{O_B + O_C} = \frac{10}{3+10} = \frac{10}{13}$$
$$W_{VOL}^{in-recent}(A,C) = \frac{I_C}{I_B + I_C} = \frac{3}{3+3} = \frac{3}{6}$$
$$W_{VOL}^{in}(B,C) = \frac{I_C}{I_C} = \frac{10}{10} = \frac{10}{10} = 1$$
$$W_{VOL}^{out}(B,C) = \frac{O_C}{O_C} = \frac{10}{10} = \frac{10}{10} = 1$$
$$W_{VOL}^{in-recent}(B,C) = \frac{I_C}{I_C} = \frac{3}{3} = 1$$

$$CT(B) = 4 \qquad CT(C) = 6 \qquad CT(A) = 5$$

إذا كانت d=0.85 عندئذ يكون تصنيف الصفحات كالتالي :

$$PR(A) = (1 - 0.85) + 0.85\left(\frac{1 \times 1 \times 1 \times 1}{6}\right) = 0.29167$$

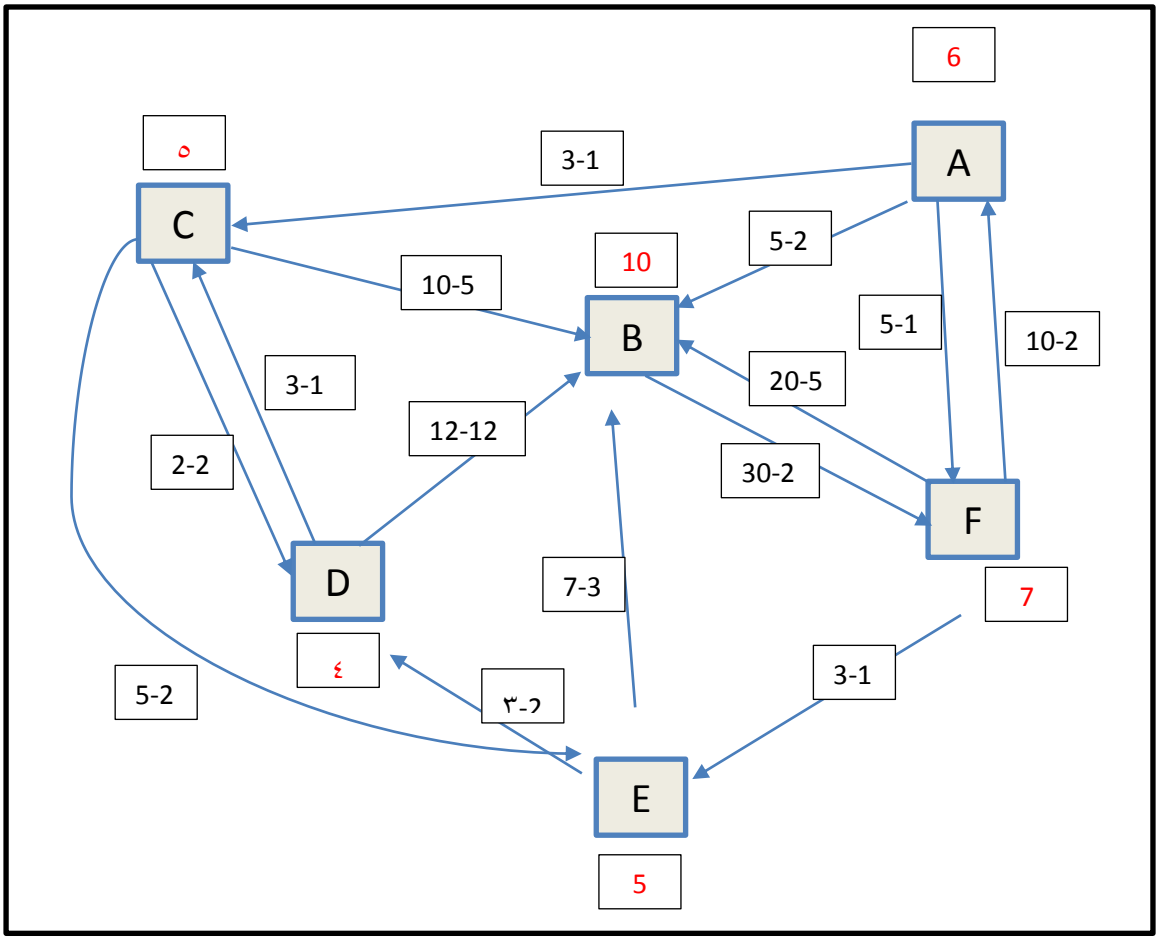
$$PR(B) = (1 - 0.85) + 0.85\left(\frac{0.29167 \times \frac{4}{14} \times \frac{3}{13} \times \frac{3}{6}}{5}\right) = 0.15163$$

$$PR(C) = (1 - 0.85) + 0.85\left(\frac{0.29167 \times \frac{10}{14} \times \frac{10}{13} \times \frac{3}{6}}{5} + \frac{0.15163 \times 1 \times 1 \times 1}{4}\right)$$
$$= 0.19584$$

ثم يتم تكرار هذه العملية حتى تتساوى قيم التصنيف للصفحات لتكرارين متتاليين ويحدث هذا التطابق عند التكرار السادس ويصبح تقييم الصفحات كالتالي :

$$PR(A) = 0.17696$$
$$PR(B) = 0.15099$$
$$PR(C) = 0.19035$$

قمنا بتجربة الخوارزمية المقترحة على مخطط أكبر – المخطط (٢) – وتم مقارنة النتائج مع تطبيق خوارزمية Enhancement in Weighted PageRank Algorithm Using VOL على نفس المخطط



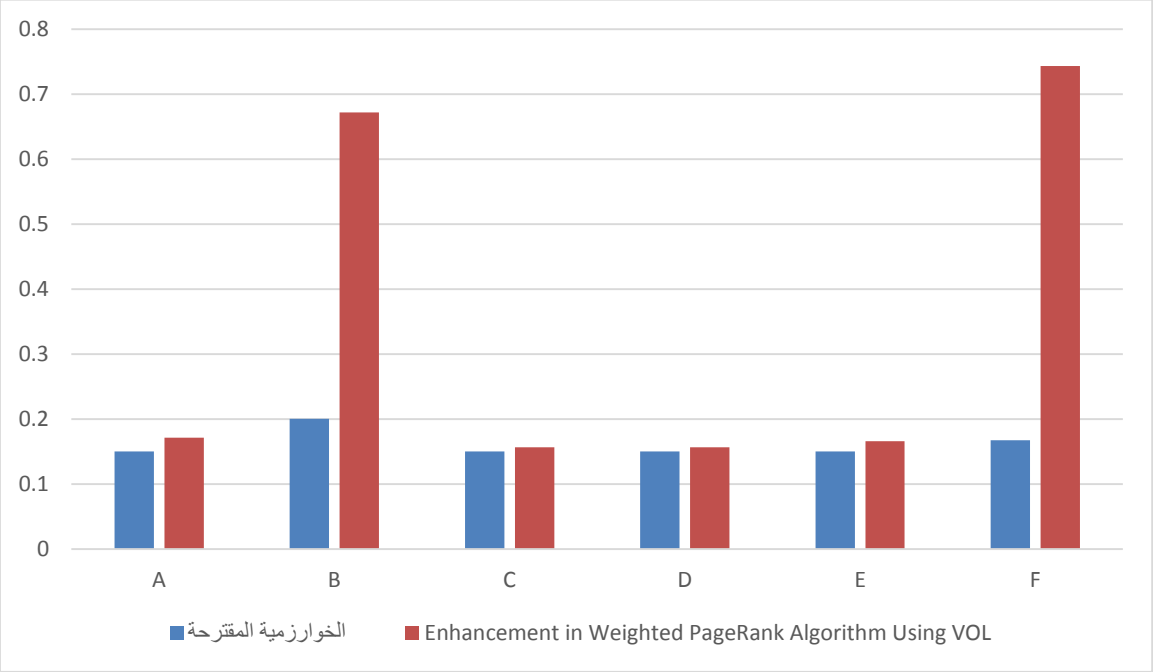
المخطط (٢) – بنية أكثر تعقيداً لصفحات الويب

عند تطبيق الخوارزمية المقترحة و أيضا الخوارزمية Enhancement in Weighted PageRank Algorithm Using VOL وذلك من أجل المقارنة عند d=0.85 وبعد ٧ تكرارت نصل ألى النتيجة التالية وهي تمثل تقييم صفحات الويب بعد تطبيق الخوارزميتين

الجدول (١) – تقييم صفحات الويب بعد تطبيق الخوارزميات

تقييم صفحات الويب بعد تطبيق الخوارزمية المقترحة					
F	E	D	C	B	A
٠.١٦٧٣٥٦٨٩	٠.١٥٠٠٨٨٨٢	٠.١٥٠١٥٤١	٠.١٥٠٠٩٧٤٢	0.2006358	٠.١٥٠٠٤٣٢٧
تقييم صفحات الويب بعد تطبيق الخوارزمية Enhancement in Weighted PageRank Algorithm Using VOL					
F	E	D	C	B	A
٠.٧٤٣١٩	٠.١٦٦١٩	٠.١٥٦٧١	٠.١٥٦٧٨	٠.٦٧١٧٩	٠.١٧١٦١

المخطط (٣) –تمثيل بياني لتقييم صفحات الويب بعد تطبيق الخوارزميات



المخطط البياني السابق – المخطط (٣) – يمثل نتائج تطبيق الخوارزمية المقترحة و الخوارزمية Enhancement in Weighted PageRank Algorithm Using VOL على بنية صفحات الأنترنت الموجودة في – المخطط (٢) – حيث تقييم الصفحات باللون البرتقالي هو نتيجة عمل خوارزمية Enhancement in Weighted PageRank Algorithm Using VOL أما تقييم الصفحات باللون الأزرق هو نتيجة عمل الخوارزمية المقترحة حيث نجد أنه في الخوارزمية المقترحة أصبح تقييم جميع الصفحات متقارب وليس كما في الخوارزمية السابقة حيث تقييم الصفحتين B و F هو أكبر بشكل كبير من باقي الصفحات ولكن الذي يهم هنا حقاً هو ترتيب هذا التقييم لأنه حسب هذا الترتيب ستظهر الصفحات في نتائج البحث وفي – الجدول (٢) – تم ترتيب هذه الصفحات من الأكبر تصنيفاً إلى الأقل تصنيفاً.

الجدول (٢) – ترتيب أهمية صفحات الويب بعد تطبيق الخوارزميات

ترتيب الصفحات من الأكبر تصنيف إلى الأقل تصنيف					
الأكبر تصنيف ← الأقل تصنيف					
الخوارزمية المقترحة					
A	E	C	D	F	B
Enhancement in Weighted PageRank Algorithm Using VOL					
D	C	E	A	B	F

تحليل النتائج

- لم يتغير ترتيب الصفحتين B و F كثيرا عند تطبيق الخوارزمية المقترحة حيث حافظا على ترتيبهما في أعلى التصنيف (أصبحت الصفحة B أكبر تصنيفا من F) حيث على الرغم من أنهما أقدم صفحتين ولكنهما حافظا على عدد زيارات كبير حديثا أي ما زالت أهميتهما كبيرة.
- الصفحة A تراجع ترتيبها كثيرا وذلك لأن أغلب عدد الزيارات لها كانت قديمة حيث أن المجموع الكلي لعدد الزيارات منذ تاريخ إنشاء الصفحة هو ١٠ ولكن عدد الزيارات الحديثة هو ٢ فقط وهذا يعني أن أهمية الصفحة حديثا أنخفضت وهذا اثر على تصنيفها بشكل واضح حيث أحتلت الترتيب الثالث في الخوارزمية Enhancement in Weighted PageRank Algorithm Using VOL ولكن الأن في الخوارزمية المقترحة أصبحت في الترتيب الأخير.
- ترتيب D يبين التأثير الأيجابي للخوارزمية المقترحة على عملية التصنيف حيث على الرغم من العمر الحديث للصفحة (٤ سنوات فقط وهي أصغر الصفحات عمراً) ولكن ترتيبها أزداد بشكل ملحوظ وأنتقلت من الترتيب الأخير عند تطبيق الخوارزمية Enhancement in Weighted PageRank Algorithm Using VOL إلى الترتيب الثالث عند تطبيق الخوارزمية المقترحة والسبب في ذلك هو عدد الزيارات الحديثة الكبير.
- ترتيب الصفحتين C و E لم يتغير كثيرا

اي نستنتج أنه عند تطبيق الخوارزمية المقترحة :

١. الصفحات القديمة ذات عدد الزيارات الحديثة كبير نسبياً ستحافظ على ترتيب تصنيف مرتفع ولكن اذا كانت عدد الزيارات الحديثة لهذه الصفحات القديمة قليل عندئذ فأن تصنيف الصفحة سينخفض بشكل ملحوظ.

٢. الصفحات الحديثة التي لها عدد زيارات مرتفع فأن تصنيفها سيرتفع بشكل كبير وهو من أهم إيجابيات الخوارزمية المقترحة

الأستنتاجات والعمل المستقبلي

الشيء المميز في هذه الخوارزمية المقترحة بأنها تحسب التصنيف ليس فقط بالأعتماد على عدد الزيارات بل أيضا على تاريخ هذه الزيارات وتاريخ إنشاء الصفحة مما يعطي نتائج أكثر دقة لأنها تهتم بالأهمية الحالية المتنامية للصفحة وليس فقط بأهمية الصفحة بشكل عام منذ تاريخ إنشائها وذلك لأن الصفحة التي كانت مهمة سابقا ممكن أن تصبح أقل أهمية حاليا وذلك يمكن ملاحظته من خلال عدد الزيارات الحديثة لان عدد الزيارات الكلي دون تحديد تاريخ حدوث هذه الزيارات يمكن أن يعطي معلومات مضللة نوعا ما.

بعض التحسينات التي يمكن تطبيقها على الخوارزمية مستقبلا:

- يمكن أولا تطبيق بعض الخوارزميات التي تُستخدم في أنظمة إستعادة المعلومات وذلك لتجميع الصفحات حسب المواضيع التي تحتويها [7] ثم يتم إعادة تشكيل المخطط الذي يمثل الصفحات والروابط بين الصفحات لكل مجموعة صفحات لها نفس الموضوع على حدى وبعد ذلك يتم تطبيق الخوارزمية المقترحة
- الخوارزمية الحالية تستخدم فقط معلومات زيارات الروابط من المستخدم يمكن مثلا إضافة معلومات أخرى مثل معلومات يمكن الحصول عليها من محركات البحث مثل أي صفحة أختارها المستخدم من بين النتائج المعادة.
- يمكن تنفيذ تجارب أكثر على مجموعة أكبر من صفحات الويب لكي تعطي نتائج أكثر دقة مما يعطي تقييم أفضل لعمل هذه الخوارزمية
- يمكن تحسين دقة الخوارزمية من خلال إضافة عوامل أخرى عند حساب تصنيف الصفحات مثل وقت تحميل الصفحة وعدد الصور والوسائط التي تحويها ووقت الاستخدام ووقت القراءة.

المراجع

1. S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. Computer Networks and ISDN Systems, 30(1–7):107–117, 1998

2. Wen u Xing and Ghorbani Ali, Weighted PageRank Algorithm, Proceedings of the Second Annual Conference on Communication Networks and Services Research NS ’04 , IEEE, 2004

3. Gyanendra Kumar, Neelam Duahn, and Sharma A. K., Page Ranking Based on Number of Visits of Web Pages, International onference on Computer & Communication Technology (ICCCT), 2011

4. Neelam Tyagi, Simple Sharma, Weighted Page Rank Algorithm Based on Number of Visits of Links of Web Page, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-3, July 2012

5. Sonal Tuteja, Enhancement in Weighted PageRank Algorithm Using VOL, IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661, p-ISSN: 2278-8727Volume 14, Issue 5. 2013

6. Isha Mahajan, Sachin Gupta, Ms. Harjinder Kaur, Dr. Darshan Kumar, Extended Weighted Page Rank Based on VOL by Finding User Activities Time and Page Reading Time, Storing them Directly on Search Engine Database Server, International Journal of Engineering Works, Vol. 4, Issue 2, PP. 41-48, February 2017

7. Loubna Ali, Shkelqim Hajrulla, Improving Information Retrieval Systems' Efficiency, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, Vol. 11 Issue 11, November-2022