

**Using Web-links Information and Blocking
Percentages to Improve the Internet Filtering
in Saudi Arabia**

BY

Abeer Basheer Zimmo

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF COMPUTER SCIENCE

KING ABDULAZIZ UNIVERSITY

JEDDAH
1427 H
2007 G

ACKNOWLEDGMENTS

All praise is for Allah. We praise Him and seek His help. I am grateful to Allah for helping me choosing this thesis and giving me the strength to complete it.

I thank my supervisor, Dr. Mohammad Ameen for his guidance. He allowed me to form my own opinions, while providing helpful pointers that were instrumental to the completion of this thesis.

This thesis was completed with the help of many great people. First and foremost I would like to thank the members of King Abdulaziz City for Science and Technology especially Mr. Hamad Al-Sulaiym the Director of Information Security Center and Mr. Abdullah Al-Abdullatif for their support and assistance by providing me valuable insight and important information.

I also have a great deal of appreciation for the help provided by my friends Suhad Qablawi and Lyla Shahbar to complete this thesis.

Last but not least, I owe a great deal of gratitude for the support provided by all my family.

I would like to dedicate this thesis to my parents, my husband and my children. To my parents I cannot ever repay them for the support and prays they gave

and are still giving me. To my husband I cannot express in words the appreciation I have for his patience, support and guidance over the past several years. To my children I thank them for their patience and understanding.

ABSTRACT

People in Saudi Arabia are using the Internet increasingly. An endless array of subjects is available on the Internet, including sex, pornography, hate, and violence is all prohibited in the Saudi Arabian culture and by the religion of Islam. Therefore, the Saudi Arabian government authorized King Abdulaziz City for Science and Technology (KACST) to control and filter the Internet before providing it to the Internet Service Providers “ISP’s” in order to protect people from undesired material. This endeavor to keep the Internet clean has been successfully done. Nevertheless in some cases, this attempt was under-inclusive “failed to block access to sites with objectionable material”, or over-inclusive “blocked access to sites with non-objectionable material”. This shortcoming is inherent in any filtering device though; in other words, there is no full proof filter of any sort.

This thesis investigated the aforementioned issues and developed a program entitled “Suggest Unblocking Blocking Lists Creator program” or “**SUBLC**” to generate two lists that propose some web pages to be unblocked and others to be blocked. A Web site or page often has links to other web sites or pages in the same domain of interest. Hence, information obtained from these reference sites can be used to assist in the filtering process.

SUBLC took advantage of the relationship between web pages by using their links to each other, and the blocking percentage of each one to suggest blocking and unblocking of web pages. The experiment was done on a sample of 1911 web pages,

which were drawn from the Internet from Yahoo categories and sub-categories and from searching for some specific objectionable and non-objectionable words by using different search engines such as Google, AltaVista and HotBot. The experiment depended heavily on the blocking percentages of the web page and on the links between these web pages.

Upon examination of these lists by KACST it was found that incorporating this new method of filtering with the existing method would improve the current filtering capability. In fact, KACST has taken advantage of the SUBLC results which implies that improvement of existing filtering systems can be achieved.

إستخدام معلومات روابط الإنترنت والنسب المئوية للحجب
لتحسين فعالية ترشيح الإنترنت
في المملكة العربية السعودية

مقدمة من

عبير بشير زمو

قدمت هذه الرسالة استكمالاً لمتطلبات الحصول على درجة الماجستير
في العلوم

قسم علوم الحاسب الآلي - كلية الحاسب وتقنية المعلومات
جامعة الملك عبد العزيز
جدة

1427هـ - 2007م

شكر وتقدير

الحمد لله تعالى الذي أعانني على إختيار وتكملة هذه الرسالة على أكمل وجه. الحمد لله تعالى على أن مدني بالصبر لأكمل هذه الرسالة. الحمد لله تعالى على أن سخر لي من يساعدي من خلقه.

أتوجه بالشكر الجزيل للمشرف على رسالتي الدكتور محمد أمين على جهوده العظيمة ومساعدتي لإتمام هذه الرسالة وذلك بتقديم التوجيهات اللازمة لذلك.

أتقدم أيضاً بالشكر والعرفان بالجميل للعاملين في مدينة الملك عبد العزيز للعلوم والتقنية. وأخص بالشكر مدير مركز أمن المعلومات المهندس حمد السليم، والسيد عبد الله العبد اللطيف اللذان بذلا جهداً كبيراً في مدّي بما يفيد ويلزم لإكمال هذه الرسالة.

أتوجه بالشكر الجزيل لجميع من قدم لي المساعدة أو الرأي أو ساهم لإكمال الرسالة وأخص بالشكر صديقاتي سهاد قبلاوي وليلى شحبر.

وأخيراً وليس آخراً أتوجه بالشكر والإمتنان العظيمين لجميع أفراد عائلتي وخاصة والدي العزيزين وزوجي الكريم وأولادي وبناتي وإخواني وأخواتي وأم زوجي ولجميع من قام بالدعاء لي بالتوفيق.

أود أن أهدي هذه الرسالة إلى والديّ العزيزين وإلى زوجي العزيز وإلى أولادي
وبناتي الأحباء. فمهما قلت فلن تسعني الكلمات ومهما عملت فلن أوفيهم حقهم على صبرهم
ودعائهم ودعمهم لي حتى أكمل هذه الرسالة على أكمل وجه، فجزاهم الله عني خير الجزاء.

المستخلص

لقد زاد استخدام الناس للإنترنت في المملكة العربية السعودية. يوجد على الإنترنت العديد من الموضوعات المتنوعة بما في ذلك موضوعات متعلقة بالجنس والكرهية والعنف وغيرها من الموضوعات المحظورة في المملكة العربية السعودية بسبب الدين والعادات. لذلك فإن الحكومة السعودية أوكلت مدينة الملك عبد العزيز للعلوم والتقنية (المدينة) بمهمة ترشيح الإنترنت والتحكم بها قبل توفيرها لمقدمي خدمات الإنترنت من أجل حماية الناس من الموضوعات الغير مرغوب فيها. وقد نجحت المدينة في ترشيح الإنترنت وتقديمه للناس بشكل يليق بدينهم وعاداتهم ، إلا أن هناك بعض المواقع والصفحات الإباحية والغير مرغوب فيها قد فشلت المدينة في حجبتها وهذا ما يسمى " بحجب غير كاف أو حجب أقل من المطلوب" بالإضافة إلى أن هناك بعض المواقع والصفحات المفيدة والعلمية والتي قامت المدينة بحجبها خطأ وهذا ما يسمى "حجب أكثر من المطلوب". وهذا الخلل كامن وموجود في أي برنامج ترشيح معروف. وبعبارة أخرى "ليس هناك برنامج ترشيح كامل من دون أخطاء من أي نوع".

لقد قمنا في هذه الرسالة بعمل برنامج يسمى "برنامج إنشاء قوائم لإقتراح الحجب أو الفتح" أو "سبلك" ليقوم بإنشاء قائمتين لإقتراح حجب بعض صفحات الإنترنت المفتوحة وفتح غيرها من الصفحات المحجوبة.

تحتوي مواقع وصفحات الإنترنت عادة على روابط تنقل متصفح الموقع إلى مواقع وصفحات أخرى تحتوي على نفس المواضيع ونفس مجال الصفحة الأساسية . لذلك فإن

المعلومات المستقاة من هذه المواقع والصفحات المرجعية يمكن استخدامها من أجل المساعدة في عملية الترشيح.

قام برنامج "سبلك" باستغلال العلاقة بين صفحات الإنترنت من حيث الروابط بين هذه الصفحات بعضها ببعض واستخدام النسبة المئوية للحجب لكل صفحة وذلك لإقتراح حجب أو فتح هذه الصفحات . أجريت التجربة على عينة من 1911 صفحة من الإنترنت مستمدة من الفئات الأساسية والفرعية لموقع Yahoo ومن البحث عن بعض الكلمات المعترض وغير المعترض عليها باستخدام محركات البحث مثل Google, AltaVista, HotBot. وقد اعتمدت التجربة اعتماداً كبيراً على النسب المئوية للحجب في هذه الصفحات وعلى الروابط بين الصفحات وبعضها البعض.

بعد فحص هذه القوائم المقترحة من "سبلك" بواسطة المدينة تبين أن إضافة هذه الطريقة الجديدة للترشيح إلى الطريقة المتبعة حالياً يمكن أن تحسن نظام الترشيح المقام حالياً. وبالفعل استخدمت المدينة النتائج المقترحة من برنامج "سبلك" في الترشيح مما يعني أنه بالإمكان تحسين أنظمة الترشيح القائمة حالياً.