

Aplicación del algoritmo de PageRank para determinar la importancia de los estándares del Comité 802 del IEEE

Luis Enrique Balderas A., Salvador Martínez D., Gonzalo Mena M., Ricardo Chaparro S.,
Universidad Autónoma de Querétaro, Facultad de Informática

RESUMEN

La importancia de los estándares de la IEEE es un tema inherentemente subjetivo, que depende de los intereses del lector, de sus conocimientos y actitudes. Sin embargo, mucho se puede decir objetivamente de la importancia de cada estándar. Este artículo describe la utilización del algoritmo PageRank, un método de evaluar páginas web objetiva y mecánicamente, para determinar la relevancia relativa de cada uno de los estándares vigentes del comité 802 de la IEEE. Se calculó el puntaje de PageRank de cada estándar en base a las referencias explícitas y a las menciones simples hacia otros estándares, respectivamente. Se analizaron los resultados, se compararon entre ellos y contra la importancia que se esperaría que tuviera cada estándar en relación a su posición en la jerarquía de los documentos.

Palabras claves

IEEE, 802, PageRank, referencias, menciones, importancia, organización

INTRODUCCIÓN

El Comité 802 del Instituto de Ingenieros en Eléctrica y Electrónica (IEEE) definió los estándares de redes de área local (LAN) y área metropolitana (MAN). La mayoría de los estándares fueron establecidos por el comité en la década de 1980 cuando apenas comenzaban a surgir las redes entre computadoras personales. De entonces a la fecha muchos de esos estándares han sido derogados y otros más han sido actualizados.

Se buscaba una manera de organizar los estándares vigentes del Comité 802 de acuerdo a su importancia. Donde la importancia pudiera ser medida objetivamente.

Una forma de medir la importancia objetiva de un documento es en base a las citas

que se hacen de él en otros documentos. Uno de los métodos para medir dicha importancia es el algoritmo de PageRank [1].

Aunque existe una amplia literatura sobre análisis de citas de publicaciones académicas, se escogió un algoritmo que nació para la web y los aplicamos a los estándares vigentes del Comité 802 de la IEEE.

Este algoritmo fue creado por Sergey Brin y Lawrence Page cuando estudiaban su doctorado en la Universidad de Stanford, ambos posteriormente fundaron la compañía Google, cuyo motor de búsqueda esta basado en PageRank.[2]

PageRank es un algoritmo con gran aceptación y ha sido ampliamente analizado por varios investigadores.

PageRank se basa en la naturaleza democrática de la Web, al utilizar la basta estructura de enlaces como un indicador del valor de una página individual. PageRank interpreta el enlace de una página A a una página B como un voto, de la página A para la página B pero PageRank va más allá de contar el número de votos o enlaces que recibe una página; también analiza la página que emite el voto. Los votos emitidos por páginas que son “importantes” pesan más y ayudan a volver a otras páginas “importantes”.

Suponga un pequeño universo de 4 páginas: A, B, C, D. Si todas esas páginas enlazan a A, entonces el PR (PageRank) de la página A sería la suma del PR de las páginas B, C y D.

$$PR(A) = PR(B) + PR(C) + PR(D)$$

Pero supóngase que B también tiene un enlace a C y que D tiene enlaces a las otras 3 páginas. Una página no puede votar 2 veces, y por esa razón se considera que B da medio voto a A y medio voto a C. De la misma manera solo

un tercio del voto de D se cuenta para el PageRank de A.

$$PR(A) = PR(B)/2 + PR(C)/1 + PR(D)/3$$

En otras palabras divídase el PR entre el número total de enlaces que salen de esa página.

$$PR(A) = PR(B)/L(B) + PR(C)/L(C) + PR(D)/L(D)$$

Finalmente todo esto se reduce por un cierto porcentaje al multiplicarlo por un factor q . Por razones explicadas en el artículo de Brin y Page [1] ninguna página puede tener un PR de 0, de hecho el PR mínimo de una página es $1 - q$.

$$PR(A) = q[PR(B)/L(B) + PR(C)/L(C) + PR(D)/L(D)] + 1 - q$$

DESARROLLO

Para calcular el puntaje de PageRank para cada documento y ordenar los estándares de acuerdo al mismo se llevaron a cabo los siguientes pasos:

Lista de estándares a evaluar

Debido a que algunos de los estándares ya fueron derogados se determinó usar los que se encuentran vigentes a octubre de 2004 y que están disponibles en el sitio web de la IEEE [3]. La lista comprende 40 estándares que se encuentran contenidos en 39 documentos.

Obtención de las referencias de cada documento

Cabe mencionar que algunos de los documentos no incluían referencias explícitas, esto es una sección definida de referencias o de bibliografía en formato estándar, por lo que se consideraron también las menciones simples de otros estándares que aparecen a lo largo de todo el documento.

Con esto se obtuvieron dos juegos de datos, uno conformado por referencias explícitas y otro por menciones simples (que incluye a las primeras). Sólo se consideraron las citas a los documentos que se encuentran definidos en el punto anterior.

Las referencias explícitas se obtuvieron inspeccionando manualmente las secciones de referencias, bibliografía y apéndices. Las

menciones simples fueron obtenidas semiautomáticamente buscando la cadena “IEEE Std 802”, esto es suficiente debido a que los documentos son totalmente consistentes en la manera de citarse entre ellos.

Forma de calcular la importancia de cada estándar

Debido a que el algoritmo de PageRank es ampliamente difundido se decidió buscar una implementación ya existente, y se encontró una biblioteca de procesamiento de grafos en Java llamada JUNG [4] la cual incluye el algoritmo de PageRank y otras herramientas que permiten la representación gráfica de los resultados.

Formato para registro de las referencias

Debido a que la biblioteca JUNG acepta el formato PajekNet [5], que es un estándar muy utilizado para representar redes y grafos, se decidió emplearlo para almacenar las referencias.

Programa para calcular el PageRank

Se desarrolló un programa en Java para realizar el cálculo del puntaje para cada estándar y representar de manera gráfica los resultados de acuerdo a la relevancia de cada uno. El programa permite elegir entre el juego de datos de referencias explícitas y el de menciones simples.

RESULTADOS

Las columnas en las tablas de resultados son:

1. #: Representa la posición de importancia relativa del estándar.
2. PR: PageRank, puntaje calculado del documento.
3. E: Número de citas que hacen otros documentos al estándar en cuestión.
4. S: Número de citas del documento hacia otros estándares.
5. Nombre del estándar.

En la tabla 1 se muestran los resultados obtenidos en base a las referencias explícitas.

Referencias Explícitas Internas				
#	PR	E	S	Estándar
1	0.26922	19	7	802
2	0.09379	10	1	802.2
3	0.07666	9	3	802.3
4	0.07350	4	2	802.1F
5	0.04681	5	1	802.11
6	0.04265	2	4	802.12
7	0.04265	2	4	802.1H
8	0.03879	1	2	802.1G
9	0.03734	1	0	802.3ak
10	0.03632	5	8	802.1Q
11	0.03221	5	3	802.5
12	0.01523	1	1	802.5c
13	0.01140	2	7	802.16
14	0.01126	2	4	802.1X
15	0.00928	1	0	802.11a
16	0.00869	1	1	802a
17	0.00783	1	1	802.15.1
18	0.00783	1	0	802.11b
19	0.00749	1	0	802.16.2
20	0.00749	1	2	802.16a
21	0.00749	1	3	802.16c
22	0.00610	0	1	1802.3
23	0.00610	0	1	802.11b_Cor1
24	0.00610	0	2	802.11d
25	0.00610	0	4	802.11f
26	0.00610	0	1	802.11g
27	0.00610	0	3	802.15.2
28	0.00610	0	1	802.15.3
29	0.00610	0	2	802.15.4
30	0.00610	0	2	802.3aj
31	0.00610	0	1	802.5t
32	0.00610	0	3	802.5v
33	0.00610	0	0	802.11h
34	0.00610	0	0	802.16_Conformance02
35	0.00610	0	0	802.16_Conformance01
36	0.00610	0	0	802.3ae
37	0.00610	0	0	802.3af
38	0.00610	0	0	802.5j
39	0.00610	0	0	802.5w
40	0.00610	0	0	802.5r

Tabla 1

En la tabla 2 se muestran los resultados obtenidos en base a las menciones simples.

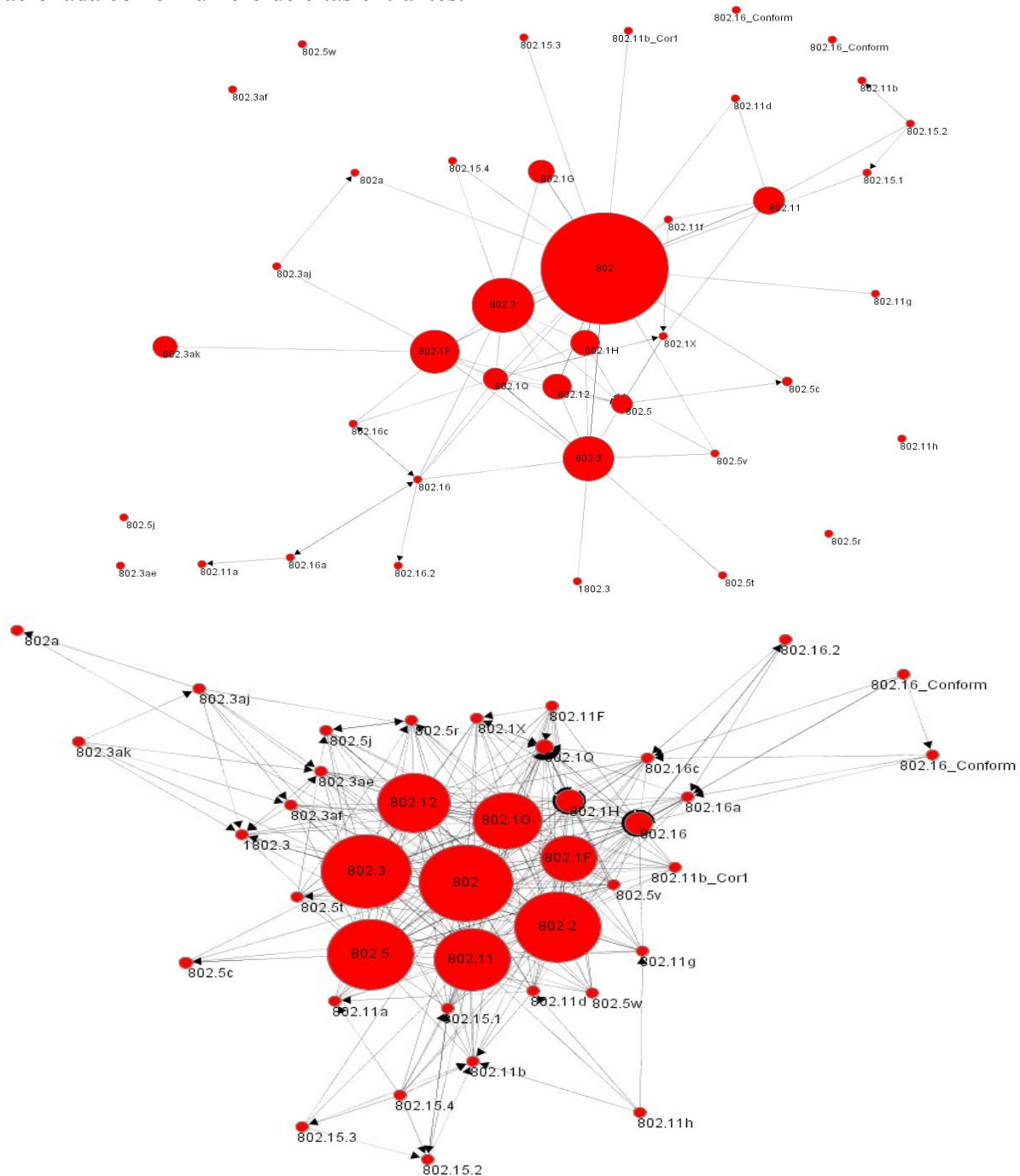
Menciones Simples Internas				
#	PR	E	S	Estándar
1	0.11191	33	10	802
2	0.10801	31	10	802.3
3	0.10378	30	7	802.2
4	0.10346	29	7	802.5
5	0.09069	31	7	802.11
6	0.08601	23	7	802.12
7	0.08191	26	6	802.1G
8	0.06776	24	4	802.1F
9	0.03279	17	6	802.1H
10	0.03255	18	11	802.16
11	0.01988	12	11	802.1Q
12	0.01631	1	4	802.5c
13	0.01489	5	8	1802.3
14	0.00761	3	10	802.16a
15	0.00761	3	11	802.16c
16	0.00759	5	8	802.11b
17	0.00626	1	3	802.16.2
18	0.00562	2	12	802.15.1
19	0.00557	2	11	802.1X
20	0.00537	3	3	802.15.2
21	0.00530	3	11	802.3ae
22	0.00492	2	11	802.3af
23	0.00476	3	9	802.5r
24	0.00454	1	13	802.11d
25	0.00454	1	10	802.11g
26	0.00454	1	3	802.16_Conformance01
27	0.00453	2	9	802.5j
28	0.00440	2	8	802.11a
29	0.00438	1	7	802.3aj
30	0.00428	1	1	802a
31	0.00410	1	4	802.15.3
32	0.00399	1	10	802.5t
33	0.00375	0	11	802.11F
34	0.00375	0	11	802.11b_Cor1
35	0.00375	0	4	802.11h
36	0.00375	0	9	802.15.4
37	0.00375	0	4	802.16_Conformance02
38	0.00375	0	5	802.3ak
39	0.00375	0	13	802.5v
40	0.00375	0	9	802.5w

Tabla 2

En las gráficas siguientes los documentos se encuentran representados como nodos y las citas como enlaces dirigidos del documento referenciante al documento referenciado. Cada nodo está etiquetado con el número del estándar y su diámetro es proporcional a su puntaje de PageRank.. La distancia de un documento al centro de la gráfica está relacionada con el número de citas entrantes.

La primera gráfica representa la relación entre los documentos en base a las referencias explícitas.

La segunda gráfica representa la relación entre los documentos en base a las menciones simples.



CONCLUSIONES

Si se comparan las dos tablas de resultados, puede verse que el estándar 802 es el documento de mayor relevancia. Esto era de esperarse ya que es el punto de partida y define el contenido de los demás estándares.

Los siguientes documentos en grado de importancia son el 802.2 y el 802.3, que cubren el control de enlaces lógicos y el modo en el que opera el método de acceso múltiple con detección de colisiones, respectivamente.

En quinto lugar de ambas tablas aparece el 802.11, lo cual podría suponerse que es un reflejo del auge que han tenido las redes inalámbricas en la actualidad.

Una discrepancia entre los dos juegos de resultados son los estándares 802.1F y 802.5 que aparecen en cuarto lugar en la tabla de referencias explícitas y la tabla de menciones simples respectivamente. Sin embargo cada uno de estos documentos es menos importante en la otra tabla. El estándar 802.5 tienen mayor número de referencias y de menciones, sin embargo el 802.1F tiene mejor puntaje en la tabla de referencias debido a que los documentos que lo citan poseen una mayor importancia que aquellos que referencian al 802.5.

Los estándares que carecen ya sea de menciones o de referencias aparecen con la menor importancia objetiva.

REFERENCIAS

- [1] Lawrence Page, Sergey Brin, Rajeev Motwani, & Terry Winograd - The PageRank Citation Ranking: Bringing Order to the Web - 1999
<http://dbpubs.stanford.edu:8090/pub/1999-66>
- [2] Sergey Brin, Lawrence Page - The Anatomy of a Large-Scale Hypertextual Web Search Engine - WWW7 / Computer Networks 30(1-7): 107-117 (1998)
<http://www-db.stanford.edu/~backrub/google.html>
- [3] Get IEEE 802® Program
<http://standards.ieee.org/getieee802/portfolio.html>

- [4] Scott White, Danyel Fisher, Joshua O'Madadhain, Yan-Biao Boey – JUNG: Java Universal Network/Graph Framework.
<http://jung.sourceforge.net/>

- [5] Vlado Batagelj, Andrej Mrvar – Pajek: Program for Large Network Analysis Connections 21(1998)2, 47-57
<http://vlado.fmf.uni-lj.si/pub/networks/pajek/>

APÉNDICE

A continuación se incluye la lista completa de los títulos de los documentos evaluados.

IEEE Std 802®-2001
IEEE Standard for Local and Metropolitan Area Networks: Overview and Architecture

IEEE Std 802.1F-1993
Common Definitions and Procedures for IEEE 802 Management Information

ANSI/IEEE Std 802.1G, 1998 Edition
Part 5: Remote Media Access Control (MAC) bridging

IEEE Std 802.1H, 1997 Edition
Part 5: Media Access Control (MAC) Bridging of Ethernet V2.0 in Local Area Networks

IEEE Std 802.1Q™, 2003 Edition
Virtual Bridged Local Area Networks

IEEE Std 802.1X-2001
Port-Based Network Access Control

ANSI/IEEE Std 802.2, 1998 Edition
Specific requirements
Part 2: Logical Link Control

IEEE Std 802.3™-2002
Part 3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications

IEEE Std 802.3ae™-2002

Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications Amendment: Media Access Control (MAC) Parameters, Physical Layers, and Management Parameters for 10 Gb/s Operation

IEEE Std 802.3af™-2003

Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications Amendment: Data Terminal Equipment (DTE) Power via Media Dependent Interface (MDI)

IEEE Std 802.3aj.-2003

Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications Amendment: Maintenance 7

IEEE Std 802.3ak.-2004

Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications Amendment: Physical Layer and Management Parameters for 10Gb/s Operation, Type 10GBASE-CX4

ANSI/IEEE Std 802.5-1998E

Part 5: Token ring access method and Physical Layer specifications

IEEE Std 802.5c-1991

Supplement to Token Ring Access Method and Physical Layer Specifications Recommended Practice for Dual Ring Operation with Wrapback Reconfiguration

ANSI/IEEE Stds 802.5r and 802.5j, 1998 Edition
Part 5: Token ring access method and physical layer specifications

Amendment 1: Dedicated token ring operation and fibre optic media

IEEE Std 802.5t-2000

Amendment to Part 5: Token Ring Access Method and Physical Layer Specifications

IEEE 802.5v-2001

Part 5: Token Ring access method and Physical Layer specifications Amendment 5: Gigabit Token Ring operation

IEEE Std 802.5w-2000

Part 5: Token ring access method and Physical Layer specifications— Corrigendum 1

ANSI/IEEE Std 802.11, 1999 Edition

Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications

IEEE Std 802.11a-1999

Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications High-speed Physical Layer in the 5 GHz Band

IEEE Std 802.11b-1999

Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications: Higher-Speed Physical Layer Extension in the 2.4 GHz Band

IEEE Std 802.11b-1999/Cor 1-2001

Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 2: Higher-speed Physical Layer (PHY) extension in the 2.4 GHz band— Corrigendum 1

IEEE Std 802.11d-2001

Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 3: Specification for operation in additional regulatory domains

IEEE Std 802.11F™-2003

IEEE Trial-Use Recommended Practice for Multi-Vendor Access Point Interoperability via an Inter-Access Point Protocol Across Distribution Systems Supporting IEEE 802.11™ Operation

IEEE Std 802.11g™-2003

Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 4: Further Higher Data Rate Extension in the 2.4 GHz Band

IEEE Std 802.11h™-2003

Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 5: Spectrum and Transmit Power Management Extensions in the 5 GHz band in Europe

ANSI/IEEE Std 802.12, 1998 Edition

Part 12: Demand-Priority access method, physical layer and repeater specifications

IEEE Std 802.15.1™-2002

Part 15.1: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Wireless Personal Area Networks (WPANs)

IEEE Std 802.15.2™-2003

Part 15.2: Coexistence of Wireless Personal Area Networks with Other Wireless Devices Operating in Unlicensed Frequency Bands

IEEE Std 802.15.3™-2003

Part 15.3: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for High Rate Wireless Personal Area Networks (WPANs)

IEEE Std 802.15.4™-2003

Part 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks (LR-WPANs)

IEEE Std 802.16-2001™

Part 16: Air Interface for Fixed Broadband Wireless Access Systems

IEEE Std 802.16.2™-2004

Coexistence of Fixed Broadband Wireless Access Systems

IEEE Std 802.16™/Conformance01-2003 IEEE Standards 802.16™ Conformance IEEE Standard for Conformance to IEEE 802.16

Part 1: Protocol Implementation Conformance Statement (PICS) Proforma for 10–66 GHz WirelessMAN-SC Air Interface

IEEE Std 802.16™/Conformance02-2003

IEEE Standards 802.16™ Conformance IEEE Standard for Conformance to IEEE 802.16

Part 2: Test Suite Structure and Test Purpose for 10–66 GHz WirelessMan-SC™ Air Interface

IEEE Std 802.16a™-2003

Amendment 2: Medium Access Control Modifications and Additional Physical Layer Specifications for 2–11 GHz

IEEE Std 802.16c™-2002

Part 16: Air Interface for Fixed Broadband Wireless Access Systems— Amendment 1: Detailed System Profiles for 10–66 GHz

IEEE Std 802a.-2003

Amendment 1: Ethertypes for Prototype and Vendor-Specific Protocol Development

IEEE Std 1802.3-2001

Specific Requirements—Part 3: Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications