

Los Indicadores de Webometrics en SeDiCI según las Normas ISO 9126 y 25010

Autor: Jose Daniel Texier R.
dantexier@gmail.com

Profesor: Luis Olsina
olsinal@ing.unlpam.edu.ar
olsinal@yahoo.com

Resumen

La calidad de productos (aplicaciones) web puede ser determinada a través de las normas ISO 9126 y 25010. Estas normas definen modelos referentes a la calidad de uso, calidad del producto y calidad de los datos, las cuales han permitido al Profesor Luis Olsina desarrollar varias estrategias (INCAMI, C-INCAMI, GOCAME, SIQinU, etc.) para medir y evaluar esa calidad a partir de las normas ISO. Por ello, en este trabajo se aplicará parte de la estrategia C-INCAMI a partir de los criterios de Webometrics para determinar la calidad del repositorio institucional SeDiCI de la Universidad Nacional de la Plata. Webometrics gestiona un ranking que se fundamenta en los criterios de enlaces web internos y externos, archivos interoperables y visibilidad de los trabajos académicos y científicos. Este trabajo determinará una correlación entre SeDiCI y los 3 primeros repositorios de Latinoamérica a partir de esos criterios. La sugerencia final será establecer un estrategia para determinar la calidad de SeDiCI según requerimientos propios y poder establecer lineamientos para mejorar la calidad del producto web.

Palabras clave: calidad de aplicaciones web, C-INCAMI, SeDiCI, Webometrics.

Fecha: 10 de septiembre de 2011.

1. Introducción

El curso del postgrado “Estrategia de Evaluación para Comprender y Mejorar Calidad en Uso de Aplicaciones Web” permitió entender estrategias para medir y evaluar la calidad de los productos de software vía web y la idea principal del curso consistió en comprender la calidad percibida por los usuarios para mejorar el producto a través de métodos específicos.

Para percibir la calidad del producto, se puede analizar bajo los modelos de calidad interna, calidad externa y/o calidad de uso de acuerdo con la norma ISO 9126[1] [2] y 25010[3], donde se evidencia que existen relaciones de dependencia e influencia entre esos modelos. El presente trabajo medirá la calidad del repositorio institucional llamado Servicio de Difusión de Creación Intelectual (SeDiCI) [4] de la Universidad Nacional de La Plata (UNLP) según el enfoque de Webometrics. SeDiCI es un desarrollo web que consta básicamente de una interfaz de usuario desarrollada en PHP, de diversos componentes de software internos en Java y PHP, y de un gestor de base de datos en MySQL. La satisfacción del cliente será analizada en este trabajo a partir de algunas

características (requerimientos no funcionales) representadas en las normas ISO y adaptadas a la propuesta de Webometrics.

Webometrics, proyecto que pertenece al Laboratorio Cybermetrics de España[5], gestiona un ranking de repositorios que pone a disposición dos ediciones anuales a partir del 2008. Según Aguillo [6], el ranking inicialmente realizaba sus cálculos clasificados en un 50% de peso para el número de páginas, archivos PDF e ítems que se encuentran en Google Scholar y el otro 50% que toma en cuenta la visibilidad a partir de los enlaces externos recibidos por el repositorio. Sin embargo, cada cierto tiempo están ajustando los indicadores para la última edición [6], en julio de 2011, la ajustaron a 4 indicadores:

1. Size con un peso de 20%. Representa los enlaces web (*links*) del repositorio en los buscadores Google, Yahoo SE (Site Explorer) y Bing.
2. Visibility con un peso de 50%. Son los diferentes *links* externos que hacen referencia al repositorio, son conocidos como *inlinks*, en Yahoo SE.
3. Rich Files con un peso de 15%. Son los archivos pdf, doc, docx, ppt, pptx, ps y eps según en el buscador de Google.
4. Google Scholar con un peso de 15% representando el trabajo científico presente en el repositorio.

La clasificación o ranking Webometrics se calcula a partir del máximo valor para el *size*, *visibility*, *rich files* y *Google Scholar* de todos los repositorios. El trabajo medirá la calidad de SeDiCI a partir de esos 4 valores establecidos por Webometrics. Para observar la mejora se debe establecer un punto de comparación que permita medir la calidad.

Antes de comenzar a desarrollar el modelo, se construirá una base conceptual permitiendo verificar los conocimientos adquiridos y entender el desarrollo del trabajo.

1.1. Requerimientos Funcionales

Son los que definen el comportamiento interno del sistema, tales como cálculos, detalles técnicos, manipulación de datos y otras funcionalidades definidas en los casos de uso, es decir, son las reglas organizacionales o del negocio.

1.2 Requerimientos No Funcionales

Están directamente relacionados con la calidad del producto ya que se usan para evaluar la operación del sistema en lugar de los comportamientos específicos. Según la norma ISO 25010 y la propuesta de extensión de la norma [7], las características a tomar en cuenta y que afectan el grado de satisfacción del cliente (calidad) son:

- Adecuación funcional.
- Eficiencia.

- Compatibilidad.
- Usabilidad.
- Confiabilidad.
- Seguridad.
- Mantenibilidad.
- Portabilidad.
- Calidad de Información.

1.3 Índice

Un índice es una expresión numérica que se obtiene por medio de la combinación de variables o indicadores que permiten medir los cambios en un determinado periodo de tiempo. Un ejemplo es el Índice de Desarrollo Humano (IDH).

1.4 Indicadores

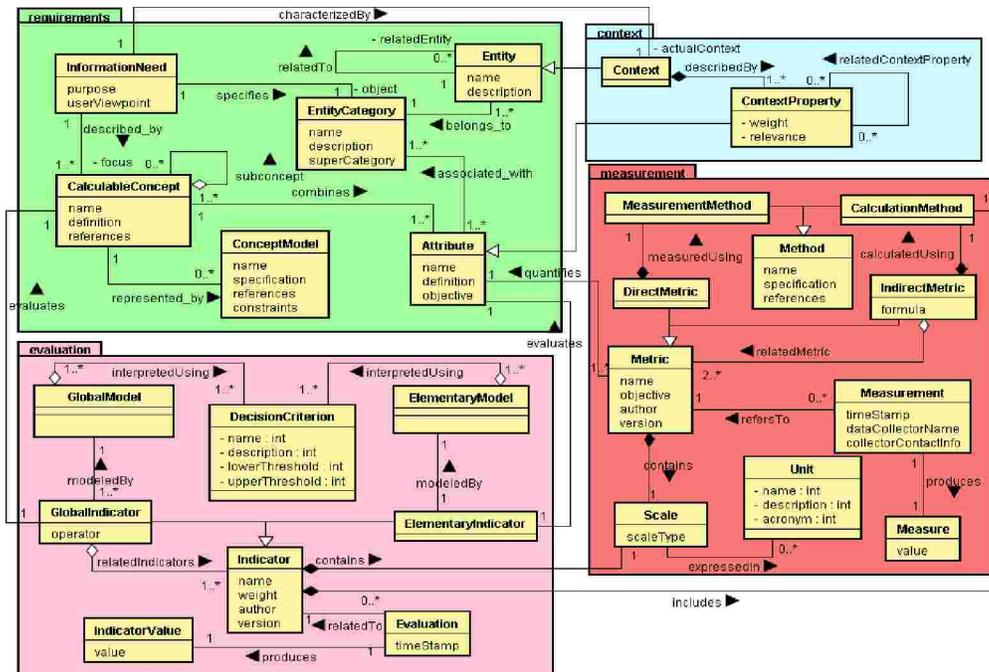
Son medidas, estadísticas o cualquier método de cálculo definido que permiten estudiar donde estamos y hacia donde nos dirigimos con respecto a determinados objetivos y metas (también conocidas como definición de necesidades de información) [8]. De acuerdo con las normas ISO 25010 se clasifican en elementales y globales.

1.4.1. Indicador elemental (IE). Es aquel que no depende de otros indicadores para evaluar o estimar un concepto calculable. Para tal fin, se debe acordar un modelo elemental, niveles de aceptabilidad, descripción y escala [8], procedente de la interpretación del valor de medida de métricas de un atributo perteneciente a una característica.

1.4.2. El indicador global (IG). Deriva de otros indicadores para evaluar o estimar un concepto calculable [8], siguiendo una estructura jerárquica y de pesos de cada uno de los atributos de las características tomadas en cuenta. El IG representa el grado de satisfacción en cumplimiento de las necesidades de información para un propósito.

1.5 Framework C-INCAMI

La calidad de un sistema es el grado que satisface las necesidades de las partes interesadas y por tanto, proporcionan valor. Estas necesidades están representadas en la serie SQuaRE de las normas ISO 25010 [3] a través de modelos que categorizan la calidad del producto dentro de una jerarquía de características y subcaracterísticas. Esta norma ISO recomienda 3 modelos de calidad: modelo de la calidad de uso, modelo de la calidad del producto y modelo de la calidad de los datos. El profesor Olsina orienta el curso y el trabajo final al modelo de calidad del producto. Por ello, según el estudio de Becker y otros[9], nos concentramos en seguir el framework C-INCAMI (ver figura 1) limitándonos a los modelos de requerimientos no funcionales, mediciones y evaluaciones.



En la figura 1, se observa que existen 4 componentes: especificación de los requerimientos no funcionales, especificación del contexto, diseño de la medición e implementación y diseño de la evaluación e implementación. Estos componentes y sus diferentes relaciones permiten sistematizar una estrategia para mejorar la calidad de uso; pero en este trabajo se limitará a seguir las siguientes actividades para medir la calidad del producto según los criterios Webometrics:

- Definición y especificación de los requerimientos no funcionales. Esta actividad se implementará a través del modelo propuesto en los estudios de Becker y otros (ver figura 2). También se construirá un árbol de requerimientos para la medición y evaluación de los atributos correspondientes a las características según la norma ISO 25010.
- Diseño e implementación de la medición de acuerdo con el modelo de la figura 3.
- Diseño e implementación de la evaluación de acuerdo con el modelo de la figura 4.

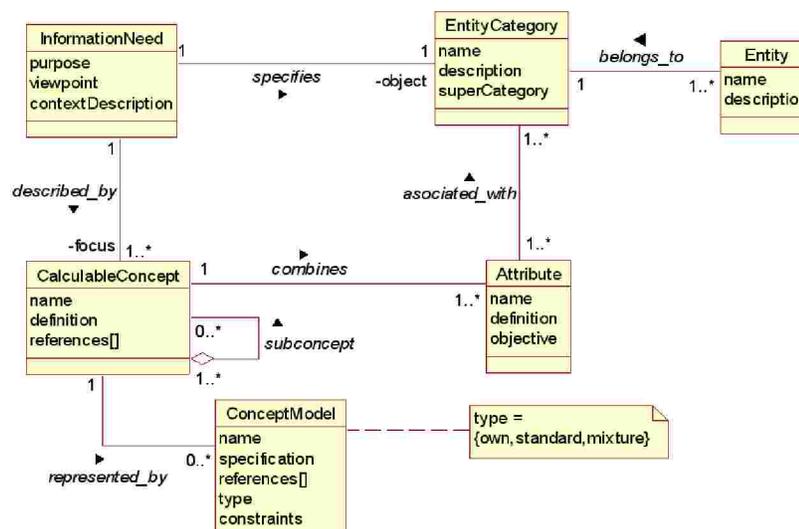
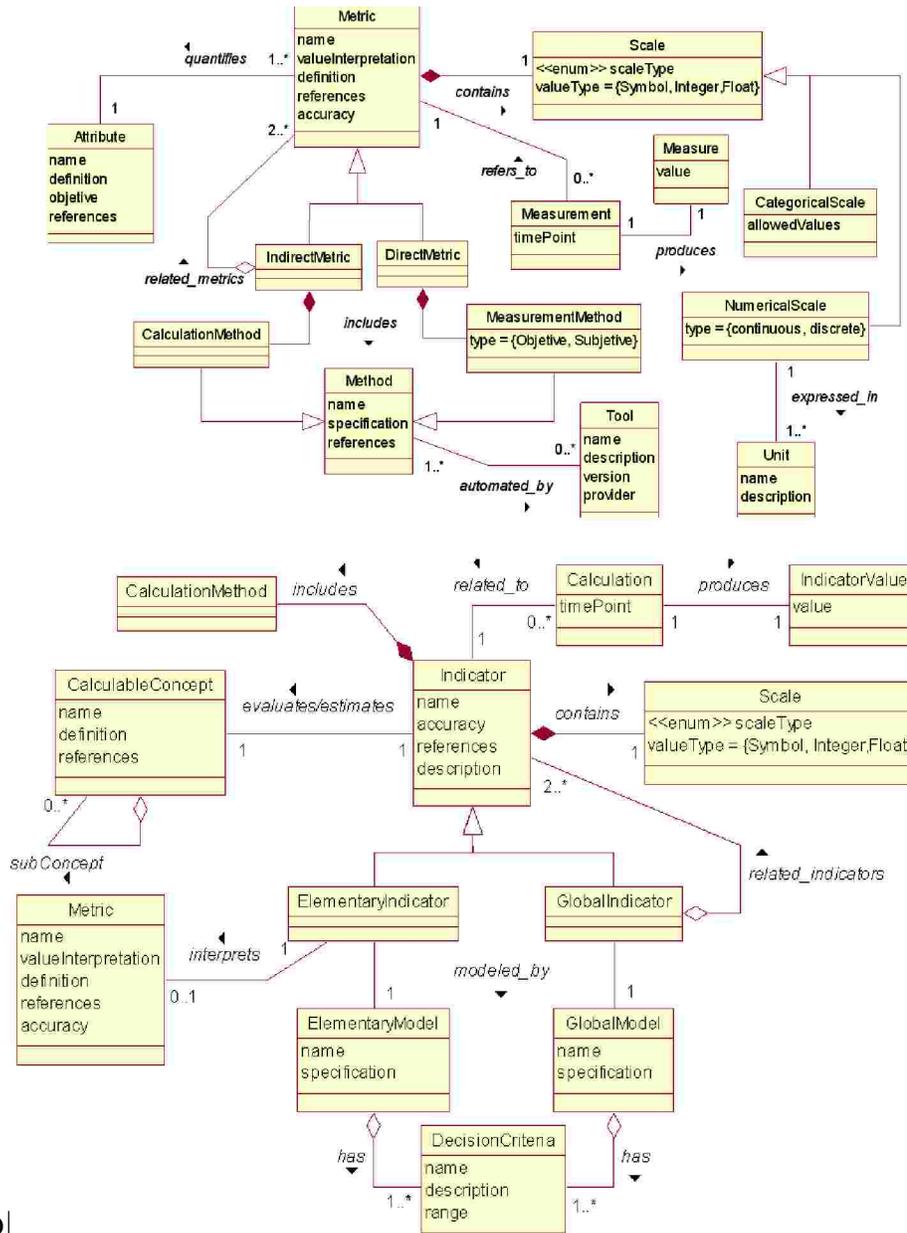


Figura 2. Modelo de definición y especificación de requerimientos.



2. Desarrollo

2.1. Definición y especificaciones de los requerimientos no funcionales.

En esta actividad se debe exponer el árbol de requerimientos de acuerdo con las características y subcaracterísticas de la norma ISO 25010 y ajustado a los 4

criterios. Por ello, a continuación se describen los criterios y se explican cómo los calcula Webometrics:

1. *Size* con un peso de 20%. Este indicador se calcula a partir del promedio de referencias en la búsqueda de enlaces que contenga el URL del sitio web del repositorio en los buscadores Google, Yahoo SE (Site Explorer) y Bing. Luego a este valor se le resta el promedio total de los *Rich Files* existentes para ese repositorio. Para Google y Bing la búsqueda se realiza con la siguiente estructura: `site:URL_Repositorio` y para Yahoo SE solamente se coloca: `URL_Repositorio`.
2. *Visibility* con un peso de 50%. Para calcularlo se toman los inlinks de Yahoo SE de: `URL_Repositorio` y se selecciona la opción "*Except from this domain to Entire Site*".
3. *Rich Files* con un peso de 15%. Se calcula a partir de la sumatoria de los archivos considerados rich files como pdf, doc, docx, ppt, pptx, ps y eps. Para obtener los valores para cada uno de los tipos se debe colocar en Google lo siguiente: `site:URL_Repositorio filetype:XYZ`.
4. *Google Scholar* con un peso de 15%. Este indicador se calcula en dos partes (un peso de 50% para cada una). En cada una de ellas hay que realizar la búsqueda en Google Scholar colocando: `site:URL_Repositorio`.
 - La primera parte (a) representa las búsquedas comprendidas entre el 2006 y el 2010 con la opción de "*at least summaries*".
 - La segunda parte (b) es sin la limitación de los años de publicación, pero sí con la opción de "*at least summaries*".

Finalmente se normaliza cada una de las partes (a y b) para después promediarlos:

- $a = (\text{datos } 2006-2010) / (\max(\text{datos del } 2006-2010))$
- $b = (\text{datos todos años}) / (\max(\text{datos de todos años}))$
- $GS = (a+b) / 2$

El ranking se calcula a partir del máximo valor para el *Size*, *Visibility*, *Rich Files* y *Google Scholar* de todos los repositorios. Cada valor del repositorio se normaliza para poder aplicárseles los pesos (porcentajes) y después sumarlos por estar normalizados. La fórmula de normalización aplicada para uno de los 4 valores es:

$N = \ln(a_k+1) / \ln(a_{\max}+1)$, donde a_k puede ser el *Size (S)*, *Visibility (V)*, *Rich Files (RF)* o *Google Scholar (GS)*, y \ln es el logaritmo natural.

El valor total para cada repositorio es calculado de acuerdo con la siguiente fórmula:

$$R_{\text{Total}_j} = N(S_j) * 0.20 + N(V_j) * 0.5 + N(RF_j) * 0.15 + N(GS_j) * 0.15$$

Donde j representa los primeros 20 repositorios de latinoamérica representados en la Tabla 1.

WR	LR	CR	REPOSITORIO	20%		50%		15%		15%		R_Total
				SIZE (TAMAÑO)		VISIBILIDAD		RICH FILE		GOOGLE SCHOLAR		
				Valor	Normal	Valor	Normal	Valor	Normal	Valor	Normal	
16	1	1	Universidade de Sao Paulo Biblioteca Digital de Teses e Dissertações	119.077	0,779	21.170	0,679	60.253	0,855	1.913	0,232	0,668
47	2	2	Universidad de los Andes Repositorio Institucional	120.204	0,780	17.484	0,666	21.828	0,776	0,706	0,116	0,623
85	3	4	Repositorio Digital de Tesis de Grado Escuela Politécnica Nacional	30.011	0,688	27.211	0,696	3.835	0,641	0,628	0,106	0,597
94	4	5	Lume Repositorio Digital Universidade Federal do Rio Grande do Sul	91.784	0,762	5.033	0,581	21.500	0,775	3,231	0,313	0,606
114	5	6	Repositorio Universidad Politécnica Salesiana	5.911	0,579	25.457	0,691	7.337	0,691	0,161	0,032	0,570
132	6	3	Repositorio de la Escuela Superior Politécnica del Ecuador	94.373	0,764	5.897	0,592	42.235	0,827	1,399	0,190	0,601
156	7	7	Repositorio de la Escuela Superior Politécnica de Chimborazo	7.831	0,598	27.326	0,696	1.084	0,543	0,110	0,023	0,563
174	8	8	Universidade de Brasilia Repository	26.626	0,680	2.342	0,529	11.601	0,727	1,303	0,181	0,536
231	9	10	Universidade Federal do Parana Biblioteca Digital de Teses e Dissertações	27.434	0,682	2.042	0,519	11.108	0,724	0,245	0,048	0,512
233	10	9	Universidad de Chile Cvbertesis	18.027	0,654	4.809	0,578	7.478	0,693	0,146	0,029	0,528
324	11	13	Maxwell Pontificia Universidade Católica do Rio de Janeiro	25.019	0,675	3.406	0,554	271	0,435	0,118	0,024	0,481
334	12	12	Biblioteca Digital Jurídica do Superior Tribunal de Justiça	10.491	0,617	1.169	0,481	10.404	0,719	0,628	0,106	0,488
343	13	15	Universidad Andina Simon Bolivar Repositorio Institucional	20.047	0,661	514	0,426	2.298	0,601	0,211	0,041	0,441
360	14	11	Universidad Nacional de Colombia Repositorio Institucional	54.599	0,727	983	0,470	11.014	0,723	0,493	0,087	0,502
414	15	14	Servicio de Difusion de la Creacion Intelectual U. Nacional de la Plata	20.094	0,661	474	0,420	3.161	0,626	0,578	0,099	0,451
415	16	19	Universidad de Chile Repositorio Academico	38.467	0,704	81	0,300	4.727	0,657	0,347	0,065	0,399
421	17	18	Biblioteca Dig. Monografias de Graduación e Especialización Univer. Brasilia	5.206	0,571	205	0,363	2.850	0,618	0,263	0,051	0,396
447	18	20	Biblioteca Digital Universidad Icesi	12.367	0,628	165	0,348	1.678	0,577	0,112	0,023	0,390
465	19	16	Biblioteca Digital Universidad Nacional de Cuyo	4.440	0,560	829	0,458	1.595	0,573	0,044	0,009	0,428
488	20	17	Universidad Colegio Mayor de Nuestra Señora del Rosario EdocUR	11.704	0,625	318	0,393	2.051	0,592	0,094	0,019	0,413
1			Social Science Research Network	763.684	0,903	1.054.762	0,945	390.007	1,000	24.216	0,699	0,908
3			CiteSeerX	3.250.028	1,000	2.359.224	1,000	1	0,054	53.856	0,868	0,838
5			Smithsonian/NASA Astrophysics Data System	2.115.590	0,971	752.154	0,922	47	0,301	100.000	1,000	0,850
MAX LATAM				120.204		27.326		60.253		3.231		
MAX				3.250.028		2.359.224		390.007		100.000		

Tabla 1. Los primeros 20 repositorios de latinoamérica. Valores tomados el 21 de agosto del 2011.

Reseña de la tabla 1:

- WR: Ranking Mundial.
- LR: Ranking de Latinoamérica.
- CR: Ranking Calculado por SeDiCI.
- Valor: Valor calculado según metodología explicada anteriormente.
- Normal: Normalización de los valores del indicador según fórmula.
- El máximo valor en Size es de 3.250.028 enlaces para el repositorio CiteSeerX.
- El máximo valor de Visibility (2.359.224) corresponde al repositorio CiteSeerX.
- El máximo valor en los Rich Files es de 390.007 archivos correspondiente al repositorio Social Science Research Network.
- El máximo valor en Google Scholar corresponde al repositorio Smithsonian/NASA.

Entendiendo los criterios y los cálculos de Webometrics, el árbol de requerimientos es:

Requerimientos	
Indicador de Calidad Global	
1. Eficiencia	
1.1. Productividad	
1.1.1. Size	
2. Usabilidad	
2.1. Reconocibilidad	
2.1.1. Visibility	

3. Portabilidad
3.1. Intercambiabilidad
3.1.1. <i>Rich Files</i>
4. Calidad de la Información
4.1. Accesabilidad
4.1.1. <i>Google Scholar</i>

Tabla 2. Árbol de Requerimientos según el enfoque de Webometrics.

A continuación se instancia el modelo de definición y especificación de requerimientos (ver figura 2) para los 4 criterios. Las definiciones generales son:

- Ente

- Nombre: Repositorio Institucional de SeDiCI.
- Descripción: aplicación web encargada de facilitar el depósito de las producciones de las unidades académicas de la UNLP, además de darles mayor visibilidad y preservarlas en el tiempo. SeDiCI se encuentra en la dirección web: `sedici.unlp.edu.ar`.

- Categoría de la entidad

- Nombre: Servicios de Repositorios Institucionales.
- Descripción: aplicación web.
- Super Categoría: producto.

- Necesidad de Información

- Propósito: comprender la calidad del repositorio SeDiCI desde la filosofía de Webometrics en comparación con los 3 mejores repositorios de latinoamérica.
- Punto de vista: usuario general sin autenticación.
- Descripción del contexto: el usuario final es una persona que utiliza los diferentes servicios ofrecidos por el repositorio, tales como:
 - leer sinopsis de los trabajos (artículos, presentaciones, tesis, libros, etc.),
 - bajar los archivos relacionados a cada trabajo,
 - realizar el envío de un trabajo,
 - leer alguna información de las páginas estáticas de SeDiCI.

- Concepto Calculable

- Nombre: calidad externa.
- Definición: es el grado en el cual el sistema satisface las necesidades cuando es usado bajo condiciones específicas [1].

Ahora, se presentan los 4 atributos (ver el árbol de requerimientos en la tabla 2) que representan los 4 criterios de Webometrics, que comparten las definiciones anteriores:

- **Atributo 1**

- Nombre: *Size*.
- Definición: representa la característica de eficiencia y la subcaracterística de productividad según la norma ISO 25010.
- Objetivo: representar el tamaño real del repositorio de acuerdo con el promedio de links de los buscadores Google, Bing y Yahoo SE.

- **Atributo 2**

- Nombre: *Visibility*.
- Definición: representa la característica usabilidad y la subcaracterística de reconocibilidad según la norma ISO 25010.

- Objetivo: medir el impacto del repositorio a través de los enlaces externos del repositorio.

- Atributo 3

- Nombre: *Rich Files*.
- Definición: representa la característica portabilidad y la subcaracterística de intercambiabilidad según la norma ISO 25010.
- Objetivo: determinar la cantidad de archivos estándar que permitan la interoperabilidad para los usuarios en general y para otros repositorios en Google.

- Atributo 4

- Nombre: *Google Scholar*.
- Definición: representa la característica de calidad de la información y la subcaracterística de accesabilidad según la norma ISO 25010.
- Objetivo: usar la información académica y científica de los repositorios a través de la recopilación realizada por Google Scholar.

2.2. Diseño e implementación de la medición.

Ahora se instancia el modelo para la medición (ver figura 3) y sus definiciones son:

Métrica 1

- Nombre: *Size_Metric*.
- Definición: esta métrica es indirecta, ya que requiere de otras métricas para realizar la medición. Con esta medida se cuantifica el **atributo Size**.
- Escala
 - Tipo del valor: integer.
 - Escala numérica: discreta.
 - Tipo de escala: ordinal.
- Unidad
 - Nombre: cantidad de enlaces.
- Método
 - Especificación: se hace una consulta en Google con la siguiente estructura: `site:URL_Repositorio`. Se destaca que para este caso, no se tomo la información de Bing ni de Yahoo SE.
 - Herramienta: Google.
 - Fórmula: `site:URL_Repositorio - site:URL_Repositorio filetype:pdf`.

Métrica 2

- Nombre: *Visibility_Metric*.
- Definición: esta métrica es directa, ya que requiere se hace una consulta de forma directa en el buscador. Con esta medida se cuantifica el **atributo Visibility**.
- Escala
 - Tipo del valor: integer.
 - Escala numérica: discreto.
 - Tipo de escala: ordinal.
- Unidad
 - Nombre: cantidad de enlaces externos.
- Método medición
 - Tipo: objetiva.
 - Especificación: se hace una consulta en Yahoo SE del `URL_Repositorio`.
 - Herramienta: Yahoo SE.

Métrica 3

- Nombre: *RichFile_Metric*
- Definición: esta métrica es directa, ya que requiere se hace una consulta de forma directa en el buscador. Con esta medida se cuantifica el **atributo Rich File**.

- Escala
 - Tipo del valor: integer.
 - Escala numérica: discreto.
 - Tipo de escala: ordinal.
- Unidad
 - Nombre: cantidad de archivos.
- Método medición
 - Tipo: objetiva.
 - Especificación: sumatoria de las consultas obtenidas a través de Google para todos los tipos de archivo siguiendo el siguiente patrón `site:URL_Repositorio filetype:XYZ`.
 - Herramienta: Google.

Métrica 4

- Nombre: GoogleScholar_Metric.
- Definición: esta métrica es indirecta, ya que requiere de una fórmula para la medición. Con esta medida se cuantifica el **atributo Google Scholar**.
- Escala
 - Tipo del valor: flotante.
 - Escala numérica: continua.
 - Tipo de escala: porcentaje.
- Unidad
 - Nombre: normalización de artículos científicos
- Método
 - Especificación: hacer la consulta en *Google Scholar* en 2 partes: una en el lapso del 2006 al 2010 y la otra sin la limitación de lapso. También, es necesario tener el máximo para cada una de las partes en los repositorios de latinoamérica.
 - Herramienta: Google Scholar.
 - Fórmula: se debe normalizar cada una de las partes y luego si obtiene el promedio entre ellas.

$$a = (\text{datos } 2006-2010) / (\max((\text{datos del } 2006-2010)))$$

$$b = (\text{datos todos años}) / (\max((\text{datos de todos años})))$$

$$\text{GoogleScholar_Metric} = (a+b) / 2$$

2.3. Diseño e implementación de la evaluación.

A continuación se instancia el modelo para la evaluación (ver figura 4) a partir del modelo de medición desarrollado anteriormente:

Indicador Elemental 1

- Nombre: porcentaje de enlaces internos.
- Acrónimo: IE_S.
- Descripción: determinar la diferencia porcentual con el valor esperado.
- Especificación: establecer la correlación entre el *Size* de SeDiCI y el promedio de los 3 mejores repositorios de latinoamérica según Webometrics.
- Cálculo: $(\text{Size_Metric_SeDiCI} * 100) / \text{Size_Metric_Promedio}$
- Peso: 20%.
- Niveles de aceptabilidad:
 - No satisfactorio: 0 - 10.
 - Marginal: 10 - 50.
 - Aceptable: 50 - 75.
 - Satisfactorio: 75 - 100.
- Métrica: Size_Metric
- Escala:

- Tipo de valor: float.
- Tipo de escala: porcentaje.
- Rangos:
 - { 0 - 10, 0 }
 - { 10 - 50, 1 }
 - { 50 - 75, 2 }
 - { 75 - 100, 3 }

Indicador Elemental 2

- Nombre: porcentaje de enlaces externos.
- Acrónimo: IE_V.
- Descripción: determinar la diferencia porcentual con el valor esperado.
- Especificación: establecer la correlación entre el *Visibility* de SeDiCI y el promedio de los 3 mejores repositorios de latinoamérica según Webometrics.
- Cálculo: $(\text{Visibility_Metric_SeDiCI} * 100) / \text{Visibility_Metric_Promedio}$
- Peso: 50%.
- Niveles de aceptabilidad:
 - No satisfactorio: 0 - 10.
 - Marginal: 10 - 50.
 - Aceptable: 50 - 75.
 - Satisfactorio: 75 - 100.
- Métrica: *Visibility_Metric*
- Escala:
 - Tipo de valor: float.
 - Tipo de escala: porcentaje.
 - Rangos:
 - { 0 - 10, 0 }
 - { 10 - 50, 1 }
 - { 50 - 75, 2 }
 - { 75 - 100, 3 }

Indicador Elemental 3

- Nombre: porcentaje de archivos interoperables.
- Acrónimo: IE_RF.
- Descripción: determinar la diferencia porcentual con el valor esperado.
- Especificación: establecer la correlación entre el *Rich File* de SeDiCI y el promedio de los 3 mejores repositorios de latinoamérica según Webometrics.
- Cálculo: $(\text{RichFile_Metric_SeDiCI} * 100) / \text{RichFile_Metric_Promedio}$
- Peso: 15%.
- Niveles de aceptabilidad:
 - No satisfactorio: 0 - 10.
 - Marginal: 10 - 50.
 - Aceptable: 50 - 75.
 - Satisfactorio: 75 - 100.
- Métrica: *RichFile_Metric*
- Escala:
 - Tipo de valor: float.
 - Tipo de escala: porcentaje.
 - Rangos:
 - { 0 - 10, 0 }
 - { 10 - 50, 1 }
 - { 50 - 75, 2 }

- { 75 - 100, 3 }

Indicador Elemental 4

- Nombre: porcentaje de producción académica y científica.
- Acrónimo: IE_GS.
- Descripción: determinar la diferencia porcentual con el valor esperado.
- Especificación: establecer la correlación entre el *Google Scholar* de SeDiCI y el promedio de los 3 mejores repositorios de latinoamérica según Webometrics.
- Cálculo: $(\text{GoogleScholar_SeDiCI} * 100) / \text{GoogleScholar_Metric_Promedio}$
- Peso: 15%.
- Niveles de aceptabilidad:
 - No satisfactorio: 0 - 10.
 - Marginal: 10 - 50.
 - Aceptable: 50 - 75.
 - Satisfactorio: 75 - 100.
- Métrica: GoogleScholar_Metric
- Escala:
 - Tipo de valor: float.
 - Tipo de escala: porcentaje.
 - Rangos:
 - { 0 - 10, 0 }
 - { 10 - 50, 1 }
 - { 50 - 75, 2 }
 - { 75 - 100, 3 }

Indicador Global

- Nombre: grado de satisfacción según Webometrics.
- Acrónimo: IG.
- Descripción: de acuerdo con los criterios establecidos por Webometrics, se calcula a partir de los IE obtenidos.
- Cálculo: $IG = IE_S * 0.20 + IE_V * 0.50 + IE_RF * 0.15 + IE_GS * 0.15$
- Niveles de aceptabilidad:
 - No satisfactorio: 0 - 10.
 - Marginal: 10 - 50.
 - Aceptable: 50 - 75.
 - Satisfactorio: 75 - 100.
- Escala:
 - Tipo de valor: float.
 - Tipo de escala: porcentaje.
 - Rangos:
 - { 0 - 10, 0 }
 - { 10 - 50, 1 }
 - { 50 - 75, 2 }
 - { 75 - 100, 3 }

3. Resultados

Luego de desarrollar parte de la estrategia para la medición y evaluación de la calidad de una aplicaciones web [8] y [9], se instanciaron los modelos para los requerimientos, medición y evaluación. A continuación se muestran los valores obtenidos por seguir la estrategia propuesta:

REPOSITORIO	20%		50%		15%		15%		R_Total
	SIZE (TAMAÑO)		VISIBILIDAD		RICH FILE		GOOGLE SCHOLAR		
	Valor	Normal	Valor	Normal	Valor	Normal	Valor	Normal	
Universidade de São Paulo Biblioteca Digital de Teses e Dissertações	119.077	0,779	21.170	0,679	60.253	0,855	1.913	0,232	0,658
Universidad de los Andes Repositorio Institucional	120.204	0,780	17.484	0,666	21.828	0,776	0.706	0,116	0,623
Repositorio Digital de Tesis de Grado Escuela Politécnica Nacional	30.011	0,688	27.211	0,696	3.835	0,641	0,628	0,106	0,597
Promedio para los tres primeros repositorios de Universidades en Latam	89.764	0,75	21.955	0,68	28.639	0,76	1,08	0,15	0,63
Servicio de Difusión de la Creación Intelectual Universidad Nacional de la Plata	20.094	0,661	474	0,420	3.161	0,626	0,578	0,099	0,451
Indicadores Elementales i (Diferencia entre SeDiCi y el Promedio)	22,39%		2,16%		11,04%		53,43%		
IG									15,23%

Tabla 3. Indicadores Elementales (Size, Visibility, Rich Files y Google Scholar). Valores calculados el 21 de agosto del 2011.

Los indicadores elementales se calculan de la siguiente manera:

$$IE_i = (\text{Valor_SeDiCi}_i * 100) / \text{Valor_Promedio}_i$$

Donde i representa el valor *Size*, *Visibility*, *Rich Files* y *Google Scholar*.

Luego el indicador global (IG) se calcula al aplicar la fórmula IG (similar a R_Total_i) a partir de los indicadores elementales de *Size*, *Visibility*, *Rich Files* y *Google Scholar* calculados anteriormente:

$$IG = IE_S * 0.20 + IE_V * 0.50 + IE_RF * 0.15 + IE_GS * 0.15$$

Requerimientos	Medida	Valor del IE	Valor de IG
Indicador de Calidad Global			15,23
1. Eficiencia		22,39	
1.1. Productividad		22,39	
1.1.1. <i>Size</i>	20.094		
2. Usabilidad		2,16	
2.1. Reconocibilidad		2,16	
2.1.1. <i>Visibility</i>	474		
3. Portabilidad		11,04	
3.1. Intercambiabilidad		11,04	
3.1.1. <i>Rich Files</i>	3.161		
4. Calidad de la Información		53,43	
4.1. Accesabilidad		53,43	
4.1.1. <i>Google Scholar</i>	0,578		

Tabla 4. Indicadores elementales y globales en el árbol de requerimientos.

4. Trabajos a Futuro

La elaboración de este trabajo permitió explorar el mundo de la calidad en aplicaciones web según las normas ISO. Por tanto, los trabajos a futuro que se pueden realizar a partir de este trabajo son:

- Definir una estrategia que permita realizar una medición y evaluación de SeDiCI a partir de los requerimientos específicos de SeDiCI, todo dentro de los modelos de las normas ISO 9126 y 25010. Algunas estrategias sugeridas serían la C-INCAMI y SIQinU [7].
- Construir una herramienta que permita el cálculo automático de los criterios de Webometrics, para realizarlo cada 6 meses y poder observar las mejoras correspondientes, así como medir y evaluar la calidad de SeDiCI.
- De acuerdo con los resultados obtenidos a través de los indicadores elementales y el indicador global, planificar trabajos de mejoras a corto, mediano y largo plazo en SeDiCI.
- Realizar un estudio que determine si son los correctos los criterios evaluados por Webometrics y en caso contrario generar una propuesta de criterios a evaluar diferente a ellos.

5. Conclusiones

Este trabajo representa el primer paso de definición de una estrategia para la mejora de la calidad del repositorio en estudio. Por ello, no debe sorprendernos los valores obtenidos en los indicadores, sino debe servir de estímulo a construir planes de mejoramiento para ofrecer un excelente servicio a los usuarios de la UNLP y de la comunidad en general. Este servicio se ofrece, con éxito, desde el 2003 y se puede observar en el anexo 1 que se tiene un promedio de 7.000 visitas mensuales, reportadas en los últimos 6 meses, valores que representan la importancia del repositorio.

Referente a Webometrics, se ha convertido en una referencia para la evaluación de los repositorios institucionales públicas y privadas, destacándose como un proyecto académico y sin interés de instituciones ni de empresas. El único punto en que se tiene discordia, por los momentos, es en la no publicación de los valores crudos de los 4 criterios, obtenidos a partir de las consultas de los diferentes buscadores, situación que consolidaría más al ranking.

6. Referencias

- [1] ISO/IEC FCD 9126-1.2, *Software Engineering - Software product quality - Part 1: Quality models*. 1998.
- [2] ISO/IEC DTR 9126-4, *Software Engineering - Software product quality - Part 4: Quality in use metrics*. 2001.
- [3] ISO/IEC 25010, *Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software qualitymodels*. 2011.

- [4] "SeDiCI - Universidad Nacional de La Plata." [Online]. Available: <http://sedici.unlp.edu.ar/>. [Accessed: 10-Sep-2011].
- [5] Jose Antonio Prieto Valverde and Jose Antonio Prieto Valverde, "internetlab.cchs.csic.es." [Online]. Available: <http://internetlab.cchs.csic.es/>. [Accessed: 10-Sep-2011].
- [6] I. F. Aguillo, J. L. Ortega, M. Fernández, and A. M. Utrilla, "Indicators for a webometric ranking of open access repositories," *Scientometrics*, vol. 82, pp. 477-486, Feb. 2010.
- [7] P. Becker, P. Lew, and L. Olsina, "Strategy to improve quality for software applications: A process view," in *Proceedings - International Conference on Software Engineering*, 2011, pp. 129-138.
- [8] L. Olsina, F. Papa, and H. Molina, "How to Measure and Evaluate Web Applications in a Consistent Way," in *Web Engineering: Modelling and Implementing Web Applications*, Springer, 2008.
- [9] P. Becker, M. Molina, and L. Olsina, "Measurement and Evaluation as quality driver," *Journal JISI*, vol. 15, no. 6, pp. 33-62, 2010.