**Identificación y trazabilidad de Atributos de Calidad en requerimientos**

Tesis entregada para el grado de

**Ingeniero de Sistemas**

en la Facultad de Ciencias Exactas



Por

**Francisco Bertoni**

**Sebastián Villanueva**

Bajo la supervisión de la

Dra. Claudia Marcos

Dr. Andrés Díaz Pace

**Universidad Nacional del Centro**

**de la Provincia de Buenos Aires**

Tandil, Argentina

Agosto 2010

Tabla de contenidos

[Capítulo I - Introducción 5](#_Toc272822449)

[Capítulo II - Conceptos relacionados 6](#_Toc272822450)

[II.1. Arquitectura de Software 6](#_Toc272822451)

[II.1.1. Definición de Arquitectura de Software 6](#_Toc272822452)

[II.1.2. Ejemplo de una arquitectura 7](#_Toc272822453)

[II.2. Atributos de calidad 10](#_Toc272822454)

[II.2.1. Principales atributos de calidad 11](#_Toc272822455)

[II.2.1.1. Disponibilidad 11](#_Toc272822456)

[II.2.1.2. Performance 11](#_Toc272822457)

[II.2.1.3. Seguridad 12](#_Toc272822458)

[II.2.1.4. Usabilidad 12](#_Toc272822459)

[II.2.1.5. Modificabilidad 13](#_Toc272822460)

[II.2.1.6. Testeabilidad 13](#_Toc272822461)

[II.2.2. Atributos de calidad en una arquitectura 13](#_Toc272822462)

[II.3. Aspectos tempranos y atributos de calidad 14](#_Toc272822463)

[II.4. Especificación de requerimientos 15](#_Toc272822464)

[II.4.1. Casos de uso 16](#_Toc272822465)

[II.4.2. Escenarios de calidad 17](#_Toc272822466)

[Capítulo III - Trabajos Relacionados 20](#_Toc272822467)

[III.1. Enfoques y técnicas existentes 20](#_Toc272822468)

[III.1.1. Métodos de elicitación 20](#_Toc272822469)

[III.1.1.1. Eliciting Efficiency Requirements with Use Cases 20](#_Toc272822470)

[III.1.1.2. Crosscutting Quality Attributes for Requirements Engineering 21](#_Toc272822471)

[III.1.2. Métodos de detección semiautomáticos 22](#_Toc272822472)

[III.1.2.1. NFR Classifier 22](#_Toc272822473)

[III.1.2.2. Using classification techniques for informal requirements in the requirements analysis-supporting system 24](#_Toc272822474)

[III.1.2.3. Identiﬁcation of non-functional requirements in textual speciﬁcations: A semi-supervised learning approach 26](#_Toc272822475)

[III.2. Conclusiones 27](#_Toc272822476)

[Capítulo IV - Identificación de atributos de calidad 28](#_Toc272822477)

[IV.1. Enfoque propuesto 28](#_Toc272822478)

[IV.2. Tokens Generation 29](#_Toc272822479)

[IV.2.1. Input Processor 30](#_Toc272822480)

[IV.2.2. Tokens Filter 32](#_Toc272822481)

[IV.2.2.1. Filtro Lower Case 34](#_Toc272822482)

[IV.2.2.2. Filtro Stop Words 34](#_Toc272822483)

[IV.2.2.3. Filtro Stemming 35](#_Toc272822484)

[IV.2.2.4. Filtro Ocurrencias 36](#_Toc272822485)

[IV.2.2.5. Filtro Pesos 36](#_Toc272822486)

[IV.3. Token Analysis 38](#_Toc272822487)

[IV.3.1. QAT Identification 38](#_Toc272822488)

[IV.3.1.1. Maps Identification 39](#_Toc272822489)

[IV.3.1.1.1. Descripción algorítmica 40](#_Toc272822490)

[IV.3.1.1.2. Descripción matemática 42](#_Toc272822491)

[IV.3.1.1.3. Ejemplo 43](#_Toc272822492)

[V.3.1.2 Maps Combination 46](#_Toc272822493)

[IV.3.2. Word-QA Association 47](#_Toc272822494)

[IV.3.2.1. Ontología 48](#_Toc272822495)

[IV.3.2.1.1. Componentes de una ontología 48](#_Toc272822496)

[IV.3.2.1.2. Ontología definida 49](#_Toc272822497)

[IV.3.2.2. Word-Instance Association. 51](#_Toc272822498)

[IV.3.2.3. Instance-QA Map Association 52](#_Toc272822499)

[Capítulo V - Evaluación de la técnica propuesta 54](#_Toc272822500)

[V.1. Casos de estudio 54](#_Toc272822501)

[V.1.1. HWS (Health Watcher System) 54](#_Toc272822502)

[V.1.1.1. Casos de Uso 55](#_Toc272822503)

[V.1.1.2. Aspectos tempranos identificados 55](#_Toc272822504)

[V.1.1.3. Análisis con la técnica propuesta 55](#_Toc272822505)

[V.1.1.4. Resumen de resultados del caso de estudio HWS 55](#_Toc272822506)

[ANEXO I 56](#_Toc272822507)

[Definición de “mapa” 56](#_Toc272822508)

[Suma de dos mapas 56](#_Toc272822509)

[División de un mapa por un número real 56](#_Toc272822510)

[Multiplicación de un mapa por un número real 56](#_Toc272822511)

[Bibliografía 57](#_Toc272822512)

# Introducción

*Acá iría el capítulo 1*

# Conceptos relacionados

En este capítulo se presentarán los conceptos que se utilizarán a lo largo del presente trabajo. Se introducirán los conceptos de “Arquitectura de software”, “Atributo de calidad”, “Escenario” y “Early Aspect” (Aspecto Temprano). Además, se introducirán algunas técnicas de especificaciones de requerimientos, centrándose en el concepto de “Caso de Usos” y su relación con los aspectos tempranos y atributos de calidad.

## Arquitectura de Software

Durante mucho tiempo los diseñadores de software han construido sistemas basados mayormente en requerimientos funcionales. De esta manera, el diseñador debía basar su diseño en estos requerimientos, para luego crear el sistema a partir del mismo. A pesar del feedback que podría haber entre el diseñador y el analista, existía implícitamente la idea que los requerimientos funcionales eran suficientes para construir un diseño satisfactorio.

En este contexto, la *Arquitectura de Software* ha emergido como un eslabón crucial en el proceso de diseño, especialmente para el desarrollo exitoso de sistemas de software complejos [1].

La visión arquitectónica de un sistema es abstracta, concentrándose en el comportamiento e interacción de componentes, prestando poca atención a detalles de implementación, algoritmos o representación de datos. La arquitectura de software constituye el primer paso hacia el diseño de un sistema que tiene un conjunto de propiedades deseadas.

### Definición de Arquitectura de Software

En la actualidad no hay ninguna definición de arquitectura de software que esté unánimemente respaldada por la totalidad de la comunidad. En general, las definiciones entremezclan el trabajo dinámico de estipulación de la arquitectura dentro del proceso de ingeniería o el diseño (su lugar en el ciclo de vida), la configuración o topología estática de sistemas de software contemplada desde un elevado nivel de abstracción y la caracterización de la disciplina que se ocupa de uno de esos dos asuntos, o de ambos. [3].

Una definición reconocida es la de Clements [4]:

*La arquitectura de software es, a grandes rasgos, una vista del sistema que incluye los componentes principales del mismo, el comportamiento de esos componentes según se la percibe desde el resto del sistema y las formas en que los componentes interactúan y se coordinan para alcanzar la misión del sistema.*

Sin embargo se debe mencionar que, algunos años después, el mismo Clements, junto con Bass, Kazman y Northrop, definen a la arquitectura de software de la siguiente forma [5]:

*La arquitectura de software de un programa o sistema computarizado es la estructura o estructuras del sistema, que involucra elementos de software y las propiedades externamente visibles de esos elementos.*

Por lo tanto, a pesar de la gran cantidad de definiciones del campo de la arquitectura de software, existe en general acuerdo de que ella se refiere a la estructura a grandes rasgos del sistema, compuesta por elementos de software, junto con sus propiedades y relaciones entre ellos. Estas cuestiones estructurales se vinculan con el diseño, pues la arquitectura de software es una forma de diseño de software que se manifiesta tempranamente en el proceso de creación de un sistema. Este diseño ocurre a un nivel más abstracto que el de los algoritmos y las estructuras de datos. En este trabajo se tomará como referencia la última definición mostrada. Por ello vale la pena profundizar en algunas implicaciones de tal definición.

La primera implicancia es que la arquitectura define *elementos de software*. Es decir que la arquitectura incorpora información acerca de cómo esos elementos se relacionan entre ellos. Esto último implica que la arquitectura omite cierta información acerca de elementos que no pertenecen a la interacción. Se puede ver a la arquitectura como una abstracción de un sistema que suprime ciertos detalles que no afectan a cómo los elementos son usados, se relacionan o interactúan con otros elementos.

La segunda implicancia es que la definición deja bien en claro que los sistemas poseen más de una estructura y que ninguna de estas puede adjudicarse ser “la arquitectura” del sistema.

La tercera implicancia es que la definición implica que todo sistema de software tiene una arquitectura, porque todo sistema puede ser descripto como un conjunto de componentes y relaciones entre estos componentes.

Finalmente, la cuarta implicancia es que el comportamiento de cada componente es parte de la arquitectura en tanto y en cuanto este comportamiento pueda ser observado o discernido desde el punto de vista de otro elemento. Ese comportamiento es lo que permite a los elementos interactuar entre ellos, lo que es parte de la arquitectura.

### Ejemplo de una arquitectura

A continuación se muestra un ejemplo de una arquitectura simple, utilizando una vista de componentes y conectores. El mismo es un ejemplo de una arquitectura de software para el problema de KWIC (Key Word in Contex). En su artículo de 1972, Parnas propuso el siguiente problema [6]:

*El KWIC [Kew Word in Context] acepta un conjunto ordenado de líneas, donde cada línea es un conjunto ordenado de palabras, y cada palabra es un conjunto ordenado de caracteres. Toda línea debe ser "desplazada circularmente'' varias veces, quitando la primera palabra y añadiéndola al final de la línea. El sistema de KWIC tiene como salida un listado, ordenado alfabéticamente, de todos los posibles desplazamientos circulares de todas las líneas.*

En la se muestra una arquitectura propuesta como solución para el problema citado anteriormente. La misma se basa en el estilo de “Pipe and Filters”[7]. En este esquema, a los elementos se los denomina filtros (filters), mientras que a los conectores se los denominan tuberías (pipes).

Cada filtro realiza una tarea específica, procesando los datos de entrada y emitiéndolos por su salida. El control es distribuido: cada filtro puede procesar los datos apenas le llegan, independientemente de lo que estén haciendo los otros filtros. Idealmente los filtros no comparten entre sí variables o estados internos. Cada filtro tiene una entrada, por donde le llegan datos, y una salida donde escribe datos producidos. La comunicación entre los filtros es restringida, sólo se realiza a través de las tuberías. Éstas conectan la salida de los filtros con las entradas de otros, transmitiendo los datos de una punta a la otra.



Figura II.1 Arquitectura propuesta para el problema de KWIC

En la arquitectura propuesta, se muestran cuatros filtros: “Input”, “Circular Shifter”, “Alphabetizer” y “Output”.

El filtro “Input” parsea la entrada y escribe las líneas parseadas en su salida. Una tubería conecta la salida de este filtro con la entrada del filtro “Circular Shifter”. De esta forma, las líneas parseadas por el filtro “Input” sirven como entrada del segundo filtro. Éste procesa las líneas de entrada produciendo desplazamientos circulares sobre esas líneas. Las nuevas líneas desplazadas son escritas en la salida del filtro, en donde nuevamente se encuentra una tubería que conecta la salida del filtro con el filtro “Alphabetizer”. De esta manera, las líneas producidas por el filtro “Circular Shifter” sirven como entrada para el filtro “Alphabetizer”. Éste ordena las líneas alfabéticamente, escribiéndolas en su salida. Análogamente, una tubería conecta este filtro al filtro “Output”, que mediante otra tubería escribe los resultados finales en la salida estándar.

## Atributos de calidad

La calidad de software se define como el grado en el cual éste posee una combinación deseada de atributos, tales como: Performance, Disponibilidad, Portabilidad, Seguridad, Usabilidad, etc. [8]. Estos atributos son requerimientos adicionales del sistema [9] que hacen referencia a características o restricciones que éste debe satisfacer, y complementan los requerimientos funcionales del mismo. Estas características o atributos se conocen con el nombre de “atributos de calidad” [10].

La arquitectura de un sistema de software es determinante para que éste posea una combinación deseada o requerida de atributos de calidad. La arquitectura es la que exhibe o inhibe los atributos de calidad de un sistema, pues una arquitectura representa las decisiones más tempranas del diseño, que impactarán de manera decisiva sobre las etapas de desarrollo posteriores. De esta manera, es posible hacer predicciones sobre cuáles atributos de calidad, y cuáles no, poseerá el sistema de software mediante la evaluación de su arquitectura.

En términos generales, Bass y otros autores [5] establecen una clasificación de los atributos de calidad en dos categorías:

* Observables en tiempo de ejecución: aquellos atributos que se determinan del comportamiento del sistema en tiempo de ejecución (por ejemplo disponibilidad, performance, seguridad, etc.)
* No observables en tiempo de ejecución: aquellos atributos que se establecen durante el desarrollo del sistema (por ejemplo modificabilidad, reusabilidad, portabilidad, escalabilidad, etc.).

La “funcionalidad” es la habilidad de un sistema para realizar el trabajo para el que fue pensado. Los atributos de calidad de un sistema son independientes de la funcionalidad del mismo [5]. Si esto no fuera así, una funcionalidad seleccionada dictaría los niveles de seguridad, modificabilidad, disponibilidad, etc. de un sistema. Esto no significa que cualquier nivel de cualquier atributo de calidad puede ser alcanzado para cualquier funcionalidad. Manipular gráficos complejos o realizar complejos cálculos matemáticos, por ejemplo, ciertamente tienen un impacto negativo sobre la performance. Sin embargo, lo que sí es posible es, bajo determinada funcionalidad, tomar ciertas decisiones que determinen un nivel relativo de calidad.

En general, es en la funcionalidad en donde se pone el foco de atención durante las distintas etapas de desarrollo, dejando a los atributos de calidad, con mucha suerte, en un segundo plano. De esta manera, la mayoría de los enfoques tienen en cuenta a los atributos de calidad separadamente de la funcionalidad, dejando su integración para las etapas finales del ciclo de desarrollo [10]. Sin embargo, los atributos de calidad deben ser considerados en todas las etapas del proceso, ya que ninguno es completamente dependiente de una sola etapa[5].

Particularmente, la arquitectura es fundamental para la realización de varios atributos de calidad de un sistema. Es la primera etapa del ciclo de desarrollo en donde las estructuras deben ser diseñadas satisfaciendo a los atributos requeridos, y en donde estos pueden ser incluso testeados. Es por esta razón, que el tener un conocimiento exacto de los atributos de calidad de un sistema en las etapas tempranas de desarrollo es fundamental para lograr un sistema exitoso [11].

En este trabajo se denominará QAR (Quality Attribute Requirement) a un requerimiento de un atributo de calidad [12]. En la bibliografía los QARs también son conocidos como “Non Functional Requirements”, o NFRs. Sin embargo, se adopta como convención la primera denominación.

### Principales atributos de calidad

A continuación se describen algunos atributos de calidad que se mencionarán luego a lo largo del trabajo. Estos son disponibilidad (availability), performance (performance), seguridad (security), usabilidad (usability), modificabilidad (modifiability) y testeabilidad (testeability).

#### Disponibilidad

La disponibilidad se relaciona con las fallas o interrupciones que pueden ocurrir en un sistema y sus consecuencias asociadas. Una falla del sistema ocurre cuando el mismo no puede brindar un servicio consistente con sus especificaciones. Dicha falla puede ser observada por usuarios del sistema - ya sean humanos u otros sistemas.

Algunas de las áreas concernientes son: identificar cómo la falla del sistema es detectada, cuán frecuentemente ésta puede ocurrir, qué sucede cuando la falla ocurre, cuánto tiempo el sistema está fuera de operaciones luego de que ésta haya ocurrido, cuántas fallas pueden ser prevenidas o que clases de notificaciones se deben producir cuando una falla ocurre.

De esta manera, la disponibilidad está fuertemente relacionada con la confiabilidad, es decir, con la habilidad del sistema de mantenerse operativo todo el tiempo.

#### Performance

Performance se relaciona con el tiempo. Eventos (por ejemplo, interrupciones, mensajes, pedidos de usuarios o simplemente el paso del tiempo) ocurren, y el sistema debe responder a ellos. Hay una gran variedad de caracterizaciones de eventos que arriban y sus respuestas, pero básicamente la performance tiene que ver con cuánto tiempo le lleva al sistema responder cuando un evento ocurre.

La performance a nivel arquitectónico se puede entender, modelar y analizar, observando: la distribución y la tasa de llegada de los servicios de pedidos, los tiempos de procesamiento, el tamaño de las colas, y la latencia. Se puede simular la performance de un sistema mediante la construcción de un modelo de encolado estocástico, basado en escenarios de carga de trabajo anticipado.

#### Seguridad

La seguridad es una medida de la capacidad del sistema para resistir el uso no autorizado sin dejar de ofrecer sus servicios a los usuarios legítimos. Un intento de violación de la seguridad se denomina “ataque” y puede adoptar diversas formas. Puede ser un intento no autorizado para acceder a datos o servicios o para modificar datos, o podría ser el intento de negar servicios a los usuarios legítimos.

Los ataques pueden ser de una amplia variedad, como el robo de dinero por transferencia electrónica, modificación de datos sensibles, robo de números de tarjetas de crédito, destrucción de los archivos en los sistemas informáticos o ataques de denegación de servicio ocasionados por gusanos o virus.

#### Usabilidad

La usabilidad concierne a la facilidad que se le brinda al usuario para completar una tarea deseada y al soporte de usuario que el sistema provee. Puede ser descompuesto en las siguientes áreas:

* Aprender las características del sistema.
* Usar el sistema eficientemente.
* Minimizar el impacto los errores.
* Adaptar el sistema a las necesidades del usuario.
* Aumentar la confianza y satisfacción del usuario.

#### Modificabilidad

La modificabilidad de un sistema es la habilidad para hacer cambios en forma rápida con un costo razonable. En general, esta capacidad deriva directamente de la arquitectura: este atributo es, casi por completo, función de la localización de cada cambio. Hacer un cambio disperso en el sistema es, frecuentemente, más costoso que hacer un cambio en un solo componente y que los restantes permanezcan sin cambios.

La regla anterior tiene excepciones. Un solo componente que es excesivamente grande y complejo, puede ser más costoso de modificar que cinco componentes más simples. Sin embargo, también es fácil imaginar un cambio global, simple y sistemático: por ejemplo, modificar el valor de una constante que aparece en todos lados. De esta manera, se considera como principio general que los cambios localizados son los mejores. [5]

Dado que la arquitectura define los componentes y las responsabilidades de cada uno, también define las circunstancias bajo las cuales cada componente tendrá que cambiar. Una arquitectura clasifica los posibles cambios en tres categorías, de acuerdo a si el cambio se realizará sobre un solo componente, sobre más de un componente, o algo más drástico como el cambio del estilo de la arquitectura.

#### Testeabilidad

La testeabilidad del software se refiere a la facilidad en la cual éste puede ser sujeto a pruebas (generalmente basadas en la ejecución) para demostrar sus defectos. Al menos el 40% del costo de desarrollo de sistemas bien diseñados es absorbido por las pruebas. Si el arquitecto de software puede reducir este coste, la recompensa es significativa.

En particular, la capacidad de prueba se refiere a la probabilidad, en el supuesto de que el software tiene al menos un defecto, de que se produzca una falla en la ejecución de la siguiente prueba.

Para que un sistema sea testeable, debe ser posible controlar el estado y la entrada de cada componente interno y después observar su salida.

### Atributos de calidad en una arquitectura

Siguiendo con el ejemplo de la arquitectura mostrada en la Figura II.1 se puede analizar en qué medida este diseño impacta en los atributos de calidad mencionados.

En primera medida, se aprecia que ese diseño posee un alto nivel de modificabilidad. Cada filtro posee una funcionalidad específica, por lo que a la hora de realizar una modificación a la funcionalidad actual es sencillo ubicar el lugar indicado. Al ser cada módulo independiente del otro, el cambio en uno no afecta al resto. También es sencillo agregar, modificar o eliminar módulos sin afectar al resto de sistema.

El hecho de que dos o más módulos puedan realizar su tarea simultáneamente tiene un impacto positivo sobre la performance, al agregar concurrencia.

Sin embargo, como atributos de calidad en los cuales la arquitectura impacta negativamente, se debe mencionar que el diseño dificulta la interacción con el usuario, resultando en un decremento del atributo de usabilidad. Como los filtros no tienen estado, se debe transmitir toda la información de un filtro a otro. Si es un volumen grande de información, se produce un alto overhead, que impacta negativamente sobre la performance.

Como se ve, ningún atributo de calidad se logra en aislamiento. Generalmente, satisfacer un atributo dado tiene un impacto negativo sobre otro atributo.

## Aspectos tempranos y atributos de calidad

Se define un “concern” como “cualquier asunto de interés en un sistema de software” [13]. El desarrollo de software tiene que tratar con un gran número de concerns. Algunos de estos están relacionados al producto (el software) que debe ser creado, como la funcionalidad y la performance. Otros concerns están relacionados con el proceso de desarrollo en sí mismo, como los tiempos y costos del desarrollo.

Mientras que, por medio de técnicas convencionales de modularización u orientación a objetos, un tipo de concern puede ser encapsulado dentro de artefactos como módulos, clases y operaciones a nivel de diseño o implementación, esto mismo no es posible para otro tipo de concern. Estos cortan transversalmente (crosscut) el diseño o implementación de unos cuantos o incluso muchos artefactos y por lo tanto son llamados crosscutting concerns [13].

En el contexto de la Ingeniería de Requerimientos (*Requirements Engineering*, RE) los concerns de interés central son los requerimientos. Un requerimiento es un tipo especial de concern. Una especificación de requerimientos bien escrita está caracterizada por el hecho de que cada declaración de requerimientos no mezcla varios requerimientos sino que representa exactamente un requerimiento, por ejemplo, solamente un concern.

Sin embargo, esto no siempre sucede en la práctica, donde ocurre que ciertos concerns se encuentran presentes en más de un requerimiento. Los requerimientos que cortan transversalmente a otros son denominados *requerimientos crosscutting* (crosscutting requirements) [13] o aspectos tempranos (early aspects). De esta manera, se define a un “aspecto temprano” como un concern que atraviesa el diseño de un sistema [14], y se manifiesta generalmente en las especificaciones de requerimientos u otros documentos preliminares producidos en el análisis de requerimientos.

La Ingeniería de Requerimientos orientada a Aspectos (Aspect-oriented requirement Engineering, AORE) se enfoca en la identificación, especificación y representación de los aspectos tempranos. Ejemplos de estos últimos incluyen a concerns que se identifican con persistencia, distribución, seguridad o restricciones de tiempo real.

Muchos aspectos tempranos se corresponden con QARs de alto nivel como seguridad, performance, portabilidad y usabilidad [15]. De hecho, el descubrimiento tanto de aspectos tempranos como de atributos de calidad tiene una significativa importancia a la hora de diseñar la arquitectura. Además de que la correcta identificación de aspectos tempranos es significativa no sólo para propósitos de diseño arquitectónico, sino para que puedan ser evaluados y modelados en la etapa de diseño y codificación del sistema, evitando tener que ser “minados” y refactorizados en etapas posteriores.

En consecuencia, dada la similitud entre aspectos tempranos y atributos de calidad, el descubrimiento de los primeros puede proporcionar pistas para identificar QARs. De esta manera, dado un aspecto temprano pude ser posible realizar un razonamiento acerca del potencial atributo de calidad al que éste hace referencia.

## Especificación de requerimientos

Tanto los requerimientos funcionales como los requerimientos de atributos de calidad son capturados en un documento denominado “Especificación de Requerimientos de Software” (Software Requirements Specification), o SRS [16].

El SRS es una completa descripción del comportamiento de un sistema a ser desarrollado. Principalmente, está conformado por un conjunto de casos de uso, que especifican la funcionalidad del sistema y como éste responde a distintos estímulos externos. Idealmente, el SRS también incluye un una especificación de los requerimientos de atributos de calidad, como así también un prototipo del sistema a ser desarrollado. Sin embargo, esto último rara vez se cumple ó, en caso contrario, los QARs no son especificados de una manera formal mediante escenarios de calidad, sino que son expresados de una manera informal utilizando lenguaje natural.

De hecho, en un estudio realizado [15], en base a 15 especificaciones de requerimientos, se manifestó una escasez de menciones a QARs. Esto puede indicar que los desarrolladores fallan a la hora de analizar la importancia de los QARs, o que falsamente asumen que éstos son sobreentendidos y aceptados por todos los stakeholders.

En general, los documentos son organizados en términos de funcionalidad, poniendo muy poco énfasis en los QARs. De esta manera, éstos quedan diseminados o esparcidos en varios documentos. Como se mencionó antes, ésto conlleva a que los atributos de calidad no sean tenidos en cuenta en las primeras etapas del desarrollo, originando diseños arquitectónicos que no cumplen con los requerimientos solicitados y productos de software que no satisfacen las necesidades del cliente.

Más aún, la identificación de los QARs no es una tarea sencilla y puede consumir gran cantidad de tiempo y esfuerzo. En general, esta identificación es realizada “ad-hoc”, analizando varios documentos que describen al sistema. Como los QARs pueden estar esparcidos entre muchos de estos documentos, se corre el peligro de malinterpretarlos fácilmente o, directamente, ignorarlos por completo.

### Casos de uso

En ingeniería del software, un caso de uso es una técnica para la captura de requisitos de un nuevo sistema o una actualización de software [17]. Cada caso de uso proporciona uno o más escenarios que indican cómo debería interactuar el sistema con el usuario ó con otro sistema para conseguir un objetivo específico. Normalmente, en los casos de usos se evita el empleo de términos técnicos, prefiriendo en su lugar un lenguaje más cercano al usuario final. En ocasiones, se utiliza a usuarios sin experiencia junto a los analistas para el desarrollo de casos de uso.

Los casos de uso se convirtieron en una de las prácticas más comunes para la captura de requisitos funcionales, especialmente con el desarrollo del paradigma de la programación orientada a objetos. También, como se mencionó anteriormente, representan la espina dorsal de los SRS y es en donde generalmente se pone el mayor énfasis, dejando en un segundo plano la especificación de los requerimientos de los atributos de calidad.

Los casos de uso evitan típicamente los términos técnicos, prefiriendo la lengua del usuario final o del experto del campo del saber al que se va a aplicar. Los casos del uso son a menudo elaborados en colaboración por los analistas de requerimientos y los clientes. Cada caso de uso se centra en describir cómo alcanzar una única meta o tarea de negocio.

Desde una perspectiva tradicional de la ingeniería de software, un caso de uso describe una característica del sistema. Para la mayoría de proyectos de software, esto significa que quizás a veces es necesario especificar diez o centenares de casos de uso para definir completamente el nuevo sistema. El grado de la formalidad de un proyecto particular del software y de la etapa del proyecto influenciará el nivel del detalle requerido en cada caso de uso. Los casos de uso pretenden ser herramientas simples para describir el comportamiento del software o de los sistemas. Un caso del uso contiene una descripción textual de todas las maneras que los actores previstos podrían trabajar con el software o el sistema. Los casos de uso no describen ninguna funcionalidad interna (oculta al exterior) del sistema, ni explican cómo se implementará. Simplemente muestran los pasos que el actor sigue para realizar una tarea.

Un caso de uso debe:

* tener un nivel apropiado del detalle;
* describir una tarea del negocio que sirva a una meta de negocio;
* ser bastante sencillo como que un desarrollador lo elabore en un único lanzamiento.

### Escenarios de calidad

Los escenarios de calidad son una forma de representar, de una manera formal, los QARs de un sistema [11]. Un escenario consiste de seis partes [5] (Figura II.2 Escenario de calidad :



Figura II.2 Escenario de calidad

1. Fuente del estímulo (Concrete Source). Alguna entidad (un humano, un sistema de computación u otro actor) que genera un estímulo.
2. Estímulo (Concrete Stimulus). Una condición que necesita ser considerada cuando arriba al sistema.
3. Ambiente (Concrete Enviroment). El estímulo ocurre bajo ciertas condiciones. El sistema puede estar sobrecargado o puede estar ejecutándose normalmente cuando el estímulo ocurre, o alguna otra condición puede ser verdadera.
4. Artefacto (Concrete Artifact). Algún artefacto es estimulado. Este puede ser el sistema completo o alguna parte de él.
5. Respuesta (Concrete Response). La respuesta es una actividad llevada a cabo luego del arribo del estímulo.
6. Medida de respuesta (Concrete Response Measure). Cuando la respuesta ocurre, ésta debería ser medida de alguna manera así el requerimiento puede ser testeado.

Se distingue entre escenarios generales y concretos de calidad. Escenarios son aquellos que son independientes del sistema y pueden, potencialmente, pertenecer a cualquier sistema. Los escenarios concretos son específicos de un sistema en particular que están bajo consideración. La caracterización de atributos de calidad puede ser hecha como una colección de escenarios generales, sin embargo, para traducir la caracterización de los atributos a requerimientos para un sistema específico, los escenarios generales deben transformarse en específicos.

Un escenario general para el atributo de calidad de disponibilidad, por ejemplo, se describe en la Figura II.3. Se muestran sus seis partes, indicando el conjunto de valores que pueden tomar. El escenario general se forma al elegir, para cada parte, uno de estos posibles valores.



Figura II.3 Conjunto de valores para escenarios generales del atributo disponibilidad

En la Figura II.4 se muestra un ejemplo de un escenario concreto de disponibilidad, derivado del escenario general mostrado en Figura II.3 Conjunto de valores para escenarios generales del atributo disponibilidad. Esto se logra haciendo a cada parte del escenario general específica del sistema en consideración. En este caso, el escenario representado es "An unanticipated external message is received by a process during normal operation. The process informs the operator of the receipt of the message and continues to operate with no downtime."



Figura II.4 Escenario concreto del atributo disponibiliad

Una colección de escenarios concretos puede ser usada como los QARs de un sistema. La información contenida en el conjunto de escenarios indica qué atributos de calidad deben estar presentes en la arquitectura de un sistema de software.

En general, las definiciones de un atributo de calidad no son operacionales. No tiene significado decir que un sistema será, por ejemplo, modificable, ya que todos los sistemas son modificables respecto a un conjunto de cambios y no modificable con respecto a otro conjunto. Los escenarios brindan una solución a esta cuestión al hacer posible su testeo a un nivel arquitectural.

# Trabajos Relacionados

En este capítulo se realizará una reseña de técnicas (o enfoques) existentes para interiorizar y comprender las ideas exploradas por los investigadores. Éste no pretende ser un estudio exhaustivo acerca del conjunto de técnicas actuales, sino más bien mencionar las que sean más representativas de las líneas seguidas.

Finalmente, se hará una comparación entre estos estudios y la motivación de la metodología seguida en la técnica propuesta en este trabajo.

## Enfoques y técnicas existentes

En general, el problema de identificación y/o clasificación de QARs es abordado por los investigadores desde dos perspectivas distintas [15]. La primera se basa en métodos de elicitación, proveyendo soporte para la identificación de QARs a medida que estos son descriptos, negociados o explicados con los stakeholders. La segunda incluye métodos de detección para la extracción semi-automática o manual de QARs a partir de una variedad de documentos existentes

.

### Métodos de elicitación

Los métodos de elicitación basan su enfoque en la provisión de cuestionarios, checklists o plantillas para indagar a los stakeholders acerca de los atributos de calidad involucrados. Los QARs son identificados a medida que los documentos de especificaciones de requerimientos están siendo creados. Este hecho brinda la posibilidad de que los QARs puedan ser negociados con los stakeholders en las etapas tempranas de desarrollo.

#### Eliciting Efficiency Requirements with Use Cases

En Dörr y otros [18] se presenta un enfoque tendiente a lograr un mínimo, completo y focalizado conjunto de medibles y trazables QARs. “Mínimo” se refiere a que solo los QARs relevantes son detectados para no restringir a la arquitectura prematuramente. “Completo” indica que todos los QARs de los stakeholders son identificados. “Focalizado” se refiere a que es claro el impacto de los QARs en la solución final, soportando ambigüedad en el sentido del estándar IEEE 830 [16]. “Medibles” significan que una métrica es dada para poder verificar que el sistema final satisface el QAR detectado. Por último, “trazable” se refiere a que el enfoque propuesto también provee la lógica efectuada por la cual el QAR en cuestión fue identificado.



Figura III.1Creación de un modelo basado en la experiencia

Para lograr estos objetivos la solución parte de un metamodelo de referencia propuesto. Este incluye los conceptos de atributo de calidad, QAR, sistema, tarea, entre otros. A partir de este modelo una organización puede instanciarlo, total o parcialmente, en un modelo concreto siguiendo sus políticas propias y experiencia de desarrollo de proyectos.

La Figura III.1 muestra lo afirmado anteriormente, en donde un modelo de referencia es “instanciado” en un modelo basado en la experiencia. A partir de este modelo la tarea “Derive Facilities” crea los checklists y templates especiales para identificar QARs. En esto caso, éstos estarán asociados con el atributo de calidad para el cual el metamodelo fue creado.

Finalmente, este conjunto de documentos creados en este proceso se utilizan para identificar QARs en la etapa de elicitación.

#### Crosscutting Quality Attributes for Requirements Engineering

En el trabajo propuesto por Moreira y otros [18] se pone el foco en los atributos de calidad que atraviesan distintas funcionalidades descriptas en los requerimientos. Es por ello que se propone un modelo para identificar y especificar QARs que atraviesen los requerimientos, incluyendo su integración semántica en la descripción funcional.

El modelo de proceso propuesto es compatible con UML y está compuesto por tres actividades principales: *identificar* (identify), *especificar* (specify) e *integrar* (integrate). La primera actividad consiste en identificar todos los requerimientos de un sistema y seleccionar, de éstos, los atributos de la calidad relevantes al dominio de aplicación y a los stakeholders. La segunda actividad se divide en dos partes principales: (1) especificar los requisitos funcionales utilizando un enfoque basado en casos de uso, (2) describir los atributos de calidad utilizando plantillas especiales e identificar aquellos que atraviesan los requerimientos funcionales (es decir, los atributos de calidad transversales). La tercera actividad propone una serie de modelos para representar la integración de los atributos de calidad transversales y requerimientos. La Figura III.2 muestra las principales actividades del modelo.



Figura III.2 Modelo de requerimientos para atributos de calidad

A su vez, el modelo seleccionado para especificar atributos de calidad está influenciado por los enfoques de Chung y otros [19] y Malan y Bredmeyer [20]. En la Figura III.3 se puede ver un fragmento del mismo.



Figura III.3 Fragmento del temaplate propuesto para atributos de calidad

### Métodos de detección semiautomáticos

Otros enfoques propuestos consisten en desarrollar herramientas semiautomáticas que “minen” QARs a partir de una variedad de documentos creados en la etapa de requerimientos. En estos casos se suelen utilizar técnicas de recuperación de la información (Information Retrieval, IR) o procesamiento natural del lenguaje (Natural Language Process, NLP) [21].

#### NFR Classifier

En [15] el problema de detección y clasificación de QARs es abordado desde una perspectiva de clasificación supervisada. Este enfoque utiliza métodos de recuperación de la información para encontrar e identificar QARs. El método supone que los diferentes tipos de QARs se caracterizan por el uso de palabras claves, relativamente distintas, a las que denominan "términos indicadores”. Cuando los términos indicadores se aprenden de un tipo específico de QAR, pueden ser utilizados para detectar requerimientos, oraciones, o frases relacionadas con ese tipo. El proceso, que se representa en la Figura III.1, incluye las tres fases principales de *Training (Entrenamiento)*, *Classification* (*clasificación)*, y *Application* (*aplicación)*.



Figura III.4 Proceso de clasificación de NFRs (QARs)

Durante la fase de entrenamiento los términos indicadores son extraídos de especificaciones de requerimientos existentes, en las que los QARs han sido manualmente categorizados por tipo. Esto lo realiza la actividad “Indicator miner”, formando un conjunto de “Indicator Terms”. Estos términos luego se utilizan durante la fase de clasificación para detectar y clasificar otros QARs, mediante la actividad “NFR Classifier”. Finalmente durante la fase de aplicación, los requerimientos clasificados se utilizan para apoyar actividades más avanzadas de la etapa de desarrollo, como la negociación o diseño arquitectónico. En esta fase, las actividades “Architectural Design”, “Aspect Oriented Design Tools”, etc. son externas a la técnica propuesta y muestran algunos ejemplos que pueden tomar como entrada los NFRs clasificados. Una fase adicional, marcada en el diagrama como I*terative Training Phase* refina los términos indicadores previamente clasificados al proveer feedback indicando la correctitud de éstos resultados.

#### Using classification techniques for informal requirements in the requirements analysis-supporting system

En el trabajo realizado por Park et al [18] se propone un sistema de soporte para el analista que permite la clasificación de requerimientos bajo distintas categorías.

Los proyectos de gran escala deben lidiar con un gran número de requerimientos, generalmente provenientes de fuentes distribuidas. Estos requerimientos en general son escritos en lenguaje natural, sin una adecuada organización o formalidad. Es por ello que se hace necesaria una correcta clasificación de los mismos, para su adecuado manejo en las etapas tempranas de desarrollo.

La herramienta propuesta en este trabajo presenta una técnica automática de clasificación de requerimientos en distintas categorías, utilizando técnicas de procesamiento del lenguaje natural (Natural Language Process, NLP). Estás categorías pueden ser, por ejemplo, costo, prioridad, tiempo de desarrollo, criticidad, atributo de calidad o cualquier otra en la que se haya entrenado a la misma.



Figura III.5 Área de aplicación del sistema propuesto

En la Figura III.2 se muestra el área de aplicación del sistema propuesto. Varios requerimientos, potencialmente provenientes de distintas fuentes, son ingresados al sistema. Éstos forman un conjunto de requerimientos desorganizados. Este grupo es justamente la entrada de la técnica propuesta que, luego de su aplicación, devuelve un conjunto de requerimientos correctamente organizados según algún criterio y listos para un análisis posterior.

El proceso seguido para la clasificación de requerimientos consta de tres etapas: *pre-procesamiento*, *construcción de un conjunto de “oraciones-centrales” para el entrenamiento de cada categoría* y *aprendizaje del clasificador*.

* Pre-procesamiento: En esta etapa se extraen las sentencias de un grupo de requerimientos preseleccionados. Por cada sentencia se sepan las palabras y se realiza un etiquetado POS [19].
* *Construcción de un conjunto de “oraciones-centrales” para el entrenamiento de cada categoría:* La técnica propuesta se basa en el hecho de que un analista puede realizar una lista de *topic-word*. Éstas son palabras más representativas de cada categoría o tópico por el que se quiere clasificar los requerimientos. Mediante esta lista de *topic words* se crea una lista de *key words*. Estas *key words* son palabras, provenientes de los requerimientos utilizados en el pre-procesamiento, altamente relacionadas con las *topic words.* Estos dos conjuntos de palabras son utilizados para obtener un conjunto de *oraciones centrales* para cada categoría. Este conjunto esta formado por las oraciones mas representativas de cada categoría y contienen una o más *key words* o *topic words.*
* *Aprendizaje del clasificador:* Las oraciones centrales obtenidas en el paso anterior sirven como base para el entrenamiento de un clasificador bayesiano.

#### Identiﬁcation of non-functional requirements in textual speciﬁcations: A semi-supervised learning approach

El enfoque propuesto en [] es un enfoque de *learning* semi-supervisado. En este enfoque, un clasificador reconoce automáticamente diferentes tipos de NFRs ocultos en un conjunto de documentos, cada uno describe un requerimiento para el sistema de forma textual, y los resultados son presentados al analista para su inspección.

En principio, en la etapa de *Learning*, algunos requerimientos categorizados son utilizados en conjunto con los no categorizados para entrenar el clasificador utilizando un algoritmo semi-supervisado.

Una vez que el clasificador fue entrenado, es utilizado para categorizar próximos unlabeled requerimientos. Luego, opcionalmente, se retroalimenta el clasificador con los resultados obtenidos.



Pre procesamiento

El método más común en el campo de recuperación de información (IR) para obtener representaciones de documentos efectivos es el modelo de vector de espacio (VSM). En este modelo, cada documento es identificado por un feature vector en el espacio en el cual cada dimensión corresponde a un disctinct término con un valor numérico o peso indicando su importancia.

Se realizan varios pasos de pre procesamiento para transformar requerimientos textuales en vectores according al modelo de vector espacio. Cuando los documento son desestructurados, se llevan a cabo los pasos más comunes de normalización (modificar letras capitales dígitos, hyphens y marcas de puntuación), limpieza de stop-words y stemming.

Semi-supervised classification with EM

En supervised learning, el algoritmo seleccionado (por ejemplo, naive Bayes, k-NN etc) usa algunos ejemplos clasificados de cada clase para generar una función de clasificación o hipótesis. El problema de este enfoque, es el gran número de ejemplos para entrenar un claificador provisto de aquiring…

El algoritmo de Maximización Expectativa (EM) es una clase de algoritmo iterativo para estimar máximum likehhood en problemas con datos incompletos. Consiste de dos etapas, la etapa de Expectation y la etapa de Maximización. Básicamente, la primer etapa, llena `



## Conclusiones

Se han visto distintas técnicas para la identificación y clasificación QARs en etapas tempranas del desarrollo. La mayoría de estos enfoques utilizan estrategias de elicitación de requerimientos o la utilización de herramientas semi-automáticas, frecuentemente basadas en técnicas de NLP o IR.

En general, ninguna de las propuestas utilizando información aportada por aspectos tempranos para la identificación de QARs.

En los métodos de elicitación es muy difícil detectar los aspectos tempranos del sistema como para que estos puedan, en esta misma etapa, ser utilizados para la creación de templates, checklists o cuestionarios.

En las herramientas semi-automáticas analizadas no se utiliza información de los aspectos tempranos para identificar QARs. Algunas, como la propuesta de Cleland-Huang, poseen una técnica que identifica indistintamanete aspectos tempanos y QARs, dada la similitud entre ambos. Pero no se utiliza información aportada por los primeros para “descubrir” a los últimos.

En la actualidad existen herramientas semi-automáticas que “minan” aspectos tempranos a partir de una variedad de documentos creados en las etapas tempranas de desarrollo. Es por ello que se ve la necesidad de investigar si, a partir de estos aspectos tempranos detectados, es posible descubrir QARs de del sistema.

# Identificación de atributos de calidad

En este capítulo se describe en detalle la técnica de identificación de atributos de calidad en especificaciones de requerimientos. Esta técnica ofrece una asistencia extra al analista en la identificación de los atributos de calidad del sistema en etapas tempranas del desarrollo.

EXPLICAR LA IDEA . CONCEPTUALMENTE LA INTUCICION QUE LUEGO IMPLEMENTAMOS.

## Enfoque propuesto

En este trabajo se ha definido un proceso por medio del cual es posible identificar el grado de relación que tiene un conjunto de early aspects (aspectos tempranos) y casos de uso con los atributos de calidad del sistema, tal como se muestra en la Figura IV.1. Este proceso utiliza como entrada una lista de casos de uso y un early aspect, y agrega a cada conjunto <early aspect, casos de uso relacionados> información del atributo de calidad dominante para el conjunto en cuestión.



Figura IV.1 Diagrama de actividades principales

Básicamente, el proceso está constituido por dos etapas. La primera etapa, denominada *Tokens Generation* (Generación de Tokens), realiza un procesamiento sobre los artefactos de entrada (use cases y early aspect) para extraer de los mismos la lista de tokens que los conforman y luego, mediante un filtrado, establecer cuáles son relevantes para la etapa de análisis.

En la segunda etapa, *Tokens Analysis* (Análisis de Tokens), se calcula el porcentaje de pertenencia de cada elemento de entrada con los diferentes atributos de calidad del sistema. Posteriormente, esa información es combinada para obtener el atributo de calidad dominante para el conjunto de entrada. De esta manera, la salida de la técnica es un *Quality Attribute Theme* (QAT). El mismo está formado por un conjunto de casos de uso relacionados con un early aspect. Adicionalmente, un QAT contiene un atributo de calidad al que esos casos de uso y el early aspect hacen referencia.

Para soportar dicho proceso se ha implementado una herramienta como una extensión - plugin - del entorno de desarrollo integrado (IDE) Eclipse [25]. La herramienta, permite al analista seleccionar como entrada el early aspect que desea analizar, junto a los casos de uso que los relacionan. A partir del resultado obtenido a través de la ejecución de la herramienta, el análista podrá contar con la información acerca de los atributos de calidad identificados, además de tener la posibilidad de visualizar la trazabilidad del proceso completo. Esto útlimo se refiere a la posibilidad que tiene el analista de ver como cada token influye en la formación del QAT.

## Tokens Generation

El conjunto de actividades correspondientes a esta etapa, tiene como finalidad llevar a cabo un análisis léxico y sintáctico sobre los textos definidos como entrada para la técnica propuesta. Esta entrada, es una lista de casos de uso y el early aspect que los relaciona. Existen herramientas que generan esta información, o sobre las cuales se puede adaptar la salida para este propósito. Una de las herramientas utilizada para el desarrollo de este enfoque fue Aspect Extractor Tool [26], la cual, a través de un análisis semántico de los casos de uso, identifica los early aspects ocultos en la especificación [27]. La salida producida por esta herramienta es la lista de early aspects detectados, los casos de uso analizados y las relaciones entre ellos.

La especificación textual de los casos de uso respeta el estándar establecido por Rational [28]. De igual modo, la especificación de los early aspects sigue el formato definido en Aspect Extractor Tool, el cuál es semejante al de los casos de uso. Con toda esta información, mediante la técnica propuesta, se confeccionan las listas de tokens que serán suministradas a la próxima etapa.



Figura IV.2 Etapa 1 – Tokens Generation

Para el análisis de la entrada, se han definido dos actividades secuenciales. En primer lugar, un procesamiento de los datos de entrada para representarlos en un formato común, denominado *Input Processor*. En segundo lugar se requiere un filtrado sobre los mismos para identificar qué información es relevante para la próxima etapa, *Tokens Filter*. En estas actividades, se logrará determinar un subconjunto del conjunto de palabras que forman la especificación de los casos de uso y la definición de los early aspects, las cuales serán relevantes para la identificación del atributo de calidad involucrado.

A continuación se describen en detalle cada una de las actividades que se desarrollan durante esta etapa.

### Input Processor

El objetivo de esta actividad es procesar la información de entrada para llevarla a una representación interna uniforme. Este procesamiento es realizado a través de la separación de la información de las entidades de entrada (especificaciones de los use cases y del early aspect) en tokens. Se define un token como una unidad básica de texto que puede ser enriquecida con diferentes pares <atributos, valor>, como por ejemplo: <peso: 1>, <ocurrencias: 4>, etc. Estos tokens representan al conjunto de palabras de entrada.



Figura IV.3 Actividad Input Processor

El resultado de este procesamiento es una lista de tokens para las palabras obtenidas de los casos de uso y otra lista de tokens para las palabras del aspecto temprano en cuestión. Al realizar la división de la entrada en tokens, se agregan los siguientes atributos para cada uno de ellos:

* Id: identificador único del token.
* Tipo: este atributo representa el tipo de documento del cual se extrajo el token (en este punto será early aspect o use case).
* Sección: se registra la sección a la cual pertenece el token. Para los casos de uso, será alguna de las secciones establecidas por el template de Rational (nombre, descripción, flujo básico, actor, etc.) mientras que para el early aspect será alguna de las partes que conforman un aspecto temprano según lo establecido en la herramienta Aspect Extractor Tool[[1]](#footnote-2).

Por ejemplo, considérese la fracción de la especificación de un caso de uso mostrada en la Tabla IV.1: Fracción de caso de uso. Tomando ésta como entrada, el resultado de aplicar el procesamiento descripto anteriormente generaría como resultado la lista tokens mostrada en la Tabla IV‑2.

|  |  |
| --- | --- |
| **Nombre** | Register User |
| **Descripción** | Record user information in the system |

Tabla IV.1: Fracción de use case

|  |  |
| --- | --- |
| **Nombre** | Distribution |
| **<verbo, objeto directo>** | <sent, server> , <sending, data> |

Tabla IV.1: Fracción de early aspect

|  |  |
| --- | --- |
| **Palabra** | **Atributos** |
| Register | <id,1>, <tipo, use case >,<sección,name> |
| User | <id,2>, <tipo, use case >,<sección, name > |
| Record | <id,3>, <tipo, use case >,<sección,description> |
| user | <id,4>, <tipo, use case >,<sección, description> |
| information | <id,5>, <tipo, use case >,<sección, description> |
| in | <id,6>, <tipo, use case >,<sección, description> |
| the | <id,7>, <tipo, use case >,<sección, description> |
| system | <id,8>, <tipo, use case >,<sección, description> |
| Distribution | <id,9>, <tipo, early aspect>,<sección, name> |
| sent | <id,10>, <tipo, early aspect>,<sección, pair> |
| server | <id,11>, <tipo, early aspect >,<sección, pair> |
| sending | <id,12>, <tipo, early aspect >,<sección, pair> |
| data | <id,13>, <tipo, early aspect >,<sección, pair> |

Tabla IV‑2: Tokens generados luego del análisis de la entrada

### Tokens Filter

Como se mencionó en la sección anterior, el filtrado de los tokens es otra de las tareas que se realiza sobre la información de entrada. En este caso, se ejecutan una serie de filtros sobre los tokens de las listas de use cases y del early aspect.

El objetivo de esta actividad es filtrar la información proveniente de la etapa anterior, *Tokens Generation*, eliminando información irrelevante y dejando sólo aquella que es útil al análisis. Además, los filtros enriquecen cada token con diferentes atributos (por ejemplo peso, ocurrencias) que serán concluyentes para la etapa de análisis.



Figura IV.4 Actividad Tokens Filter

Se considera que un filtro es una unidad de procesamiento que realiza una modificación (enriquece, refina o transforma) sobre los datos de entrada y los copia a la salida para que otro filtro trabaje sobre los mismos datos. Mediante los filtros se pueden realizar transformaciones independientes sobre el flujo de datos.

El patrón de arquitectura Pipes & Filters [7] provee una estructura para procesar flujos de datos. Cada paso de procesamiento se encapsula en un filtro y es independiente del resto. Los datos se transmiten usando los pipes entre filtros adyacentes y mediante la combinación de éstos últimos se pueden conseguir diferentes salidas.

En este caso, la entrada para los filtros es una lista de tokens, por lo que el filtro ejecutará acciones sobre los tokens de la lista (modificar la palabra, agregar atributos), para luego devolver la lista de tokens modificada.

Para esta actividad, hemos definio una serie de filtros que realizan transformaciones sobre los tokens de las listas. Los mismos fueron definidos de forma tal que puedan ser reutilizados y combinados para darle un formato común a cualquier otro texto que represente información relevante durante el proceso.

#### Filtro Lower Case

Este filtro es el encargado de pasar todos los caracteres de las palabras de los tokens a minúscula. De esta manera se reconocería que, por ejemplo, “Avión” y “avión” son la misma palabra.

Continuando con los tokens generados en la Tabla IV‑2, sólo se modifican los caracteres en mayúscula de las palabras de los tokens. En la Tabla IV‑3 se observa la lista de tokens luego de la transformación aplicada por este filtro.

|  |  |
| --- | --- |
| **Palabra** | **Atributos** |
| register | <id,1>, <tipo, use case >,<sección,name> |
| user | <id,2>, <tipo, use case >,<sección, name > |
| record | <id,3>, <tipo, use case >,<sección,description> |
| user | <id,4>, <tipo, use case >,<sección, description> |
| information | <id,5>, <tipo, use case >,<sección, description> |
| in | <id,6>, <tipo, use case >,<sección, description> |
| the | <id,7>, <tipo, use case >,<sección, description> |
| system | <id,8>, <tipo, use case >,<sección, description> |
| distribution | <id,9>, <tipo, early aspect>,<sección, name> |
| sent | <id,10>, <tipo, early aspect>,<sección, pair> |
| server | <id,11>, <tipo, early aspect >,<sección, pair> |
| sending | <id,12>, <tipo, early aspect >,<sección, pair> |
| data | <id,13>, <tipo, early aspect >,<sección, pair> |

Tabla IV‑3: Estado de los tokens luego del filtro LowerCase

#### Filtro Stop Words

Este filtro elimina las denominadas stop-words, que son palabras que, desde el punto de vista no lingüístico, no contienen información relevante. Algunas de las stop words más comunes son los artículos, las preposiciones, etc. (por ejemplo: a, an, in, the, he, she, them).

Por lo tanto, este filtro elimina de la lista de tokens aquellos en los cuales su palabra aparezca en la lista de stop-words dada. El analista/desarrollador tiene la posibilidad de incluir en esta lista las palabras, que para el dominio en el cuál este analizando, no representen información valiosa para el análisis. De igual modo, se puede modificar la lista quitando aquellas palabras que en el domino resulten relevantes y el desarrollador desee que sean procesadas.

La Tabla IV‑4 muestra el estado de los tokens de la Tabla IV‑3 luego de haberse aplicado este filtro.

|  |  |
| --- | --- |
| **Palabra** | **Atributos** |
| register | <id,1>, <tipo, use case >,<sección,name> |
| user | <id,2>, <tipo, use case >,<sección, name > |
| record | <id,3>, <tipo, use case >,<sección,description> |
| user | <id,4>, <tipo, use case >,<sección, description> |
| information | <id,5>, <tipo, use case >,<sección, description> |
| system | <id,8>, <tipo, use case >,<sección, description> |
| distribution | <id,9>, <tipo, early aspect>,<sección, name> |
| sent | <id,10>, <tipo, early aspect>,<sección, pair> |
| server | <id,11>, <tipo, early aspect >,<sección, pair> |
| sending | <id,12>, <tipo, early aspect >,<sección, pair> |
| data | <id,13>, <tipo, early aspect >,<sección, pair> |

Tabla IV‑4: Lista de tokens pos filtro Stop Words

#### Filtro Stemming

Diferentes palabras pueden tener el mismo significado desde el punto de vista semántico, pero no tener exactamente la misma secuencia de caracteres. Un ejemplo de esto pueden ser las palabras aprender, aprenden y aprendió, que, desde el enfoque en este trabajo, tienen el mismo significado. Por ello se decidió utilizar la técnica stemming [29], que es el proceso de transformar una palabra en su raíz (stem). Para el ejemplo anterior, todas esas palabras estarían identificadas por su raíz, es decir, “aprend”. Para realizar el stemming se utiliza uno de los algoritmos más populares, el algoritmo de Porter [30].

Continuando con el ejemplo observado en la Tabla IV‑4, se modifican las palabras de los tokens, llevando cada una de ellas a su raíz. En la Tabla IV‑5 se muestra el estado de los tokens luego de que se aplicó el filtro de stemming.

|  |  |
| --- | --- |
| **Palabra** | **Atributos** |
| regist | <id,1>, <tipo, use case >,<sección,name> |
| user | <id,2>, <tipo, use case >,<sección, name > |
| record | <id,3>, <tipo, use case >,<sección,description> |
| user | <id,4>, <tipo, use case >,<sección, description> |
| inform | <id,5>, <tipo, use case >,<sección, description> |
| system | <id,8>, <tipo, use case >,<sección, description> |
| distribution | <id,9>, <tipo, early aspect>,<sección, name> |
| sent | <id,10>, <tipo, early aspect>,<sección, pair> |
| server | <id,11>, <tipo, early aspect >,<sección, pair> |
| sending | <id,12>, <tipo, early aspect >,<sección, pair> |
| data | <id,13>, <tipo, early aspect >,<sección, pair> |

Tabla IV‑5: Estado de los tokens luego del filtro Stemming

#### Filtro Ocurrencias

Este filtro elimina tokens duplicados en la lista de entrada y enriquece cada token con el número de ocurrencias. Se agrega el siguiente atributo al token:

* Ocurrencias: número de ocurrencias del token en la lista. Para los tokens obtenidos de los casos de uso se considera un token duplicado en caso de que el token sea la misma palabra y además aparezca en la misma sección. Esta particularidad se debe a que luego los tokens son ponderados y debe respetarse la sección en la cual aparece a la hora de asignarle su peso. Por lo tanto, existirán tokens con la misma palabra pero diferente sección.

Igualmente, para continuar con el ejemplo, por simplicidad sólo se tendrá en cuenta la palabra del token para que se considere repetido. De esta forma, se observa en la Tabla IV‑6 el nuevo atributo que se agrega a cada token.

|  |  |
| --- | --- |
| **Palabra** | **Atributos** |
| regist | <id,1>, <tipo, use case >,<sección, name>, <ocurrencias, 1> |
| user | <id,2>, <tipo, use case >,<sección, name>, <ocurrencias, 2> |
| record | <id,3>, <tipo, use case >,<sección,description>, <ocurrencias, 1> |
| inform | <id,4>, <tipo, use case >,<sección, description>, <ocurrencias, 1> |
| system | <id,5>, <tipo, use case >,<sección, description>, <ocurrencias, 1> |
| distribution | <id,9>, <tipo, early aspect>,<sección, name> |
| sent | <id,10>, <tipo, early aspect>,<sección, pair> |
| server | <id,11>, <tipo, early aspect >,<sección, pair> |
| sending | <id,12>, <tipo, early aspect >,<sección, pair> |
| data | <id,13>, <tipo, early aspect >,<sección, pair> |

Tabla IV‑6: Lista de tokens luego de aplicar el filtro Ocurrencias

#### Filtro Pesos

Este filtro es aplicado para enriquecer los tokens de la lista de tokens de casos de uso, asignando un peso a cada token. Este peso depende de la sección donde aparece la palabra. En la técnica propuesta las palabras tienen distinta importancia según la sección del caso de uso en la que aparezcan. Por ejemplo, se asigna un peso alto a una palabra perteneciente al nombre o descripción del caso de uso, mientras que se asigna un peso más bajo a una palabra perteneciente a las precondiciones.

Hemos realizado un estudio de la importancia que tiene cada una de las componentes del template que conforman los casos de uso para nuestra problemática, y por defecto se asignan los pesos mostrados en la Tabla IV‑7 de acuerdo a la importancia de los textos y palabras que conforman las diferentes secciones. Sin embargo, el desarrollador podría modificar este peso dependiendo de las características de la definición de sus casos de uso. La herramienta soporta esta parametrización a través de un archivo de configuración donde se establecen estos pesos.

Considerando los pesos por defecto mostrados en la Tabla IV‑7, la lista de tokens final quedaría tal como se muestra en la Tabla IV‑8.

|  |  |
| --- | --- |
| **Sección** | **Peso** |
| <Nombre>, <Descripción>,<Prioridad>,<Actor> | 1 |
| <Flujo Básico> | 2 |
| <Flujo Alternativo>,<Trigger><Requerimientos especiales>,<Precondiciones>,<Postcondiciones> | 3 |

Tabla IV‑7: Ponderación de las secciones de los casos de uso

|  |  |
| --- | --- |
| **Palabra** | **Atributos** |
| regist | <id,1>, <tipo, use case >,<sección, name>, <ocurrencias, 1>,<peso, 5> |
| user | <id,2>, <tipo, use case >,<sección, name>, <ocurrencias, 2>,<peso, 5> |
| record | <id,3>, <tipo, use case >,<sección,description>, <ocurrencias, 1>,<peso, 4> |
| inform | <id,4>, <tipo, use case >,<sección, description>, <ocurrencias, 1>,<peso, 4> |
| system | <id,5>, <tipo, use case >,<sección, description>, <ocurrencias, 1>,<peso, 4> |

Tabla IV‑8: Estado final de los tokens, luego del filtro Pesos

Para el early aspect no es necesario aplicar el filtro de peso debido a que se considera que todas las palabras dentro del early aspect tienen la misma importancia. Mientras que el caso de uso está dividido en secciones, y se considera que hay secciones que deben ser ponderadas sobre otras.

## Token Analysis

En la etapa *Token Analysis* se encuentran las tareas requeridas para identificar los atributos de calidad a partir de las listas de tokens extraídos de la información de entrada. En este punto, la entrada ha sido analizada y se han construido dos listas de tokens. Además, la palabra de cada token ha sido pasada a minúscula, llevada a su raíz (stem), se ha ponderado según su importancia en el contexto y se ha dejado constancia del número de ocurrencias.



Figura IV.5. Etapa Token Analysis

La etapa *Token Analysis* (Figura IV.5. Etapa Token Analysis) se compone de dos actividades principales *QAT Identification* y *Tokens-QA Association*. La actividad *QAT Identification* se nutre de la información suministrada por la actividad *Tokens-QA Association*, que a su vez se nutre de información de una ontología (Ontology) para su ejecución.

El objetivo de esta etapa es identificar el atributo de calidad al que las listas de entrada hacen referencia, para así, finalmente, formar el Quality Attribute Theme. Adicionalmente, en el proceso de identificación del atributo de calidad dominante se podrán visualizar posibles trade-offs con otros atributos de calidad.

### QAT Identification

Esta actividad tiene como objetivo asignar un conjunto de atributos de calidad, con un porcentaje asociado a cada uno, al conjunto de la entrada suministrada por la etapa *Token Generation* (Figura IV.6. Actividad QAT Identification). Como se mencionó anteriormente, la entrada de esta actividad son dos listas de tokens: una extraída de los use cases y otra del early aspect. Cada uno de estos tokens posee diferentes atributos, como sección, peso, ocurrencias, etc. que serán determinantes en la identificación del QAT.



Figura IV.6. Actividad QAT Identification

La actividad QAT Identification está compuesta por dos sub-actividades: *Maps Identification* y *Maps Combination*. La primera calcula un mapa a partir de cada una de las listas de entrada, mientras que la segunda combina estos dos mapas en un mapa final. Ambas sub-actividades se detallan a continuación.

Para poder llevar a cabo esta actividad es necesario obtener información mediante la invocación a la actividad *Tokens-QA Association (*Figura IV.5. Etapa Token Analysis*)*. Esta última actividad relaciona un token con un conjunto de pares <atributo de calidad, porcentaje>, indicando, para los distintos atributos de calidad, el porcentaje de asociación de ese token con el mismo. A este conjunto de pares se lo ha denominado con el término “mapa” y además se han definido las operaciones de *suma de dos mapas*, *división de un mapa por un número real* y *multiplicación de un mapa por un número real. [[2]](#footnote-3)*

#### Maps Identification

Esta sub-actividad tiene como objetivo calcular un mapa de atributos de calidad y porcentajes para cada una de las listas de entrada (Use Case Map y Early Aspect Map). Estas listas son la lista de tokens extraídos de los casos de uso y la lista de tokens extraídos del early aspect.



Figura IV.7 Sub-actividad Maps Identification

A continuación, se muestra el algoritmo que sigue ésta sub-actividad para calcular el mapa a partir de los tokens extraídos de una lista. Luego se muestra el mismo razonamiento, pero expresado en forma matemática y definiendo ecuaciones formales. Finalmente, se presenta un ejemplo del funcionamiento de esta sub-actividad.

##### Descripción algorítmica

La idea subyacente de esta sub-actividad es realizar una iteración sobre cada uno de los tokens de la lista de entrada (use cases list o early aspect list). Por cada uno de los mismos se calcula el porcentaje de asociación del atributo “palabra” con los distintos atributos de calidad. Los resultados parciales de cada token se van acumulando y, cuando finalmente concluye la iteración, se calcula el promedio entre de todos los tokens.

Para un mejor entendimiento, a continuación se muestra un fragmento de pseudocódigo correspondiente a la sub-actividad “Maps Identification”.

1: Mapa <QA,porcentaje> porcentaje\_total\_map = crear mapa con todos los atributos de calidad e igualar porcentajes a 0;

2: cantidad\_de\_palabras = 0;

3: por c/ token de la lista de tokens

4: *inicio loop*

5: palabra = token.palabra

6: Mapa<QA,porcentaje> porcentaje\_palabra\_map = word-qa association(palabra);

7: n = token.ocurrencias \* token.peso;

8: porcentaje\_palabra\_map = porcentaje\_palabra\_map \* n;

9: cantidad\_de\_palabras = cantidad\_de\_palabras + n

10: porcentaje\_total\_map = porcentaje\_total\_map + porcentaje\_palabra\_map

11: *final loop*

12: porcentaje\_total\_map = porcentaje\_total / cantidad\_de\_palabras

En la primera línea, se declara un mapa de pares <atributo de calidad, porcentaje>. A este mapa se le agregan como pares todos los nombres de los atributos de calidad con los que se trabajará (performance, disponibilidad, modificabilidad, etc.), con los porcentajes de cada uno inicializados en 0.

En la línea 2, se declara la variable *cantidad\_de\_palabras* en 0. Esta variable cuenta la cantidad de palabras analizadas de los tokens de la lista. Esto último es cierto siempre y cuando el número de ocurrencias de un token y su peso sean igual a 1. En caso contrario, esta variable aumenta en la magnitud del producto entre el número de ocurrencias del token y el peso del mismo (valor de la variable n). Esta información se extrae de los atributos “peso” y “ocurrencias” pertenecientes a cada token.

En la línea 3, se inicializa una iteración por cada uno de los tokens de una lista de tokens. Por cada uno de éstos se recupera el atributo “palabra”, como se muestra en la línea 5, y se la asocia un conjunto de pares <atributo de calidad, porcentaje>, como se observa en la línea 6. Como se mencionó anteriormente, esta asociación entre la “palabra” del token y un conjunto de atributos de calidad es realizada por la tarea “Word-QA Association”, en adelante denominada *WQA*, que toma como entrada un token (aunque solamente utiliza la palabra del mismo).

En la línea 7 se declara la variable n, que es un número que representa el producto entre los atributos “peso” y “ocurrencias” que posee el token analizado en esa iteración.

En la línea 8 cada porcentaje de cada atributo de calidad del mapa al que se le asocia una palabra se multiplica por el valor de n, que, como se mencionó anteriormente, representa el número de ocurrencias de ese token y el peso del mismo. Se supone que el número de ocurrencias es mayor o igual a uno (si no fuera así el token no aparecería en la lista). El peso también se supone mayor a cero, aunque un peso igual a cero no afectaría al algoritmo, ya que en este caso se interpretaría como que el token se encontraba en una sección que no debe ser tenida en cuenta para el análisis.

Los porcentajes de cada palabra, para cada atributo de calidad, se suman y almacenan en el mapa *porcentaje\_total*.

La línea 9 indica que la cantidad de palabras analizadas se va almacenando en la variable *cantidad\_de\_palabras.* Ya que cada valor del mapa es multiplicado n veces, la cantidad de palabras no aumenta necesariamente en una unidad por cada iteración, sino que aumenta n unidades. Esto se reproduce en la línea 9.

En la línea 10 se suman los resultados parciales de porcentaje\_palabra\_map a los totales acumulados en porcentaje\_total\_map.

Una vez terminadas las iteraciones, en la línea 12 se divide cada porcentaje de cada atributo de calidad del mapa *porcentaje\_total\_map* por el valor de la variable cantidad\_de\_palabras.

##### Descripción matemática

La metodología de la sub-actividad *Maps Identification* también puede ser detallada matemáticamente. De esta manera, sea UCL[] un arreglo compuesto por tokens extraídos, por ejemplo, de los casos de uso.

Como se ha mencionado, cada token tiene varias propiedades, y en las siguientes ecuaciones se denomina como *token*.palabra al conjunto de caracteres del token que representan la palabra, *token*.ocurrencias a la propiedad que indica el número de ocurrencias de ese token y *token*.peso al valor o peso que se le asigna al mismo.

Además, sea i un número entero, UCL[i] se refiere al token en la posición i del arreglo UCL. Por último, se define como “WQA” las siglas que denominan a la tarea “Word-QA Association”. Entonces se calcula el mapa porcentaje\_total\_map de la siguiente manera:

Esta ecuación muestra lo mismo que el fragmento de pseudocódigo explicado en la subsección anterior. Por cada palabra de cada token se invoca la actividad WQA y los resultados obtenidos son multiplicados por el peso y las ocurrencias del token. A su vez el producto entre estos dos valores se va sumando para cada token, como se muestra en el dividendo de la ecuación.

Para la lista de tokens extraídos del early aspect se modifica ligeramente la ecuación anterior. En este caso, el peso del token siempre es igual a 1, ya que, como se explico anteriormente, el peso de los tokens sólo se aplica a los extraídos de los casos de uso y depende de la sección en donde se encuentren. De esta manera, para el caso de la lista de tokens extraídos del early aspect, la ecuación anterior se puede ver como:

##### Ejemplo

A continuación se verá un ejemplo de lo explicado anteriormente. En el mismo se supone una lista de sólo dos tokens extraídos de los casos de uso (Tabla IV‑9 Lista de dos tokens). El primero es un token con la palabra “fast”, tiene un peso de 1 y un número de ocurrencias también de 1. El segundo token de la lista posee la palabra “second”, con un valor de 2 para el número de ocurrencias y un valor de 3 para el peso.

|  |  |
| --- | --- |
| Palabra | Atributos |
| fast | <id,1>, <tipo, use case>,<sección,description>,<ocurrencias,1>,<peso,1> |
| second | <id,2>, <tipo, use case >,<sección, name>, <ocurrencias,2>, <peso,3> |

Tabla IV‑9 Lista de dos tokens

A fin de que se aprecie mejor el ejemplo, las palabras de los tokens aparecen sin la transformación generada por el filtro de stemming.

Supóngase que se tiene sólo los atributos de calidad Modifiability, Performance y Availability. Considerando el pseudocódigo presentado en la subsección V.3.1.1.1, la línea 1 inicializa un mapa con estos tres atributos en los pares, cada uno con valores en 0. En la línea 2 se inicializa el contador de palabras en 0. De esta manera, después de las inicializaciones, las variables quedan en el siguiente estado:

*porcentaje\_total\_map = (<Modifiability,0>; <* Availability*,0>; <Performance,0>).*

*cantidad\_de\_palabras = 0;*

En este punto comienza la iteración sobre los tokens de lista. El primer token contiene la palabra “fast”. Supóngase que la actividad *WQA* relaciona, para esa palabra, el siguiente mapa:

*WQA(“fast”) = (<Modifiability,0.25>; <* Availability*,0.0>; <Performance,0.75>)*

Este resultado significa que la actividad asigna a la palabra “fast” una relación de 0.25 o 25% con el atributo de calidad “Modifiability”, un 0% con “Availability” y un 75% con el atributo “Performance”.

La línea 7 indica que la variable n es igual al producto entre el peso del token multiplicado por el número de ocurrencias. Por lo que el valor de n es igual a 1 en la primera iteración.

Los valores de *porcentaje\_total\_map* no varían luego del producto de este por n, y la variable cantidad\_de\_palabras aumenta en 1, ya que este es el valor de n.

De esta manera, el estado de las variables luego de la primera iteración es el siguiente:

*porcentaje\_total\_map = (<Modifiability,0.25>;<* Availability*,0.0>;<Performance,0.75>).*

*cantidad\_de\_palabras = 1;*

*cantidad\_de\_palabras = 0;*

Distinto es el caso del segundo token, que posee la palabra “second”, con un peso de 3 y un número de ocurrencias igual a 2.

Supóngase que en este caso el mapa asociado para la palabra posee un valor de 0.25 para Modifiability, 0.50 para Availability y 0.25 para Performance. Estos valores deben ser multiplicados por n=6 (ya que este es el valor del producto entre el peso y el número de ocurrencias). Además, la cantidad de palabras también aumenta 6 unidades durante la segunda iteración. De esta manera, el estado de las variables luego de la segunda iteración es el siguiente:

*porcentaje\_total\_map = (<Modifiability,1.75>;<* Availability*,3.0>;<Performance,2.25>).*

*cantidad\_de\_palabras = 7;*

*cantidad\_de\_palabras = 0;*

Luego de esta iteración no hay más tokens, por los que los valores del mapa *porcentaje\_total\_map* deben ser divididos por *cantidad\_de\_palabras*, dando como resultado el siguiente mapa:

*porcentaje\_total\_map = (<Modifiability,0.25>;<* Availability*,0.43>;<Performance,0.32>)*

La interpretación de este resultado es que la lista de tokens se relaciona en un 25% con Modifiability, un 43% con Availability y un 32 % con Performance.

Como se puede notar, el peso y el número de ocurrencias del segundo token, que en su conjunto son seis veces mayores que el primero, hacen inclinar la balanza hacia Availability, que es el atributo de calidad con el que más se relaciona la palabra “second”.

#### V.3.1.2 Maps Combination

En la sub-actividad precedente se calcularon dos mapas, uno proveniente de la lista de tokens extraídos de los use cases y otro a partir de la lista de tokens extraídos del early aspect. De esta forma, la función de esta sub-actividad es combinar estos dos mapas para formar el mapa resultante que será parte del QAT.



Figura IV.8. Sub-actividad Maps Combination

La solución propuesta en este trabajo combina los resultados parciales de ambos mapas a partir de un factor k. Este factor es un número real entre 0 y 1 e indica la importancia que se le debe dar a cada uno de los mapas de entrada. A su vez, la herramienta desarrollada en este trabajo permite que el desarrollador sea capaz de ingresar este valor, según sus preferencias.

Cada valor de cada atributo de calidad del mapa de la lista de tokens del caso de uso (use cases map) se multiplica por k, mientras que cada valor de la lista del mapa de tokens del early aspect (early aspect map) se multiplica por 1-k. Luego, para cada atributo de calidad de cada mapa, se suman los valores y se almacenan en el mapa resultante.

Supóngase, por ejemplo, que el mapa A= (<Modificability,0.25>; < Availability,0.25>; <Performance,0.50>) es el relacionado con la lista de los casos de uso y el mapa B= (<Modificability,0.10>; <Availability,0.30>; <Performance,0.60>) es el relacionado con el early aspect. Eligiendo un valor de k=0.4 se tendría que multiplicar cada valor de A por 0.4 y cada valor de B por 1 - 0.4 = 0.6, para luego sumar, para cada atributo de calidad, los valores de ambas listas. Estos daría como resultado un mapa C=(<Modificability,0.16>; < Availability,0.28>; <Performance,0.56>).

Nótese que eligiendo un k=1 el mapa resultante es el mismo que el mapa que se relaciona con los casos de uso (en el caso anterior el mapa resultante, C, sería igual al A). Sin embargo, con un k=0, el mapa resultante es el mismo que el mapa identificado a partir de la lista de tokens del early aspect. Un valor de k=0,5 implica que cada mapa aporta la misma proporción al mapa resultante.

Para describir está sub-actividad de forma matemática supóngase a useCaseMap como el mapa calculado a partir de los casos de uso y earlyAspectMap como el mapa calculado a partir del early aspect. Entonces el mapa resultante, totalMap, es igual a

en donde k es un número real entre 0 y 1.

### Word-QA Association

Esta actividad relaciona una secuencia de caracteres que representa una palabra con un mapa de atributos de calidad y porcentajes. Cada porcentaje indica el grado de asociación o pertenencia que tiene esa palabra con cada atributo de calidad.



Figura .. Sub-actividad Word-QA Association

Básicamente, esta actividad consta de dos sub-actividades. La primera es averiguar a qué parte del escenario de calidad corresponde la palabra. Es decir, descubrir si corresponde a una “Fuente de estímulo”, a un “Estímulo” o a alguna otra. Este comportamiento esta agrupado en la sub-actividad *Word-Instance Association*. Una vez reconocida la instancia a la que se corresponde la palabra, se averigua cuántos y cuáles escenarios se relacionan con esa instancia. A partir de esto último, y de calcular los porcentajes en que ese grupo de escenarios se relacionan con los atributos de calidad, se obtiene un mapa de atributos de calidad y porcentajes. Este comportamiento se agrupa en la sub-actividad *Instance-QA Map Association*.

Ambas sub-actividades consultan las instancias de una ontología definida para el dominio en cuestión. Esta ontología se considera que ha sido modelada por un experto. Es por ello que a continuación ésta es detallada primero, para luego describir las sub-actividades que componen a esta actividad.

#### Ontología

La literatura contiene varias definiciones de ontologías, muchas veces contradictorias entre ellas. En este trabajo se acuerda en definir a una ontología de la siguiente manera:

*“Una ontología es un modelo de datos que describe conceptos (también llamados clases) en un dominio del discurso, propiedades de cada concepto que describen las diversas características y atributos del concepto, y restricciones sobre esas propiedades.”* [18]*.*

Un dominio es un área de temática específica o de conocimiento, tal como medicina, fabricación de herramientas, bienes inmuebles, reparación automovilística, gestión financiera, etc. [19]. Las ontologías incluyen definiciones de conceptos básicos del dominio, y las relaciones entre ellos. También codifican el conocimiento de un dominio específico y el conocimiento que extiende de los dominios.

Un dominio específico es la parte del mundo que se quiere modelar. Representa el significado aplicado a los términos usados en la construcción de la ontología. Una ontología es la descripción de los conceptos que forman parte del dominio según un punto de vista. Un sistema sólo conoce lo que puede representar en algún lenguaje, por lo tanto, todo lo que no se exprese en la ontología no será conocido para el sistema que use la ontología.

##### Componentes de una ontología

Las ontologías cuentan con los siguientes componentes que sirven para representar el conocimiento de algún dominio [20]. Los principales son:

1. Conceptos o Clases Generalmente el foco de cualquier ontología, las clases son las ideas básicas que se intentan formalizar. Por ejemplo, en una ontología de vinos, la clase “Vino” representa todos los vinos. Una clase también puede tener sub-clases que representan conceptos que son más específicos que las superclases. Por ejemplo, subclases de la clase “Vino” podrían ser “Tinto”, “Blanco”, “Rosado”, etc.
2. Instancias Las instancias son representaciones de objetos determinados de un concepto. Siguiendo el ejemplo anterior, vinos específicos son instancias de la clase “Vino”. Una botella de New Age que se vende en el supermercado es una instancia de la clase “Vino”.
3. Propiedades o Slots Los slots describen propiedades de clases o instancias. En el ejemplo anterior “New Age” es la marca del vino. Por lo tanto, esa instancia posee un slot (propiedad), que poseen todas las instancias de la clase “Vino”, que podría denominarse “marca”, y cuyo valor es la cadena de caracteres “New Age”. El valor de una propiedad puede ser una cadena de caracteres, un número, otra instancia, etc. En el caso de que fuera otra instancia se forma lo que se denomina una relación.

##### Ontología definida

En la Figura IV‑10 se muestra un diagrama con la ontología que servirá de soporte de las sub-actividades siguientes. Su construcción se basa principalmente en conceptos, definiciones, y relaciones extraídas del libro “Software Architecture in Practice” [5] y explicadas en el Capítulo II.

Esta ontología, definida por nosotros, representa el dominio de atributos y escenarios de calidad sobre el cual se está trabajando. La técnica propuesta, se basa en esta ontología para consultar información o grado de relación de las palabras de cada token con los atributos de calidad.

Esta ontología es cargada por un experto en el dominio. Su carga consiste de introducir escenarios de calidad, respetando las partes propuestas por [], relacionados con atributos de calidad.



Figura .. Ontología definida

Como se puede apreciar en la ontología, en ella aparecen los conceptos de *Quality-Attribute Scenario*, que representa a los escenarios de calidad, y *Concrete Quality-Attribute Scenario* y *General Quality-Attribute Scenario* que representan a los escenarios concretos y generales, respectivamente.

Los atributos de calidad se representan mediante el concepto *Qualitty Attribute* y se relacionan con el concepto de *Quality-Attribute Scenario* mediante la propiedad *isSpecificOf*.

A su vez, las partes de un escenario se ven reflejadas en los conceptos *Concrete Source*, *Concrete Stimulus*, *Concrete Enviroment*, *Concrete Artifact*, *Concrete Response* y *Concrete Response Measure*. Todos estos son conceptos hijos del concepto *Concrete Scenario Part* y se relacionan con el concepto de *Quality-Attribute Scenario*.

#### Word-Instance Association.

En esta sub-actividad se busca identificar la correspondencia entre una palabra y alguna instancia de los conceptos de ConcreteSource (Fuente), ConcreteStimulus (Estímulo), ConcreteEnviroment (Ambiente), ConcreteArtifact (Artefacto), ConcreteResponse (Respuesta) y ConcreteResponseMeasure (Medida de Respuesta) de la ontología. Básicamente, a partir de una palabra dada se intenta responder a la pregunta, ¿la palabra es una “Fuente”, un “Estímulo”, un “Ambiente”, un “Artefacto”, una “Respuesta” o una “Medida de Respuesta”?

Esta tarea no es trivial. El caso ideal es que exista una sola instancia, de alguna de esas seis, que se corresponda con la palabra.



Figura IV.11 Una única instancia se corresponde con la palabra "user"

La Figura IV‑11 muestra lo afirmado anteriormente. En ella se observa en el recuadro superior una clase (“ConcreteSource”), mientras que en el recuadro inferior se representa la instancia (“Instancia\_1”). A su vez la instancia posee una propiedad (de tipo “cadena de caracteres”) que se denomina “concreteScenarioPartDescription” y cuyo valor, en este caso, es “user”. Uno o más escenario se pueden relacionar con esta instancia mediante la propiedad “concreteScenariohasSource”.

Se pretende identificar la palabra “user”. En este caso hay una sola instancia en la ontología, cuyo identificador único de instancia es “Instancia\_1”, de tipo ConcreteSource, que mediante la propiedad “concreteScenarioPartDescription” posee el valor “user”, por lo que se retorna esta instancia.

Distinto es el caso en el que la misma palabra pueda representar, simultáneamente, distintas partes de un escenario (Figura IV‑12).



Figura IV.12 Dos instancias, de diferente tipo, se corresponden con la palabra "system"

En este caso la palabra “system” podría ser un ConcreteSource o un ConcreteArtifact. Lo que lleva a la cuestión de decidir entre ambas opciones. Para ello se contabiliza con cuántos escenarios se relacionan cada una de las instancias. Será elegida como significado semántico de la palabra la instancia que se relacione con un mayor número de escenarios. En este último caso, si la “instancia\_1” se relaciona con 3 escenarios y la “instancia\_2” con 6, se retorna la “instancia\_2”, tomándose como que la palabra “system” es un ConcreteArtifact. En el caso de que haya un empate se elige, indistintamente, cualquiera de las instancias.

En este punto cabe aclarar que no existen dos instancias que se relacionen con la misma palabra (mediante la propiedad “concreteScenraioPartDescription”) y que sea del mismo tipo (ConcreteScource, ConcreteArtifact, etc.) Es decir, por ejemplo, no puede haber dos instancias en las que ambas sean estímulos y además se relacionen con la palabra “developer”. Si esto ocurriera, se pasan todas las relaciones de una instancia a la otra, y se elimina la primera. Este es un proceso que se realiza al crear la ontología. Es por ello que en la descripción anterior, con devolver la instancia ya se asegura que es única para ese parte de escenario.

#### Instance-QA Map Association

Una vez que se ha identificado la instancia de alguna parte de un escenario que se corresponde con una palabra, se dispone a obtener los porcentajes de los distintos atributos de calidad que se relacionan con esa instancia. Los atributos de calidad, juntos con los porcentajes asociados, se agrupan en un mapa.

Para encontrar estos porcentajes, primero se recuperan todos los escenarios que se relacionan con la instancia encontrada en el paso anterior. Cada uno de estos escenarios se relaciona con un atributo de calidad mediante la propiedad “isSpecificOf”. De esta manera, para cada atributo de calidad, se devuelven los porcentajes en que estos escenarios se relacionan con cada uno de ellos.

Por ejemplo, supóngase que como entrada se obtuvo el término "latency". Después del paso anterior, se encuentra que esa palabra se corresponde con la “Instancia\_48” que es de tipo ResponseMeasure. A su vez 40 escenarios se relacionan con la “Instancia\_48” mediante la propiedad “scenarioHasResponseMeasure”. De esos 40 escenarios, 30 de ellos se relacionan con el atributo de calidad "Performance", 8 con " Availability " y 2 con "Modificability". De esta manera, es posible concluir que ese concepto, “latency”, se relaciona en un 75% de certeza con "Performance", 20% con " Availability " y 5% con "Modificability". En este caso, esta sub-actividad relacionaría a la palabra “latency” con <Performance 0.75, Availability 0.2, Modificability 0.05>.

# Evaluación de la técnica propuesta

En este capítulo se realiza un análisis de la herramienta desarrollada (IMAGEN DE LA HERRAMIENTA) siguiendo la técnica propuesta en este trabajo para la identificación de atributos de calidad. Esta evaluación será llevada a cabo con (X) casos de estudios reales, los cuales serán explicados brevemente y se proveerá su especificación, como así también los aspectos tempranos detectados.

## Casos de estudio

Se estudiará individualmente cada uno de los casos de estudio que fueron seleccionados para la evaluación. Por cada uno de ellos se explicará brevemente, en primera instancia, el sistema en cuestión. En segunda instancia se indicarán los aspectos tempranos detectados utilizando la herramienta Aspect Extractor Tool. También se mostrará, por cada aspecto temprano, la lista de casos de uso que éste atraviesa. En tercera instancia se mostrarán los resultados obtenidos mediante la ejecución de la herramienta. Finalmente, y en última instancia, se realizará un resumen y análisis de los resultados obtenidos en cada caso.

TODO: EXPLICAR QUE PRIMERO SE ANALIZA CADA CASO SOLO CON LOS CASOS DE USO Y LUEGO SE ANALIZA CADA ASPECTO TEMPRANO

### HWS (Health Watcher System)

The purpose of the system is to collect then manage public health related complaints and notifications. The system is also used to notify people about important information regarding the Health System.

The Health Watcher system must also exchange information with the SSVS system (Sanitary Surveillance System). Initially, this exchange will involve the querying of sanitary licenses. Subsequently, when the SSVS has the Complaint Control module deployed, Sanitary Surveillance complaints will be exchanged between the two systems.

Broadening and related systems

With the deployment of HEALTH-WATCHER system, the Public Health System will considerably improve:

• The complaint control (registering and notifications).

•Quality of service regarding the dissemination of information; for example: vaccination campaigns, disease prevention, health guides, obtaining birth/death certificates and application details for a sanitary license.

The system will be managed by DIEVS and will exchange information with the Sanitary Surveillance system.

A citizen can access the system through the internet or dialing 1520, and make their complaint or ask information about the health services. In the event of a complaint being made, it will be registered on the system and addressed by a specific department. This department will be able to handle the complaint in an appropriate manner and return a response when complaint has been dealt with. This response will be registered on the system and available to be queried.

The system will be for public use in kiosks at several strategic points, on which the citizen will be able to register complaints and request information.

#### Casos de Uso

#### Aspectos tempranos identificados

#### Análisis con la técnica propuesta

* Parámetros con los q corrimos la herramienta (valor de las secciones de los casos de uso y valor de k) MOSTRAR IMÁGENES DEL INGRESO DE PARAMETROS
* Mostrar tabla sólo de los casos de uso y resultados obtenidos. Explicar por qué hicimos esto y que estamos buscando.
* Tabla con cada early aspect con los resultados, y las capturas que muestren

#### Resumen de resultados del caso de estudio HWS

Analizar los resultados

# ANEXO I

En este anexo se describe el concepto de mapa y sus operaciones de una manera formal y detallada.

### Definición de “mapa”

Se define un “mapa” como un conjunto de tuplas <K,V> , donde a K se lo denomina “clave” y a V se lo denomina “valor”. Un mapa no puede tener dos tuplas con claves idénticas. Es decir, dado un mapa M no existe ningún par de tuplas <K1, V>, <K2,V2> tal que K1=K2.

En los mapas definidos en este trabajo se utiliza como “claves” los nombres de los atributos de calidad y como “valores” números reales entre 0 y 1.

### Suma de dos mapas

Sean M1 y M2 dos mapas con igual número de tuplas, donde toda clave K que pertenece a alguna tupla de M1, también pertenece a alguna tupla de M2. Sea además cualquier valor de V perteneciente a M1 y M2 un número real. Definimos la suma M1+M2, como un mapa M3 en donde cada tupla que éste posee es el resultado de la suma de los valores de las tuplas de igual clave de M1 y M2.

Por ejemplo, si se tiene el mapa y el mapa se define siendo K1 y K2 claves; a, b, c y d números reales

### División de un mapa por un número real

Sea M un mapa compuesto por un conjunto de tuplas <K,V> donde K es la clave y V son números reales. Sea x un número real distinto de cero. La división de M por x da como resultado el mismo mapa, con todos los valores V divididos por ese número.

Por ejemplo, si entonces , siendo a, b y x números reales; M un mapa; K1,K2 claves del mapa.

### Multiplicación de un mapa por un número real

Sea M un mapa compuesto por un conjunto de tuplas <K,V> donde K es la clave y V son números reales. Sea x un número real. La multiplicación de M por x da como resultado el mismo mapa, con todos los valores V multiplicados por ese número.

Por ejemplo, si , entonces , siendo a,b y x números reales; M un mapa; K1,K2 claves del mapa.

# Bibliografía

1. **Clements, et al.** A Practical Method for Documenting Software Architectures. *Carnegie Mellon University.* Pittsburgh, Pennsylvania, USA : s.n., 2002.

2. **Reynoso, Carlos.** *Introducción a la Arquitectura de Software Version 1.0.* s.l. : Universidad de Buenos Aires, 2004.

3. *Survey of Architecture Description Languages.* **Clements, Paul.** 1996. Proceedings of the International Workshop on Software Specification and Design.

4. **Bass, Clements and Kazman.** *Software Architecture in practice.* s.l. : Addison-Wesley, 1998.

5. **Parnas, D. L.** On the criteria to be used in decomposing systems into modules. *Communications of the ACM.* Diciembre 1972. Vol. 15, pp. 1053-1058.

6. **Buschmann, et al.** *Pattern-Oriented Software Architecture: A System Of Patterns.* West Sus- : John Wiley & Sons, 1996.

7. *IEEE Standard 1061-1992. Standard for a Software Quality Metrics.* IEEE. 1992. Standard.

8. **Kazman, Clements and Kein.** *Evaluating Software Architectures. Methods and case studies.* s.l. : Addison Wesley, 2001.

9. **Barbacci, et al.** *Quality Attributes.* 1995. Technical Report. CMU/SEI-95-TR-021 ESC-TR-95-021.

10. **Bass, Clements and Kazman.** *Software Architecture in Practice, Second Edition.* s.l. : Addison-Wesley, 2003.

11. **Berenbach, Paulish and Kazmei.** *Software & Systems Requirements Engineering: in practice.* s.l. : McGraw Hill Professional, 2009.

12. *Identifying Crosscutting Concerns in Requirements Specifications.* **Rosenhainer, L.** Vancouver : s.n., 2004. Workshop on Early Aspects: Aspect-Oriented Requirements Engineering and Architecture Design , held in Function with OOPSLA.

13. **Sampaio, A.** EA-Miner: a Tool for Automating Aspect-Oriented Requirements Identification. *Proceedings of the 20th IEEE/ACM international Conference on Automated software engineering.* 2005.

14. **Baniassan, et al.** Discovering early aspects. *IEEE Software.* 2001. Vol. 23, 1, pp. 61-70.

15. **Cleland-Huang, Jane, et al.** Automated classiﬁcation of non-functional requirements. Londres, UK : Springer-Verlag, Abril 2007. Vol. 12, 2, pp. 103-120.

16. IEEE 830: IEEE Recommended Practice for Software Requirements Specifications. 1998.

17. [Online] http://es.wikipedia.org/wiki/Caso\_de\_uso.

18. **Dörr, J., et al.** Eliciting Efficiency Requirements with Use Cases. *9th International Workshop on Requirements Engineering. Foundation for Software Quality, REFSQ '03. Pre-Proceedings.* 2003. Vol. 3, pp. 23-32.

19. **Moreira, Ana, Araujo, Joao and Brito, Isabel.** Crosscutting quality attributes for requirements engineering. *Proceedings of the 14th international conference on Software engineering and knowledge engineering.* 2002. Vol. 27, pp. 167-174 .

20. **Chung, L., Mylopoulos, J. and Nixon, B.** Representing and using nonfunctional requirements: a process-oriented approach. Junio 1992. Vol. 18, pp. 483 - 497. 0098-5589.

21. **Malan, M. and Bredemeyer, D.** Definining Non-Functional Requirements. *http://www.bredemeyer.com/papers.htm.*

22. **Casamayor, Agustin, Godoy, Daniela and Campo, Marcelo.** Identification of non-functional requirements in textual specifications:A semi-supervised learning approach. *Information and Software Technology.* 2010. Vol. 52, pp. 436-445.

23. **Park, Sooyong, et al.** Using classification techniques for informal requirements. *Information and Software Technolog.* 2007. Vol. 49, pp. 1128-1140.

24. **Brill, E.** Transformation-based error-driven learning and natural language processing: a case study in part-of-speech tagging. *Computational Linguistics.* Diciembre 1995. Vol. 7, pp. 543 -565.

25. Fundacion Eclipse. [Online] http://www.eclipse.org/.

26. *Identificacion Temprana de Aspectos.* **Haak, B., Pryor, A. and Marcos, C.** 2005, Revista SCC (Workshop in SE), Vol. 6.

27. *Early Aspect Identification from Use Cases using NLP and WSD Techniques.* **Rago, et al.** 2009. EA '09 Proceedings of the 15th workshop on Early aspects.

28. *UML Semantics version 1.1.* Rational Software Corporation. 1997.

29. [Online] http://es.wikipedia.org/wiki/Stemming.

30. Official home page for distribution of the Porter Stemming Algorithm. [Online] http://tartarus.org/~martin/PorterStemmer/.

31. **Noy, Natalya and McGuinness, Deborah.** Ontology Development 101: A Guide to Creating Your First Ontology Escenario. *Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880.* Marzo 2001.

32. OWL Web Ontology Language - Use Cases and Requirements. *W3C Recommendation .* [Online] 2004. http://www.w3.org/TR/webont-req/.

33. **Gruber, T. R.** A translation approach to portable ontology specfications. Junio 1993. Vol. 5, 2, pp. 199-220.

34. **Shaw and Garlan.** *An introduction to software architecture.* CMU Software Engineering Institute Technical Report. 1994. CMU/SEI-94-TR-21, ESC-TR-94-21..

1. La herramienta Aspect Extractor Tool representa un early aspect mediante un nombre y un conjunto de pares formados por un objeto directo y un verbo, o simplemente un verbo. [↑](#footnote-ref-2)
2. En el Anexo II se encuentra una definición formal del concepto “mapa” y de cada una de las operaciones mencionadas anteriormente, que se utilizarán a lo largo de este trabajo. [↑](#footnote-ref-3)