



Engineering Requirements for crowds

Ingeniería de Requisitos para multitudes

Rogero Silva, Anaima Dasilva

Universidad de Porto. Portugal. [rsilva\(AT\)up.pt](mailto:rsilva(AT)up.pt), [adasilva\(AT\)up.pt](mailto:adasilva(AT)up.pt)

INFORMACIÓN DEL ARTÍCULO

Artículo de Investigación

Recibido: 10-07-2015

Correcciones: 20-10-2015

Aceptado: 25-10-2015

Keyword

Requirements Engineering;
requirements elicitation; text
mining; data mining.

Palabras clave

Ingeniería de Requisitos; elicitación
de requisitos; minería de texto;
minería de datos.

ABSTRACT

In the software project the interested parts are highly distributed and form numerous and heterogeneous groups, online or face, constituting what could be called crowds. The development of social applications and cloud computing and mobile has generated a marked increase in environments based requirements in crowds. Technical Requirements Engineering (RE) traditional face these scalability issues, and require the co-presence of interested and engineers in joint meetings that can not be made in common physical environments. While different approaches have been introduced to partially automate RE in these contexts, still is required a multi-method approach to semi-automate all activities related to work with crowds. In this paper is propose an approach that integrates existing elicitation techniques and requirements analysis and is complemented by introducing new concepts. The information is collected through direct interaction and social collaboration, and through data mining techniques.

RESUMEN

En los proyectos software las partes interesadas están altamente distribuidas y forman grupos numerosos y heterogéneos, en línea o presenciales, constituyendo lo que se podría llamar multitudes. El desarrollo de aplicaciones sociales y de la computación en la nube y móvil ha generado un aumento marcado en entornos de requisitos basados en multitudes. Las técnicas de la Ingeniería de Requisitos (RE) tradicional enfrentan estos problemas de escalabilidad, y requieren la co-presencia de interesados e ingenieros en reuniones conjuntas que no se pueden realizar en entornos físicos comunes. Si bien se han introducido diferentes enfoques para automatizar parcialmente RE en estos contextos, todavía se requiere un enfoque multi-método para semi-automatizar todas las actividades relacionadas con el trabajo con multitudes. En este trabajo se propone un enfoque que integra técnicas existentes de elicitación y de análisis de requisitos y se complementa mediante la introducción de nuevos conceptos. La información se recoge a través de interacción directa y colaboración social, y mediante técnicas de minería de datos.

© 2015 IAI. All rights reserved

1. Introducción

La oferta de servicios y aplicaciones en línea abren el camino a un mercado potencialmente grande, pero la competencia es alta en este campo. Esto genera continuamente presiones en los desarrolladores y proveedores de servicios, que deben ofrecerles a sus clientes interacciones positivas e innovaciones con el fin de evitar que se cambien a soluciones de la competencia. La Ingeniería de Requisitos desempeña un importante papel en el mapeo y la anticipación a las necesidades de las partes interesadas. Sin embargo, las técnicas RE tradicionales dependen de la co-presencia de los analistas y los grupos de interés reunidos al mismo tiempo y en el mismo lugar, y por lo tanto no se adaptan bien a los ajustes de trabajo con muchas partes interesadas distribuidas [1]. Debido a que los usuarios se distribuyen físicamente se requieren técnicas remotas que les permitan a los analistas y a los muchos interesados ser activos en diferentes lugares y en diferentes momentos [2].

Las técnicas remotas vigentes dependen de la semi-automatización de aspectos de RE [2, 3], pero el mayor desafío es salvar las diferencias entre ellas en un enfoque integrado. Hasta ahora, la investigación en este aspecto se ha centrado en determinados sub-dominios y no se ha acercado lo suficiente de manera integral a este campo, a pesar que la automatización ha elevado los límites en las cantidades de datos, o a que el número de grupos interesados en ellos se incrementa paulatinamente. Es por eso que se requieren enfoques integradores, como el de la Ingeniería de Requisitos para Multitudes, que es un enfoque semi-automatizado para elicitar y analizar cualquier tipo de comentario desde una multitud de usuarios, con el objetivo de obtener requisitos validados. El dominio de RE debe considerar a la multitud como un grupo de actores actuales y potenciales, donde los clientes comparten o intercambian su experiencia con un producto en particular y de la medida en que se ajusta a sus necesidades, porque los requisitos se pueden derivar de estas declaraciones.

Sin embargo, para la elaboración de este enfoque se necesita una taxonomía exhaustiva de los comentarios de los usuarios, de la misma manera que las herramientas RE actuales para elicitación de requisitos a través de técnicas remotas que se centran en tipos de retroalimentación específicos, y que suelen utilizar solamente un método de interacción [4]. Además, y debido a que estas herramientas no capturan sistemáticamente los diferentes tipos de comentarios de los usuarios, potencialmente pasan por alto los datos relevantes y son propensas a sesgar la selección [4]. Por lo tanto, los datos solamente serán representativos de la multitud en conjunto si se recogen desde todas las dimensiones y se combinan las diferentes preferencias de los usuarios con respecto a la retroalimentación.

Con el fin de desarrollar una plataforma RE basada en multitudes y superar los retos conocidos, primero hay que alcanzar una mejor comprensión del estado del arte de la técnica y de los conceptos centrales para este enfoque. Por lo tanto, en este trabajo se abordan las siguientes preguntas de investigación: 1) ¿Qué soluciones se han sugerido y/o desarrollado para contrarrestar las limitaciones de RE tradicional con relación a la configuración de multitudes a través de la automatización? 2) ¿Cómo puede la retroalimentación de los usuarios ser categorizada en dimensiones relevantes?

2. Trabajos relacionados

El tema central del enfoque RE basado en multitudes es la multitud, que genera retroalimentación de los usuarios a través de múltiples dimensiones y de una amplia gama de canales y medios de comunicación. Esta retroalimentación se puede obtener a través de técnicas como la colaboración abierta distribuida y analizarla mediante minería de texto y minería de datos.

El concepto multitud se origina en las observaciones psicológicas del contacto social y otros patrones de conducta que se producen entre grandes grupos de personas inherentemente heterogéneos que se agrupan físicamente en un ambiente [5]. Las relaciones recíprocas y el poder del comportamiento del grupo se desarrollan a través de una reacción en cadena, y se propagan a través de disparadores particulares [6]. El comportamiento de una multitud en línea, donde las personas están explícitamente no-reunidas en un sólo lugar, es muy similar al de una multitud real, aunque varían en velocidad, tamaño y alcance [6]. Como resultado, una multitud se define como un grupo de actores, actuales y/o potenciales, lo suficientemente grande en tamaño como para mostrar comportamiento grupal, que tienen un interés común en un servicio en particular.

La retroalimentación de los usuarios que ofrece una multitud puede ser analizada para derivar las necesidades, los deseos, las ideas, los informes de errores y las pistas acerca de las tendencias, con el fin de anticiparse a nuevas innovaciones. Dado que esta información la proporcionan muchas partes interesadas, frecuentemente se pueden superar los problemas encontrados con pequeñas muestras. Por ejemplo, debido a que las partes interesadas son de diferentes culturas y orígenes, tendrán diferentes actitudes frente a un tema en particular, esto puede anular

la retención de información deliberada que realizan otros interesados, ya sea por razones políticas, sociales o emocionales [2].

La investigación fundamental en la caracterización de los comentarios de los usuarios es escasa, a pesar que varios trabajos proporcionan ideas importantes. Entre otros, un análisis del comportamiento de la retroalimentación en AppStore de Apple [7] reveló que pocos usuarios hacen más de una revisión, que escriben más críticas después de una nueva versión y que las aplicaciones con las que construyen una relación reciben considerablemente más revisiones. Almali y sus compañeros [4] analizan la actitud de cuatro grupos de usuarios hacia la retroalimentación. Estos autores encontraron que se diferencian en factores tales como la apertura a ser interrogado para proporcionar información, la medida en que se logra la privacidad de la minería de datos, y si se proporciona la retroalimentación como una motivación intrínseca o a raíz de factores sociales. Por su parte, Morales [8] caracteriza los comentarios de los usuarios [8], mientras que Morales y sus colegas [9] proporcionan una ontología de los comentarios de los usuarios para clarificar los conceptos de este dominio.

3. Metodología

Las herramientas existentes recogen declaraciones a través de colaboraciones sociales, minería de texto, o minería de datos, pero sin una práctica sistemática [4]. Hasta el momento, ninguna de las conocidas integra enfoques de múltiples o respeta las diversas formas en que los diferentes grupos de usuarios prefieren para proporcionar retroalimentación, lo que potencialmente excluye grupos de partes interesadas. Una solución multi-método ampliaría en gran medida la cantidad de datos recogidos, reduciría el sesgo de selección y permitiría estadísticamente más comparaciones de sondeo. Por eso, la metodología de este trabajo consiste en proponer una solución integradora que combine los enfoques existentes y abarque las dimensiones que hasta ahora se han descuidado.

A partir de una retroalimentación cuantitativa se realiza seguimiento y observación a los eventos y a los patrones de análisis, y posteriormente se seleccionan y ordenan los datos. Para la retroalimentación cualitativa los usuarios ya han proporcionado información, o están motivados para hacerlo, a través de interacciones e incentivos. Las declaraciones relevantes se filtran y analizan a través de un análisis no-lineal. Algunos datos pueden ser analizados a través de ambos procesos, por ejemplo, a través de la identificación de patrones en un video y el procesamiento de la transcripción del audio o de las notas de un observador. Los resultados de los dos procesos se pueden agregar y validar para obtener los requisitos.

Para desarrollar métodos de análisis adecuados se tienen en cuenta todos los tipos de miembros de la multitud, y hasta el momento para comprender de qué manera prefieren proporcionar información se han identificado cinco dimensiones de retroalimentación (Tabla 1).

La dimensión *modo* es similar a la comunicación de la retroalimentación que propone Morales [8]. La dimensión

conciencia utiliza diferentes términos para los tipos de retroalimentación explícita y pasiva que se presentan en [4, 8], debido a la posible confusión sobre las diferentes aplicaciones de la palabra explícito, y porque una parte interesada puede no ser pasiva cuando se genera

retroalimentación. Estos tipos de retroalimentación están cubiertos por combinaciones de las dimensiones propuestas.

Tabla 1. Taxonomía de las dimensiones de retroalimentación de usuarios

| Dimensión | Tipo | Descripción |
|----------------|------------------|---|
| Conciencia | Dirigida | Proporcionada intencionalmente para mejorar el producto, o con una esperanza razonable de que será utilizada para ese fin |
| | No-dirigida | Información relevante sin esa intención o esperanza |
| | Tácita | Demasiado clara para entender a qué se refiere |
| Modo | Directo | Se suministra en una interacción con el desarrollador como destinatario |
| | Indirecto | Comunicada mediante otro soporte |
| Tipos de datos | Cualitativos | Expresiones en lenguaje natural (incluyendo transcripciones) |
| | Cuantitativos | Estadísticamente analizables y cuantificables como datos |
| Intención | Clasificar | Con una valencia particular |
| | Justificar | Explicar por qué el usuario tiene una opinión particular |
| | Exigir | Verbalizar claramente una necesidad |
| Homogeneidad | Multitud pública | Expresado en un entorno abierto |
| | Multitud privada | Expresado en un grupo cerrado |

Las herramientas existentes les solicitan a los usuarios de forma dirigida una retroalimentación cualitativa, pero no es específica o cuando se pide el proceso es muy lento. Además, recogen datos cualitativos de forma no-dirigida, lo que significa que muchas otras configuraciones de las dimensiones de retroalimentación de usuario no están cubiertas en este punto, y se requiere un enfoque que lo haga. Las declaraciones que contienen requisitos tácitos [10], a menudo consideradas no-informativas, deben ser intrínsecamente relevantes en ese nuevo enfoque. Debido a que los autores de estas declaraciones no quedan satisfechos luego del esfuerzo de proporcionar retroalimentación, el hecho de solicitar ampliaciones produciría retroalimentación más específica. Al iniciar una sesión en la que se transmite la declaración, dimensiones de la retroalimentación como en línea o fuera de línea, o el tipo de los grupos de interés que la proporciona, se pueden capturar implícitamente en la propuesta de Ingeniería de Requisitos para multitudes.

4. Conclusiones

En este trabajo se propone la Ingeniería de Requisitos para multitudes, un principio general para todas las técnicas RE automatizadas, incluyendo el *crowdsourcing*, la minería de texto y la minería de datos. Esta propuesta intenta crear un enfoque integral en el que estén representadas todas las actividades RE. Se describe cómo involucrar la multitud y la retroalimentación de los usuarios y obtener de ellos todas sus facetas. Se presenta un resumen de los conceptos y las herramientas existentes en este campo (pregunta 1) y se introduce una taxonomía inicial de respuesta de la retroalimentación de los usuarios (pregunta 2). Además de refinar los conceptos y ejecutar pasos más pequeños, el trabajo de investigación futura se orienta principalmente a superar los retos que enfrentan los equipos al desarrollar plataformas de este tipo. En particular, se buscará responder a:

- Cómo resolver las reticencias o inquietudes de las partes interesadas en relación con el suministro de información, a través de la utilización de instrumentos de motivación adecuados y de niveles apropiados de intrusión.

- Cómo tratar con las consideraciones legales y de propiedad intelectual.
- Cómo extraer los requisitos de la retroalimentación de los usuarios para cada configuración de las cinco dimensiones, a través de algoritmos adecuados que no interfieran entre sí.
- Cómo clasificar, priorizar, reunir y validar los requisitos robustos, utilizando algoritmos, mediciones de validez de datos y la interacción directa con los usuarios.
- Cómo presentar claramente el número de requisitos y declaraciones, mostrando únicamente los datos pertinentes y utilizando agregaciones como visualizaciones gráficas.
- Cómo asegurar la trazabilidad de datos para deshacerlos sin esfuerzo utilizando sus atributos.

Referencias

- [1] Todoran, I., Seyff, N. & Glinz, M. (2013). [How cloud providers elicit consumer requirements: An exploratory study of nineteen companies](#). Proceedings 21st IEEE International RE Conference (pp. 105–114). Rio de Janeiro, Brazil.
- [2] Sutcliffe, A. & Sawyer, P. (2013). [Requirements elicitation: Towards the unknown unknowns](#). Proceedings 21st IEEE International RE Conference (pp. 92-104). Rio de Janeiro, Brazil.
- [3] Goldin, L. & Berry, D. (1997). [AbstFinder: A prototype natural language text abstraction finder for use in requirements elicitation](#). Automated Software Engineering 4(4), pp. 375–412.
- [4] Almali, M., Ncube, C. & Ali, R. (2014). [The design of adaptive acquisition of users feedback: An empirical study](#). Proceedings 8th International Conference on Research Challenges in Information Science (pp. 1–12). Marrakesh, Morocco.
- [5] Henein, C. & White, T. (2006). [Information in crowds: The swarm information model](#). Lecture Notes on Computer Sciences 4173, pp. 703–706.
- [6] Russ, C. (2007). [Online crowds: Extraordinary mass behavior on the internet](#). Proceedings I-MEDIA '07 and I-SEMANTICS '07 (pp. 65-69). La Quintana, USA.

- [7] Pagano, D. & Maalej, W. (2013). [User feedback in the AppStore: An empirical study](#). Proceedings 21st IEEE International RE Conference (pp. 125–134). Rio de Janeiro, Brazil.
- [8] Morales, I. (2013). [On exploiting end-user feedback in requirements engineering](#). Proceedings 19th International Working Conference on Requirements Engineering: Foundations for Software Quality, Doctoral Symposium Programme (pp. 223–230). Essen, Germany.
- [9] Morales, I., Perini, A. & Guizzardi, R. (2014). [Providing foundations for user feedback concepts by extending a communication ontology](#). Proceedings 33rd International Conference on Conceptual Modeling (pp. 305-312). Atlanta, USA.
- [10] Chen, N. et al. (2014). [AR-Miner: Mining informative reviews for developers from mobile app marketplace](#). Proceedings 36th International Conference on Software Engineering (pp. 767-778). Hyderabad, India.