

Basado en Evidencias

Resumen

En el contexto profesional, la necesidad de tomar mejores decisiones obliga cada vez más a los líderes y gerentes a incorporar conocimiento que haya sido comprobado empíricamente, como una manera de reducir el grado de incertidumbre. Es así como los estudios empíricos han llegado a cumplir un rol importante en el proceso de toma de decisiones. Entre ellos, los experimentos controlados son el tipo de estudio más ampliamente aceptado en el contexto académico e industrial. En Ingeniería del Software (IS) la conducción de experimentos controlados comienza a tomar auge en la última década. Tales experimentos han buscado comprobar, de una manera mucho más rigurosa, las ventajas de ciertas tecnologías frente a otras; o la utilidad de una tecnología particular en una situación dada. A pesar de esto, no podemos afirmar aún que la Ingeniería de Software Experimental sea una disciplina madura. Por el contrario, existen serias debilidades en cuanto a la cantidad y calidad de los experimentos individuales que podemos localizar, así como restricciones para recuperar los mismos utilizando los medios electrónicos disponibles. Todo lo anterior nos presenta limitaciones cuando pretendemos tomar decisiones basadas en las evidencias obtenidas de los experimentos individuales. Una manera de aumentar la validez de tales evidencias es combinar los resultados individuales por medio de métodos cuantitativos o cualitativos. Es así como surgen los estudios secundarios, tales como las Revisiones Sistemáticas (RS), que combinan los resultados de los estudios primarios previamente conducidos.

A pesar de que han sido propuestos lineamientos para la conducción de RS en la IS, éstos presentan debilidades y carencias, motivado por la fuerte influencia que han recibido de propuestas equivalentes que se encuentran en uso en otras disciplinas; sin considerar un conjunto de características que son particulares a la IS (p.e. una base experimental escasa y de calidad muy diversa). Esto lleva a que las RS en IS sean muy complejas y costosas. Frecuentemente, los revisores tienen que resolver de manera intuitiva dificultades de diferente naturaleza que impiden aprovechar la base de experimentos existentes. En otras ocasiones, los resultados que se obtienen no llegan a ser significativos, como consecuencia de un número muy pequeño de experimentos y sujetos. Finalmente, esto tiene un impacto negativo en la transferencia de razonamientos empíricamente comprobados a los tomadores de decisiones en el ámbito industrial.

Área de Investigación

La Medicina Basada en la Evidencia (MBE) ha sido adoptada por varios otros dominios, incluyendo la Ingeniería de Software; donde el mecanismo normal para la identificación y la agregación de evidencia es la revisión sistemática de la literatura; dando origen a lo que se conoce, por analogía con Medicina, como Ingeniería de Software Basada en la Evidencia (EBSE), la cual puede definirse como una aproximación a los estudios empíricos en la que el investigador busca identificar e integrar la mejor evidencia disponible con la experiencia de campo para informar sobre prácticas y formulación de políticas.

ESBE tiene como objetivo principal mejorar la toma de decisiones relacionadas con el desarrollo y mantenimiento de software mediante la integración de la mejor evidencia actual de Investigación con la experiencia práctica y los valores humanos.

El nacimiento de las RS en IS es muy reciente.

A comienzos del siglo XXI aparecen las primeras revisiones:

- (Döring, 2002) presenta una RS sobre treinta estudios relacionados con el desarrollo de sitios Web personales;
- (Whitten & Stephens, 2002) describen el análisis utilizado para producir nuevo conocimiento acerca de las funciones infomediarias;
- (Premkumar, 2003) evalúa el impacto de ciertos factores en la adopción de tecnologías de comunicación basadas en el ordenador;
- (C.Wohlin, Peterson, & Aurum, 2003) realizó la combinación de 30 grupos de datos para evaluar la influencia de varios aspectos diferentes, incluyendo técnicas de lectura, tamaños de equipo y perfil de los sujetos (profesionales versus estudiantes).
- (B.Kitchenham, 2004) que se comienza a hablar de una propuesta de lineamientos para RS en IS.
- (Cruzes & Dyba, 2010) encontraron un conjunto de 31 revisiones en el período 2004-2009.
- (B. Kitchenham, Pretorius, et al., 2010) (B. Kitchenham, Pretorius, et al., 2010)se localizaron 33 RS en el período 2004-2010.

Al mismo tiempo que se dan estas revisiones, aparecen también publicados otros trabajos relacionados con aspectos que influyen en los resultados de una RS:

- Jedlitschka & D. Pfahl, 2005; B. Kitchenham et al., 2006; B. Kitchenham, Pfleeger, Pickard, Jones, Hoaglin, Emam, et al., 2002)
- (T. Dybå, Kitchenham, & Jørgensen, 2005; B. Kitchenham, Budgen, Brereton, & Linkman, 2005; B. Kitchenham, Dybå, & Jørgensen, 2004; Rainer, Jagielska, & Hall, 2005) que establecieron los beneficios y limitaciones de la IS basada en Evidencia (EBSE, por su acrónimo inglés Evidence based Software Engineering) y establecieron que el propósito de esta práctica es soportar, con teorías basadas en evidencias empíricas, el proceso de toma de decisiones de los profesionales de la IS.

Identificación del problema

La IS es un área donde el proceso de toma de decisiones involucra situaciones como la evaluación de costos, límites, calidad, y riesgos inherentes a un proyecto. Los ingenieros de software pueden tomar decisiones incorrectas, tales como la adopción de nuevas técnicas, si no consideran evidencia científica acerca de la eficacia de tales técnicas.

Para que el proceso de toma de decisiones pueda ser soportado por la evidencia científica, es necesario conducir estudios secundarios, como las revisiones sistemáticas, que permitan obtener nuevo conocimiento validado a través de los estudios primarios.

Puede observarse en la literatura como las metodologías que están en uso en otras disciplinas, han demostrado ser muy efectivas en sus respectivos contextos. Como ejemplo se puede mencionar el catálogo de la Cochrane que cuenta con más de 12000 revisiones del área de Medicina. En tales disciplinas existen metodologías reconocidas, tales como (EPPI-CENTRE, 2002; EPPI-CENTRE, 2003; Higgins, 2006; Khan et. al, 2001; NHMRC, 2000; Pai et. al, s/f), que orientan en la realización de estos procesos y el reporte de sus resultados.

Es comprensible que en las propuestas actuales en IS es notoria la influencia de las metodologías para RS de Medicina. Sin embargo, dichas metodologías no son necesariamente extrapolables a otras áreas del conocimiento y en particular a la IS.

Encontramos en la práctica que existen problemas para la aplicación de tales metodologías, ya que la IS presenta diferencias importantes con las disciplinas de las Ciencias de la Salud.

- La motivación se deriva del interés por conocer el estado de madurez del conocimiento sobre una tecnología (por ejemplo, ¿qué sabemos sobre técnicas de prueba?). Como consecuencia no se puede definir un objetivo con suficiente precisión desde el comienzo de la RS (Higgins, 2006), y es muy difícil realizar un desarrollo temprano del protocolo de revisión.
- En relación a la búsqueda, vemos que en IS la exhaustividad en las búsquedas es necesaria debido a que se cuenta con un número reducido de experimentos. Debido a que la terminología utilizada en el área no es estándar, el proponer cadenas de búsqueda predefinidas es imposible.
- La Síntesis de datos se ve afectada por la falta de estandarización. Experimentos que analizan los mismos factores y variables respuesta los pueden reportar de manera diferente, creando confusión al agregarlos.
- La extracción de datos no es sencilla ya que también se ve afectada por la falta de estandarización del vocabulario y la baja calidad de los reportes.
- Los aspectos de un estudio (por ejemplo, ocultamiento, enmascaramiento, etc.) que influyen en su calidad no son claros. Esto que ha resultado evidente en nuestros resultados preliminares, también ha sido ampliamente discutido en el caso de otras disciplinas (Concato and Horowitz, 2004) (Feinstein and Horowitz, 1977) (Lawlor et. al, 2004), en las que tampoco se tiene claro qué aspectos pueden influir directamente en los resultados del estudio.
- El contar con un número muy pequeño de experimentos, que incluyen pocos sujetos, afecta al nivel de certeza de los resultados obtenidos a través de meta-análisis tradicional (especialmente si el tamaño del efecto estudiado es muy pequeño). Además de esto, existen diferentes variables moderadoras que pueden influir en los resultados de cada experimento llevando muchas veces a obtener resultados muy heterogéneos, incluso contradictorios.

En las características anteriores podemos observar que más que aplicar metodologías importadas de otras disciplinas se requiere un proceso particular de RS para IS.

Importancia del problema de investigación

No es desconocida para la mayoría de los profesionales y académicos la complejidad que caracteriza los proyectos de IS:

- tasa de incumplimiento alta en relación a los tiempos y presupuestos.
- crisis del software
- retos continuos en cuanto al uso de las TIC con el fin de proveer mayores beneficios y prestaciones a los clientes y usuarios finales.

Todo lo anterior ha llevado a que continuamente se propongan nuevas tecnologías de desarrollo (por ejemplo; lenguajes, marcos de desarrollo, bases de dato, tipos de datos, etc.) así como técnicas, métodos y metodologías conducentes a resolver los problemas inherentes al desarrollo de software, especialmente en prácticas como la estimación, la elicitación de requisitos, la planificación, entre otros. De esta manera, las tecnologías propuestas suponen una serie de beneficios, pero también desventajas para los desarrolladores.

Las metodologías de desarrollo no incluyen necesariamente métodos específicos para su adopción (técnicas de prueba, técnicas de elicitación de requisitos, etc.), las técnicas recomendadas no son vinculadas con un determinado perfil del analista o desarrollador, se desconoce si las mismas funcionan igual de bien para diferentes dominios, etc.

En este sentido la investigación en IS es falible, por que se basa en generalizaciones que pueden ser difíciles de interpretar, y que es a menudo insuficiente para determinar los medios adecuados para la entrega de la mejor práctica.

Los profesionales y gerentes de software que buscan la calidad de su desarrollo de software suelen adoptar nuevas tecnologías sin pruebas suficientes. Estos gestores y profesionales de software a menudo tienen que tomar decisiones acerca de diversos procesos. Las compañías de software están a menudo bajo presión de adoptar tecnologías inmaduras, debido a las presiones del mercado y de gestión.

Esta complejidad de la IS, sumada a la proliferación de propuestas más o menos aisladas y de las cuales se cuenta con poca información y aplicación, genera interés en ámbitos distintos.

En primer lugar, en el ámbito académico surge la preocupación por validar tales propuestas desde un punto de vista empírico; comparando el uso de estas tecnologías entre sí, o el uso de una tecnología en particular versus no utilizar ninguna tecnología. Lo cual lleva a la conducción tanto de estudios primarios (tales

como experimentos controlados) como estudios secundarios, como las revisiones sistemáticas. Éstos últimos, son especialmente importantes ya que permiten lidiar con las limitaciones de los estudios primarios, los cuales suelen contar con pocos sujetos.

Sin embargo, la experiencia revela que la conducción de RS enfrenta diversas dificultades que limitan su aplicación en la IS. Es por ello que mientras que encontramos una cantidad creciente de estudios de mapeo, el número de RS no ha crecido considerablemente en la última década.

Se requiere entonces de orientaciones metodológicas que permitan obtener nuevas evidencias con ciertas garantías y que alimenten la base de conocimiento acerca de las nuevas tecnologías o, incluso, sobre las tecnologías existentes.

Mientras tanto, en el ámbito industrial la necesidad de evidencias que soporten la selección de las tecnologías más adecuadas para un determinado proyecto, es una constante que incide directamente en la rentabilidad del negocio y genera continuas presiones para los gerentes. Más aún cuando observamos que en la actualidad, con el avance acelerado de la tecnología, el cierre de las brechas digitales, y el dinamismo de las organizaciones, el proceso de toma de decisiones cobra una importancia crítica.

METODOLOGÍA

Para el logro de los objetivos planteados en el Capítulo XX, se ha propuesto un enfoque metodológico basado en investigación-acción (I-A) (Creswell, 2002).

De acuerdo con (Baskerville & Wood-Harper, 1996) el domino ideal de la I-A está caracterizado por una configuración social donde la investigación es un proceso (generalmente cíclico) que enlaza la teoría con la práctica. En la Figura 99.1 se muestra un resumen de las fases y actividades propuestas.

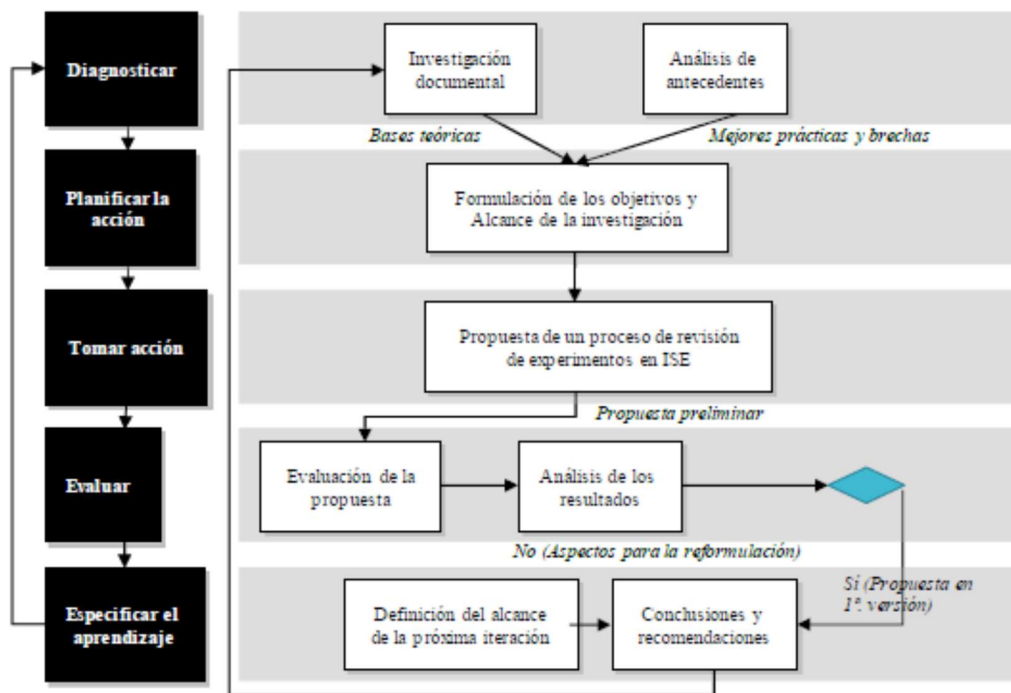


Figura 4.1. Enfoque metodológico. Adaptado de (Pérez, Grimán, Mendoza, & Rojas, 2004)

Fases

A) Fase de Diagnosticar: identifica los problemas primarios que se traducen en las razones por las cuales la organización desea cambiar. En el caso específico de este proyecto de investigación se considera como organización a cualquier comunidad de IS, bien sea de carácter investigador o profesional, en el que se requiera llevar a cabo RS de experimentos. Sin embargo, algunos de los problemas identificables en esta disciplina pueden ser de comunes a otros dominios.

Las actividades asociadas a esta fase son las siguientes:

1. Investigación documental: se refiere a la revisión de la bibliografía relacionada con RS y experimentación, para establecer un marco teórico. Implica la extracción desde fuentes diversas (incluidas las fuentes electrónicas) con el fin de construir la base conceptual que servirá de referencia a la investigación. Los productos a obtener incluyen un conjunto de conceptos, así como aspectos sociales, tecnológicos y organizacionales a ser considerados durante la investigación.
2. Análisis de antecedentes: se refiere al estudio del Estado de la Cuestión relativo al tema que es objeto de estudio: Revisiones Sistemáticas. Incluye la recolección y revisión de métodos, procesos o marcos de referencia para la conducción de RS en IS y otras disciplinas con experiencia en las prácticas empíricas. Durante esta actividad se busca establecer tareas recurrentes, ventajas/desventajas de los trabajos previos y buenas prácticas; así mismo, identificar problemas abiertos que guíen la definición de objetivos de la investigación.

B) Fase Planificar la acción, establece las acciones a través de las cuales deberían ser eliminados o mejorados los problemas principales que fueron identificados en la etapa anterior. Esta fase incluye una única actividad:

3. Formulación de los objetivos y alcance de la investigación: durante esta actividad se formula el alcance de la investigación; utilizando como entradas los resultados de las actividades precedentes.

Posteriormente, se debe acometer la fase de Tomar la acción, en la cual se implementa la acción planificada. Los participantes e investigadores colaboran en una intervención activa en la organización, canalizando los cambios. Para ello, la actividad principal es:

4. Propuesta de un proceso de RS para IS: esta actividad busca directamente la satisfacción de los objetivos establecidos en la fase anterior. El resultado esperado en el contexto de esta investigación es un proceso de RS para experimentos en IS en versión beta (prueba) junto con las consideraciones del contexto en el cual debería ser aplicado.

C) Fase de Tomar la acción, en la cual se implementa la acción planificada. Los participantes e investigadores colaboran en una intervención activa en la organización, canalizando los cambios. Para ello, la actividad principal es:

4. Propuesta de un proceso de RS para IS: esta actividad busca directamente la satisfacción de los objetivos establecidos en la fase anterior. El resultado esperado en el contexto de esta investigación es un proceso de RS para experimentos en IS en versión beta (prueba) junto con las consideraciones del contexto en el cual debería ser aplicado.

D) Fase Evaluación, los investigadores y participantes evalúan los resultados para determinar si los efectos teóricos de la acción fueron logrados y si estos efectos solventaron el problema.

Para cubrir esta fase las actividades que se deben acometer son las siguientes:

5. Evaluación de la propuesta: para determinar la mejor manera de validar el proceso propuesto, se utilizará el método DESMET propuesto por (B. Kitchenham, 1996), el cual proporciona una manera sistemática y objetiva para seleccionar un método de evaluación para herramientas y métodos en IS. Dado el carácter genérico de los métodos cualitativos y cuantitativos considerados por DESMET, se considera que éste puede ser utilizado sin riesgo para la selección del método de evaluación adecuado a este proyecto de investigación.

6. Análisis de resultados: durante esta actividad se estudian los resultados con base en los objetivos establecidos inicialmente y la evaluación previamente realizada, en términos de los productos tangibles alcanzados y los cambios producidos en el ambiente. Posteriormente, los cambios que surjan de este análisis pueden ser incorporados en una segunda versión (formal) del proceso propuesto, la cual será utilizada en futuras iteraciones de la investigación. En caso de que los resultados no fueran satisfactorios respecto a los objetivos deberá acometerse la siguiente actividad para revisar el diseño del proceso de RS o los objetivos planteados. Sin embargo, en virtud de las restricciones de tiempo establecidas en el proyecto, puede decidirse llevar a cabo un único ciclo metodológico, al final del cual los cambios requeridos por el proceso no sean acometidos.

E) Fase Especificar el aprendizaje los investigadores deben especificar el conocimiento adquirido, basados en la evaluación.

Para ello, las actividades que se realizarán son:

7. Definición del alcance de la próxima iteración: esta actividad tomará lugar en caso de que los resultados de la evaluación realizada durante la fase anterior no sean satisfactorios. De esta manera, se establece la estructura de una nueva iteración del ciclo de I-A.

8. Conclusiones y recomendaciones: en caso de que los resultados obtenidos sean satisfactorios, se establecen las conclusiones sobre el proceso de RS ya aplicado; en caso contrario, se refuerza el alcance de la próxima iteración del ciclo de I-A.

VISIÓN GENERAL DE PROCESO

Características de la IS que provocan variaciones en el proceso de RS

Una RS puede llegar a ser muy costosa en tiempo y recursos, por lo que es importante contar con una metodología de RS adecuada, bien definida y que promueva la coste-efectividad del proceso. Lo anterior se traduce en una metodología que: 1) utilice métodos adecuados, que minimicen el riesgo de incurrir en sesgos; 2) combine evidencias primarias con una calidad mínima aceptable.

Las metodologías y procesos que han sido descritos anteriormente han probado ser efectivas en sus respectivas disciplinas, especialmente aquellas que son utilizadas más popularmente en las Ciencias de la Salud y que han sido practicadas y mejoradas durante años.

Sin embargo, las características de las disciplinas relacionadas con las Ciencias de la salud (Medicina, Psiquiatría, Enfermería, etc.) parecieran distar bastante de las características de la IS. Esto hace suponer que las metodologías que se encuentran actualmente en uso en tales disciplinas no sean completamente aplicables a la IS.

Se pretende proponer un proceso de RS que resuelva aquellas características de la IS que hacen que las metodologías o procesos actuales no sean aplicables de manera efectiva. Por lo tanto, antes de definir la propuesta realizada en esta investigación, es necesario analizar las características particulares de la IS que afectan al proceso de RS. Las características que han sido identificadas en este trabajo son las siguientes:

- Debido a que la ISE es una disciplina reciente, los experimentos son escasos. Puesto que son conducidos generalmente en ambientes académicos, típicamente incluyen pocos sujetos.
- Generalmente los revisores desconocen si existen experimentos disponibles acerca de un tema de interés por lo que es difícil establecer, en una etapa temprana, una pregunta de revisión cuya respuesta sea viable en virtud de que se cuente con evidencia primaria para soportarla.
- Como consecuencia de la característica anterior, es difícil establecer un protocolo de revisión al comienzo del proceso de RS.
- Es difícil localizar los experimentos existentes debido a bases de datos bibliográficas disponibles en IS son inadecuadas.
- No se cuenta con una terminología estandarizada en relación con la tecnología bajo estudio o las mediciones que son realizadas sobre éstas.
- Típicamente, aún los experimentos que ensayan una misma tecnología, evalúan diversas variables respuesta y métricas.
- Apenas algunos experimentos recientes reportan adecuadamente los datos y resultados del estudio; la mayoría de experimentos presenta una baja calidad de reporte.
- Los experimentos son conducidos en contextos muy diversos por lo que los grupos de experimentos suelen ser muy heterogéneos.
- Debido a que la conducción de los experimentos se realiza de manera diferente a otras disciplinas relacionadas con las Ciencias de la Salud, no todos los factores que afectan la calidad de un experimento en tales disciplinas pueden ser considerados en IS.

Experimentos escasos y con pocos sujetos

En las disciplinas relacionadas con las Ciencias de la Salud y con las Ciencias Sociales, se pueden encontrar una gran cantidad de estudios empíricos, y más específicamente experimentos controlados aleatorizados (conocidos también como RCT, por su acrónimo inglés). Esto hace posible que al momento de realizar una RS se disponga de un gran número de experimentos potencialmente agregables para responder a la pregunta de revisión planteada.

Al contar con un volumen considerable de experimentos, los revisores pueden permitirse ser muy estrictos en la selección de los mismos; utilizando para ello diversos criterios de exclusión de experimentos; por ejemplo, relacionados con los sujetos, los tratamientos, los tipos de estudio o las variables respuesta.

La ISE es, en cambio, una disciplina muy reciente; por lo que no se dispone de una gran base de estudios empíricos. De los estudios empíricos disponibles, una parte considerable son estudios observacionales o casos de estudio; siendo los experimentos un tipo de estudio muy escaso en esta disciplina.

Por otra parte, la mayoría de los experimentos existentes han sido conducidos en ambientes académicos por lo que, típicamente, cuentan con un número pequeño de sujetos.

Esta situación dentro de la IS, afecta directamente la validez de las evidencias que puedan ser obtenidas en una RS ya que el tamaño de la muestra es uno de los factores que determinan el poder estadístico de una agregación (Cohen, 1977). Como resultado de esto, frecuentemente durante la agregación es difícil detectar los efectos entre los niveles que están siendo comparados; por otra parte, cuando se obtienen evidencias, éstas tienden a estar soportadas por un número muy pequeño de estudios.

Dificultad para establecer una pregunta de revisión temprana

Típicamente, la pregunta de revisión está determinada por 4 componentes básicos: participantes, intervenciones, diseños de estudios y resultados. Dichos componentes permiten planificar y plasmar en un protocolo de revisión las actividades y las decisiones principales acerca de la investigación a realizar.

Prácticamente todas las metodologías presentadas parten de la formulación de una pregunta de revisión bastante precisa. Esto tiene sentido en disciplinas experimentales maduras y que cuentan con una amplia base de experimentos, ya que es fácil conocer de antemano la disponibilidad de experimentos relacionados con los tratamientos de interés.

Este no es el caso de la IS. Por lo tanto, establecer una pregunta de revisión al comienzo del proceso frecuentemente provoca que los revisores realicen un esfuerzo inútil al intentar localizar experimentos relacionados con la pregunta de revisión.

Por lo anterior, es frecuente que los revisores (especialmente los que realizan una RS por primera vez) comiencen el proceso de RS con un objetivo exploratorio, con la intención de detectar posibles preguntas de revisión. Sólo aquellos revisores que ya han explorado y analizado la base de estudios empíricos disponibles cuentan con suficiente información como para plantearse una pregunta de revisión que eventualmente pudiera ser respondida.

Aunque actualmente existen algunas recomendaciones para sobrellevar la dificultad de establecer la pregunta de revisión, tales como realizar estudios de mapeo o búsquedas de alcance.

Dificultad para establecer un protocolo de revisión

Casi todas las metodologías analizadas, incluyendo a LRSIS y UFRJ (desarrolladas específicamente para la IS), establecen que en una etapa inicial del proceso se desarrolle un protocolo de revisión, que contribuya a gerenciar la complejidad del proceso y establecer los juicios, decisiones y recursos necesarios para la RS.

El desarrollo del protocolo de revisión exige que muchas de las decisiones relativas al objetivo a alcanzar, así como los criterios y métodos a utilizar durante la RS, sean establecidos antes de conocer la disponibilidad de experimentos relacionados con el tema de interés. Por lo tanto, las metodologías

mencionadas parten de la premisa de que los revisores cuentan con un objetivo preciso desde el comienzo de la revisión, lo cual les permite establecer tempranamente tal protocolo.

Debido a la escasez de experimentos en IS, frecuentemente es muy difícil conocer anticipadamente de qué experimentos se dispone; por lo que la mayoría de las revisiones tienen un objetivo exploratorio, puesto que se desconoce si sería posible localizar experimentos relacionados con tratamientos y variables respuesta específicos.

Nótese que aunque tales revisiones son presentadas como RS, en realidad se trata de revisiones de alcance, estudios exploratorios o de mapeo.

Como resultado, de la misma manera que pierde sentido en IS establecer una pregunta de revisión muy tempranamente; es cuando menos discutible establecer un protocolo de revisión cuando aún se desconoce qué tratamientos, variables respuesta, tipos de estudio y tipos de sujetos podrán ser localizados.

Dificultad para localizar los experimentos existentes debido a BD inadecuadas

En Medicina, cuando los revisores requieren localizar estudios primarios, pueden consultar las bases de datos bibliográficas especializadas de las cuales disponen (PubMed, MEDLINE, CENTRAL, EMBASE, AMED, CINAHL y PsycINFO).

Estas bases de datos brindan servicios que incluyen desde la indexación de artículos que reportan experimentos hasta la facilidad de realizar búsquedas por características muy específicas del experimento (p.e. tipo de estudio, condiciones, intervenciones, mediciones, localidad); pasando por la clasificación de experimentos por tópico.

A diferencia de Medicina, en IS no existen bases de datos bibliográficas especialmente destinadas al registro de estudios empíricos realizados en el área. Existen algunas iniciativas de coleccionar estudios primarios en un único sitio (como la web Evidence-based Software Engineering¹); sin embargo, esto dista mucho de una base de datos como las que han sido citadas. Por lo tanto, los revisores en IS se ven limitados a utilizar bases de datos bibliográficas de carácter generalista (ACM, Google Scholar, IEEE Xplore, Science Direct o Web of Science).

Debido a que estas bases de datos no son específicas para el registro de estudios empíricos, la localización de experimentos se hace compleja y costosa, ya que para obtener todos los artículos que reportan efectivamente experimentos relacionados con la pregunta de revisión, se debe invertir tiempo y esfuerzo considerables en descartar otros artículos que, habiendo sido recuperados durante la búsqueda, reportan ciertamente experimentos o, si lo hacen, éstos no se corresponden con el objetivo de los revisores.

Además se han reportado diferentes problemas de cobertura y desempeño de estas bases de datos.

Todo lo anterior hace pensar que se requiere de criterios para la selección de las bases de datos a utilizar durante la RS, a fin de que la búsqueda se realice de la manera más coste-efectiva posible. Sin embargo, estas limitaciones que afectan la búsqueda de experimentos en IS no son abordadas por las metodologías que han sido propuestas para esta disciplina.

Visitar el siguiente link: <http://www.dur.ac.uk/ebse/>

A diferencia de Medicina, IS es una disciplina joven. Sin embargo, debido al avance tecnológico acelerado, se ha producido un crecimiento descontrolado de la terminología. Aunque existen algunas iniciativas que intentan estandarizar el vocabulario y los conceptos relacionados con la disciplina, (ACM, 1998; IEEE, 1990, 2004a), aún no se cuenta con un vocabulario controlado que pueda ser realmente de utilidad en el manejo de la información.

De esta manera, se observan los siguientes problemas:

- Ambigüedad inherente al lenguaje humano natural (por ejemplo, método de inspección, técnica de inspección y técnica de lectura pueden significar lo mismo o estar relacionados para los expertos, pero esto es difícil de interpretar para el ordenador).
- Dificultades en el intercambio de información entre proveedores, investigadores, profesionales, etc.
- Inconsistencias en el registro de la información de un lugar a otro.
- Dificultad para clasificar la información.
- Dificultad para la manipulación simbólica de los datos; esto es, búsquedas por análisis específico.

Esta característica puede limitar fuertemente la búsqueda de estudios primarios, así como la extracción correcta de los datos; por lo que requiere especial Interés. En el caso de la búsqueda, la recuperación de toda la evidencia primaria disponible acerca de una pregunta de revisión es prácticamente imposible, ya que se requeriría conocer todos los términos equivalentes utilizados por los experimentadores para referirse a la tecnología que está siendo ensayada. En el caso de la extracción de datos, es muy difícil garantizar que los datos sean extraídos de manera correcta; lo cual finalmente terminará por afectar la síntesis de datos, al combinar resultados de métricas distintas o dejar por fuera resultados de métricas equivalentes.

Falta de una terminología estandarizada

En algunas disciplinas, como Medicina o Física, los experimentos que comparan los mismos fenómenos utilizan comúnmente la misma manera de evaluarlos; esto se traduce en que típicamente las respuestas a los tratamientos son medidas utilizando las mismas métricas. Por ello, contar con un número considerable de experimentos que soporten las evidencias relacionadas con una métrica, no es una preocupación para las metodologías de RS en estas disciplinas.

En cambio, en IS es común encontrar experimentos que miden una misma variable respuesta a través de métricas distintas; por ejemplo, la variable respuesta eficiencia del método de inspección podría ser medida en un experimento como la tasa de detección de fallos por unidad de tiempo (m_1), mientras que en otro experimento podría estar siendo medida como el tiempo transcurrido hasta encontrar un fallo (m_2). Nótese que ambas métricas son diferentes no sólo en la unidad en que son medidas (número de fallos, en el caso de m_1 ; y minutos, en el caso de m_2) sino también en la interpretación de la dirección del efecto; es decir, si el valor de m_1 es mayor para un método, puede interpretarse que tal método es más eficiente que el otro (detecta más fallos por cada unidad de tiempo), mientras que cuando el valor de m_2 sea mayor para un método deberá interpretarse que éste es menos eficiente (toma más tiempo encontrar cada fallo). Por lo anterior, m_1 y m_2 no podrían ser agregadas estadísticamente.

La situación anterior es bastante común en IS; trayendo como consecuencia que para conocer el comportamiento de una determinada tecnología se tengan que recopilar las evidencias disponibles acerca de múltiples métricas, las cuales son soportadas por muy pocos experimentos. Sin embargo, estas limitaciones no han sido abordadas aún desde el punto de vista metodológico.

Baja calidad de reporte de los experimentos

Es frecuente que un experimento, aún estando adecuadamente diseñado y habiendo sido conducido apropiadamente, sea excluido de la revisión en virtud de que no reporta toda la información necesaria. En algunas ocasiones, estas deficiencias en el reporte del experimento puede ser solventadas por los revisores comunicándose con los autores del experimento. Sin embargo, muchas veces se trata de experimentos antiguos de los cuales los mismos investigadores que condujeron el experimento no guardan los datos originales.

Estas deficiencias limitan no sólo la calidad desde el punto de vista del reporte; sino que, la mayoría de las veces, afecta también su calidad como estudio, ya que es común que los revisores respondan

a los criterios de valoración de la calidad con base en la información (por ejemplo, aleatorización del experimento, ocultamiento, etc.) que es suministrada en el reporte del experimento. Finalmente, será la SÍNTESIS DE DATOS la que más se vea perjudicada, ya que para aplicar las técnicas de meta-análisis más efectivas para la agregación, es necesario contar con un conjunto de datos cuantitativos (número de sujetos, medias y varianza). Al no contar con estos datos, otras técnicas pueden ser aplicadas, no sin comprometer la confiabilidad de las evidencias obtenidas.

Aunque parece ser común a varias disciplinas, el problema de la baja calidad de reporte en IS se suma al problema de la escasez de experimentos haciendo su impacto más dramático en la SÍNTESIS DE DATOS.

Grupos de experimentos con contextos heterogéneos

En IS la conducción de un experimento presenta condiciones que lo diferencian de un experimento en Medicina:

- Los sujetos aplican los tratamientos (en lugar de recibirlos). Para ello, los sujetos deben aplicar una técnica o método relacionado con alguna etapa del proceso de desarrollo.
- El método o técnica es aplicado a un artefacto (objeto).
- Generalmente, los resultados son medidos sobre los objetos.

Por lo anterior, otros factores deben ser tomados en cuenta en IS para establecer el contexto del experimento, ya que los mismos podrían afectar la generalización de los resultados y su aplicabilidad. Uno de estos factores particulares a la IS se refiere a la Tarea experimental que es realizada por los sujetos durante una etapa del proceso de desarrollo; y durante la cual son aplicados los tratamientos. Por otra parte, se tiene también al Objeto experimental, que se refiere al artefacto al que está siendo aplicada la técnica. Estos factores pueden afectar la SÍNTESIS DE DATOS al producir variaciones en los resultados de un grupo de experimentos.

Definición de calidad poco clara y limitaciones en las escalas de calidad

En la IS no existe mucha madurez en relación a la experimentación y más aún en relación a lo que puede definirse como un experimento de calidad. Ante este vacío los revisores suelen importar directamente de Medicina las escalas existentes, sin hacer mucha adecuación de las mismas antes de su uso en esta disciplina.

Se requiere entonces de un estudio donde los factores que inciden en la calidad de un experimento en IS sean analizados, considerando quizás aquellos que han sido tradicionalmente utilizados en otras disciplinas, pero sin perder de vista las características propias de los experimentos que se realizan en ésta.

Por ejemplo, en los experimentos que son conducidos en la IS encontramos que el enmascaramiento llega a ser una variable difícil de manejar, ya que supone que el sujeto no conozca del tratamiento que está siendo aplicado; sin embargo, la propia naturaleza de la tarea experimental lleva a que esto no sea posible, ya que el sujeto es el responsable de aplicar el tratamiento que se está comparando.

Por último, se tiene que mientras en otras disciplinas abundan las escalas de calidad y listas de verificación, en IS la cantidad de estos instrumentos es escaso; siendo que algunas de ellas contienen gran número de ítems y deben ser seleccionados para cada revisión con muy pocas orientaciones para ello; o en cambio no han sido propuestos específicamente para experimentos controlados.