

Resumen

Las patentes son documentos legales que describen el funcionamiento exacto de una invención, otorgando el derecho de explotación económica a sus dueños a cambio de dar a conocer a la sociedad los detalles de funcionamiento de dicha invención. Para que una patente pueda ser concedida debe cumplir tres requisitos: ser novedad (no haber sido expuesto o publicado con anterioridad), cumplir la actividad inventiva y tener aplicación industrial. Es por ello que las patentes son documentos valiosos, ya que contienen una gran cantidad de información técnica no incluida antes en otro tipo de documento (publicado o disponible).

Debido a las características particulares de las patentes, los recursos que éstas mencionan, así como los recursos que mencionan a las patentes, contienen enlaces que pueden ser útiles y dar apoyo a diversas aplicaciones (vigilancia tecnológica, desarrollo e innovación, Triple-Helix, etc.) al disponer de información complementaria, así como de la creación de herramientas y técnicas que permitan extraerlos y analizarlos.

Por este motivo, la presente tesis doctoral plantea la posibilidad de descubrir recursos web de calidad a través de los enlaces mediante el diseño de un método adecuado para su extracción y análisis eficiente aplicando el uso de la técnica cibernétrica de análisis de enlaces.

Para ello, se ha descrito en el estado de la cuestión dos grandes bloques: las patentes y la cibermetría, con el objetivo de ofrecer un marco de trabajo que permita entender correctamente el alcance del análisis. Tras la exposición teórica del campo de estudio, se realiza una explicación del método propuesto para alcanzar los objetivos que definen la tesis. Este método se encuentra dividido en dos bloques complementarios: Patent Outlink y Patent Inlink, que juntos conforman la técnica de Patent Link Analysis. Ambos bloques siguen un proceso de trabajo equivalente. En primer lugar, se selecciona la fuente de la cual extraer los datos, tras esto se recopilan aquellos necesarios y se almacenan en bruto con doble copia para evitar pérdidas de información, de este modo es posible recuperar en todo momento la información inicial.

Se analizan los datos brutos para entender el tipo de información y estructura en la que se encuentran y que permita preparar un sistema de extracción específico para cada bloque. En el caso del Bloque Patent Outlink, se desarrolla un programa ad hoc para la extracción de enlaces contenidos en los documentos en patentes, que permite extraer la mayor cantidad posible de enlaces con el mínimo ruido, y recopilando a su vez información relacionada con la sección de aparición o la categoría de la patente de modo que se pueda realizar un análisis en mayor profundidad.

Para realizar el estudio se selecciona la Oficina de Patentes y Marcas de Estados Unidos (USPTO), recogiendo todas aquellas patentes concedidas entre los años 2008 y 2018 (ambos incluidos). Una vez extraída la información a analizar en cada bloque se cuenta con: 3.133.247 de patentes, 2.745.973 millones de enlaces contenidos en patentes, 2.297.366 millones de páginas web de patentes enlazadas, 17.001 paginas únicas web enlazando a patentes y 990.663 patentes únicas enlazadas desde documentos web.

Los resultados del análisis de Patent Outlink muestran como tanto la cantidad de patentes que contienen enlaces (20%), como el número de enlaces contenido en patentes (mediana 4-5) es todavía bajo, pero ha crecido significativamente durante los últimos años y se puede esperar un mayor uso en el futuro. Existe una diferencia clara en el uso de enlaces entre áreas de conocimiento (42% pertenecen a Física, especialmente Computación y Cálculos), así como por secciones dentro de los documentos, explicando los resultados obtenidos y la proyección de análisis futuros.

Los resultados del análisis de Patent Inlink identifica una cantidad considerable menor de dominios webs que enlazan a patentes (17.001 frente a 256.724), pero existen más enlaces por documento enlazante (el número de enlaces total es similar para ambos bloques de análisis). Así mismo, los datos muestran una elevada dispersión, con unos pocos dominios generando una gran cantidad de enlaces. Ambos bloques muestran la existencia de una alta relación con empresas y servicios tecnológicos, existiendo diferencias relativas a los enlaces a Universidades y Gobiernos (más enlaces en Outlink).

Los resultados muestran que el modelo de análisis propuesto permite y facilita el descubrimiento y evaluación de recursos web de calidad. Así mismo, se concluye que la cibermetría, mediante el análisis de enlaces, aporta información de interés para el análisis de los recursos web de calidad a través de los enlaces contenidos y dirigidos a documentos de patentes.

El método propuesto y validado permite de un modo eficiente, eficaz y replicable la extracción y análisis de los enlaces contenidos y dirigidos a documentos de patentes. Permite, a su vez, definir, modelar y caracterizar el Patent Link Analysis como un subgénero del Link Analysis que puede ser utilizado para la construcción de sistemas de monitorización de link intelligence, de evaluación y/o de calidad entre otros, mediante el uso de los enlaces entrantes y salientes de documentos de patentes aplicable universidades, centros de investigación, así como empresas públicas y privadas.