

Michael GERVERS - Michael MARGOLIN, *El proyecto DEEDS: Hacia un análisis del contenido de los diplomas privados ingleses de los siglos XII y XIII*, «Signo. Revista de Historia de la Cultura Escrita» 15 (2005) I.S.S.N. 1134-1165, Universidad de Alcalá, pp. 7-32.

EL PROYECTO DEEDS: HACIA UN ANÁLISIS DEL CONTENIDO DE LOS DIPLOMAS PRIVADOS INGLESES DE LOS SIGLOS XII Y XIII

Michael GERVERS, PhD,
Director, DEEDS¹ Project,
Professor of History,
University of Toronto
e-mail: 102063.2152@compuserve.com

Michael MARGOLIN, BA,
Programmer / Analyst, DEEDS Project,
University of Toronto
e-mail: m.margolin@utoronto.ca

INTRODUCCIÓN

La contribución normanda a la historia administrativa de la Inglaterra medieval es objeto de estudio frecuente dentro de la investigación académica del período. Entre los muchos cambios que siguieron a la conquista de 1066 hay uno en particular cuyas consecuencias han empañado la visión del cambio social, político y económico durante cerca de dos siglos y medio que separan la conquista del final del reinado de Eduardo I en 1307. Esta particularidad era la práctica de no incluir la fecha de emisión en los documentos que registraban la transmisión de la propiedad. Dichos documentos, conocidos como escrituras o diplomas, fueron los testimonios más ubicuos del período y se estima que tan

¹ NR. Las siglas DEEDS equivalen a: *Documents of Essex England Data Set*.

sólo de los siglos XII y XIII han sobrevivido hasta hoy en día más de un millón de originales o copias en cartularios en Inglaterra.

Durante el primer milenio existía una considerable diversidad en la Europa continental en la tendencia a incluir o excluir, dependiendo del lugar y la fecha de emisión, la fecha de referencia cuando la transmisión de la propiedad era registrada por escrito. No cabe duda que cuanto más cercanos eran los lazos existentes entre el emisor y la tradición romana, mayor era la probabilidad de que el diploma llevara fecha. Cada vez hay mayor evidencia que sugiere que en Borgoña, al igual que en Normandía, la práctica de emitir fechas descendió de forma constante desde finales del siglo X hasta mediados del siglo XI, cuando apenas aparecían. Aunque situados dentro del reino franco, estos territorios se hallaban bastante lejos del control oficial del rey francés. Esta indiferencia de la intervención administrativa real podría explicar la razón por la que mientras los diplomas de la monarquía francesa seguían llevando fecha, al menos algunos de las provincias lejanas más independientes no la llevaban. A este respecto, el destino de Inglaterra hizo que la conquista normanda tuviera lugar cuando lo hizo, ya que con Guillermo I se implantó la práctica normanda del momento de no fechar diplomas en ningún nivel de la sociedad. Los duques normandos tenían un sistema administrativo desarrollado cuando los lazos con la corona francesa eran débiles y claramente sentían indiferencia hacia las convenciones formales de escritura epistolar que los monarcas europeos habían adoptado de la cancillería papal. Apenas se preocupaban por las formas tradicionales, deseaban todo lo que no era absolutamente esencial al mensaje que deseaban transmitir, y sólo ocasionalmente incluían una fecha. Buscaban brevedad y concisión en los diplomas y llevaron este principio administrativo a Inglaterra, donde se convirtió en una antigua tradición por derecho propio. Fue sin duda esta tradición, más que el deseo consciente de prescindir del concepto de tiempo, lo que llevó al perdurable fenómeno inglés de emitir diplomas sin fecha. Mientras que en Normandía se volvía a datar el texto del diploma al tiempo que la autoridad administrativa real se extendía por la región después de 1204, en Inglaterra, y con el advenimiento de Ricardo I en 1189, cuya administración se regía más por lo establecido en la Europa continental que en Inglaterra, la fecha fue regularmente introducida en los documentos que emanaban de su cancillería real. Siendo el apego a la tradición lo que era, aquellos que redactaban diplomas en otros lugares del reino tardaron más de un siglo en adoptar dicha práctica de forma sistemática. La inclusión de fechas en las transmisiones privadas no se convirtió en costumbre hasta los primeros años del reinado de Eduardo II (1307-27).

El principal objetivo del proyecto de investigación DEEDS es desarrollar una metodología para el análisis de la “carta” medieval o diploma. Lo esencial de nuestro estudio es la posibilidad de determinar las características no conocidas de un diploma concreto mediante la comparación con un conjunto de diplomas similares cuyas características conocemos. Una vez establecida la metodología, ésta puede utilizarse para realizar estimaciones cronológicas, identificar las correspondientes falsificaciones, distinguir entre y/o identificar autorías, y facilitar un amplio abanico de materias de investigación histórica basadas en análisis cuantitativos del contenido. La principal actividad del proyecto DEEDS durante los últimos ocho años ha sido desarrollar una metodología que permita datar diplomas medievales carentes de fecha.

El método tradicional utilizado para determinar como mínimo la cronología relativa a documentos no fechados se ha realizado a través de la asociación entre los nombres aparecidos en dichos documentos y sus duplicados en fuentes fechadas. Este método es bastante efectivo cuando el nombre de un funcionario (preferiblemente de dos o más), cuyas fechas de ejercicio se conocen, aparece en el mismo documento. Sin embargo, es arriesgado en el caso de individuos a quienes no se puede vincular una clara cronología definida puesto que los homónimos, incluso cuando aparecen en grupos (tal y como lo hacen de una generación a otra en ambientes rurales), a menudo pueden referirse a personas diferentes. Incluso cuando la identificación es fiable, es imposible determinar la edad que tenía esa persona cuando se emitió un documento concreto. Se pueden realizar suposiciones erróneas por falta de pruebas contextuales fiables en el caso de individuos desconocidos y fechas dudosas asignadas por deducción a documentos en los que éstos aparecen. Es más, desde el momento en que se asigna una fecha aproximada a un documento y el trabajo es llevado a imprenta, esa fecha tiende a adquirir un valor histórico propio. Los estudiosos posteriores le otorgan el beneficio de la duda como si realmente hubiera pertenecido al documento en cuestión. De esta forma se asignan fechas a diplomas que carecen de ellas, las cuales, a su vez, son utilizadas como evidencia para datar diplomas no fechados. La posibilidad de error en este proceso es algo mayor de lo que uno desearía y probablemente aún mayor de lo que uno pudiera imaginar. En un artículo reciente, publicado póstumamente por John Horace Round, el autor señala cómo el respetado editor del *Cartulario de Colchester* se equivocó en la fecha de una diploma real, que ya tenía fecha, en más de 200 años porque desconocía que la intitulación del rey, tal y como se registraba en la cláusula de fechas, “Henricus Rex Anglie et Francie et Dominus Hibernie”, se refería a Enrique VI

en lugar de a Enrique III.² Cuando es posible cometer errores de este tipo con diplomas reales ya fechados, resultan mucho más vulnerables las fechas asignadas por la asociación de nombres propios a aquellos diplomas privados que carecen de ellas.

Los problemas que surgen al identificar erróneamente a diferentes individuos con el mismo nombre, o al mismo individuo que lleva el mismo nombre, sin olvidar al mismo individuo que tiene varios nombres, se agravan en dichos documentos cuyas listas de testigos fueron omitidas, como ocurría con frecuencia cuando eran copiados a cartularios. En el caso de los documentos en los que, por lo demás, no aparecían más de 2 ó 3 nombres desconocidos, es evidente que serían necesarios otros medios para identificar los cambios cronológicos. La paleografía no era una buena opción ya que la gran mayoría de las fuentes del período han sobrevivido sólo en copias posteriores. Por este mismo motivo, la sigilografía también resultaba insuficiente, a pesar de su importancia en aquellos casos relativamente raros en los que los originales se han preservado con sello.

En la búsqueda de una solución alternativa, DEEDS siguió el camino marcado por el historiador medieval Sir Frank Stenton hace ochenta años. Afrontando el hecho de que sólo el cinco por ciento de los documentos que editaba llevaban fecha o se podían datar de forma exacta basándose en nombres reconocidos o en sucesos citados en ellos, llegó a la conclusión, tras realizar un estudio meticuloso de sus fuentes, de que la cronología del diploma estaba inextricablemente ligada al incremento y desarrollo de las fórmulas que aparecían en ellos.³ El término “fórmula” es algo ambiguo ya que sugiere un conjunto de palabras cuyo orden no cambia a pesar del contexto. Puesto que no hay nada más lejos de la realidad en términos de la fraseología legal medieval de los siglos XII y XIII, DEEDS prefiere la expresión “estructura de palabras” a “fórmula” y lo utiliza para designar a cualquier grupo de dos o más palabras que aparecen en más de un documento.

En general, todo diploma es un documento legal oficial escrito o emitido por una institución religiosa, laica, o real y, por tanto, puede ser tratado

² J. Horace Round (revisado y completado por W.R. Powell), “St. Thomas Becket’s sisters and other studies”, *Essex Archaeology and History*, 31 (2000), p. 154-160 (p. 156).

³ F.M. Stenton, ed., *Transcripts of Charters Relating to the Gilbertine Houses of Sixle, Ormsby, Catley, Bullington, and Alvingham*, Publicaciones de la Lincoln Record Society de 1920, 18 (Horncastle, 1922), p. x-xi, xxxiii-xxxiv.

como una entidad independiente al tiempo que como objeto de modelo y análisis. Es importante reconocer que la historia refleja las circunstancias específicas que determinaron los cambios en la elaboración de textos legales y la constante adopción, formulación y adaptación de palabras y expresiones. Sin embargo, el desuso y el cambio de unas palabras por otras son fenómenos comunes a toda lengua en cualquier tiempo y están implícitos en todo este proceso de cambio. Por decirlo de forma más sencilla, los cambios en el uso de las palabras y expresiones son el reflejo más inmediato del cambio social.

En 1983, haciendo uso de lo que entonces se consideraba “nueva” metodología, el proyecto DEEDS produjo una microficha de concordancia del contenido de 957 diplomas de la *secunda camera* del cartulario inglés Hospitaller de 1442⁴. Un cuidadoso examen de la concordancia dejó indudablemente claro que estos diplomas por sí solos contenían decenas de miles de estructuras de palabras identificables, es decir, grupos de dos o más palabras en un documento que eran exactamente iguales a un grupo análogo en otro documento. La amplia variedad de fechas asociadas a las diversas estructuras llevó a la hipótesis de que más que ser predominantemente estáticas, teniendo en cuenta el cambio oracional resultante de un hecho histórico o realizado como respuesta a un estatuto real, el lenguaje del diploma medieval estaba en un constante estado de cambio. De manera global, parecía razonable que si la cronología del cambio era patente, sería posible utilizar el porcentaje de aparición y desaparición de las estructuras de palabras en el documento para determinar una fecha más o menos exacta en cualquier documento del mismo tipo. Es decir, las agrupaciones cronológicas reflejadas en las estructuras de palabras a través de la concordancia indicaban la correspondencia de estructuras como medio potencialmente viable para datar un diploma sin fecha.

Creemos (y análisis previos así lo confirman) que se pueden establecer cronologías específicas al estudiar la frecuencia de la distribución de los términos y expresiones extraídas de un diploma concreto a lo largo del tiempo. Tras tener en cuenta cualquier información adicional que pudiera hallarse en el texto, como, por ejemplo, la naturaleza del contenido, la organización de la estructura del texto, las particularidades topográficas, las referencias a personas e instituciones, etc., pueden extraerse conclusiones definitivas al combinar los resultados recogidos de los análisis anteriormente mencionados de cada uno de estos componentes.

⁴ Ver Bibliografía, 1983.

EL CORPUS DE DIPLOMAS MEDIEVALES DE DEEDS

Hasta el momento, el corpus del proyecto DEEDS incluye más de 8000 diplomas medievales ingleses de los siglos XII y XIII escritos en latín procedentes de fuentes impresas. Todos los diplomas introducidos en el corpus están fechados, bien de forma interna, bien por el editor de la colección mediante pruebas internas. La mayoría de los diplomas elegidos para nuestro corpus están datados de forma interna, normalmente con el año, el año del reinado, o ambos. En ocasiones se citan los años del pontificado, casi siempre en los diplomas episcopales. A veces un acontecimiento histórico nos proporciona la fecha. También hemos añadido diplomas no fechados que pueden ser datados con un año concreto por medio de pruebas internas, como los muchos publicados por F. E. Satler para la Oxford Record Society y que contienen los nombres de los funcionarios que sólo mantuvieron su puesto de año en año. Los diplomas con fechas determinadas representan el treinta y dos por ciento del número total de diplomas del corpus de DEEDS.

Obviamente, el número de diplomas disponibles varía según el período cronológico (Fig. 1), al igual que lo hace la exactitud de la evidencia cronológica, la cual oscila del día, mes o año exacto a varios años de la fecha exacta. Un diploma medio contiene alrededor de 200 palabras (1300 caracteres), su tamaño no está específicamente relacionado con un período de tiempo concreto ni con la naturaleza de su contenido, y rara vez son muy cortos o muy largos (Fig. 2). El noventa y dos por ciento de los diplomas del corpus están representados por dieciséis tipos y cinco de ellos, por sí solos, componen casi un ochenta por ciento del total (Fig. 3). Los cinco tipos son: donaciones (33%), acuerdos (17,5%), acuerdos finales (12,4%), confirmaciones (9,4%) y renunciaciones (8,8%).

Uno de los mayores retos a los que se enfrenta el investigador que estudia documentos publicados en lugar de fuentes manuscritas es abordar los innumerables métodos y principios editoriales ideados por cada editor. Siempre que fuera preciso se utiliza el ordenador para eliminar o evitar tales diferencias. La letra “v”, por ejemplo, siempre se lee como “u” y la “j” como “i”, las letras “c” y “t” se tipifican en ciertas combinaciones, no se hace distinción entre mayúsculas y minúsculas, y se eliminan los signos de puntuación. Son más problemáticos los casos en los que los editores y, en ocasiones, también los copistas medievales, han reducido los textos o han reemplazado lo que interpretaron como meras “frases formularias” por “etc.”. Señalamos lagunas

evidentes y mantenemos las abreviaturas, aunque no las tenemos en cuenta en materia de investigación, y siempre mantenemos “etc.”.

DOCUMENTO FUENTE (SD)⁵

El documento fuente (Fig. 4) es un documento XML compuesto por el texto original extraído de la fuente impresa al escanear el texto del mismo y convertido en formato electrónico por medio de la Identificación Óptica de Caracteres (OCR), y por la información complementaria obtenida: 1) del diploma mismo, 2) de la edición correspondiente del cartulario o del conjunto de diplomas del cual es extraído, o 3) de fuentes externas. El SD es un mero documento texto definido y escrito en XML. Puede ser fácilmente analizado sintácticamente por el ordenador o leído por una persona. Desarrollamos un programa patentado llamado “gestor de documentos” para facilitar la creación y mantenimiento de los SD, los cuales son almacenados en la base de datos del proyecto DEEDS, permitiendo así poder poner índice al contenido y tener acceso propio a la información a través de Internet.

El propósito de introducir los SD es condensar la información y establecer enlaces lógicos entre el texto del diploma y la información complementaria relativa al diploma en forma compartida y portátil.

El uso de XML ha hecho posible la implementación de un almacén de datos textuales bastante flexible, así como de información auxiliar relacionada. Cada SD y sus datos claves (una combinación del código arbitrario para el cartulario en el que fue publicado y del número del diploma dentro del cartulario) ocupan una sola fila de la tabla de la base de datos. El uso de la base de datos permite acceder al índice y a las prestaciones exploratorias de los sistemas modernos de gestión de bases de datos y garantiza la coherencia e integridad de la información. Asimismo, hay que resaltar que la jerarquía usada en los SD también puede ser implementada mediante una base de datos de solución “pura”, la cual requeriría crear múltiples tablas de bases de datos y organizar enlaces complejos entre ellos. La programación para la administración, mantenimiento y aplicación de una base de datos de solución “pura” implicaría mucho más esfuerzo que una basada en XML. Por consiguiente, el uso de XML fue una elección lógica para el desarrollo de los

⁵ NR. SD equivale a *Source Document*.

documentos fuente del corpus de DEEDS. Una ventaja añadida es la disponibilidad generalizada del análisis sintáctico XML, lo que ayuda a reducir de forma significativa la complejidad de los programas de aplicación.

El SD está compuesto por cuatro elementos principales: **contenido**, **datos**, **notas** y **marcas**, siendo los tres últimos factores los que contienen información complementaria relativa al diploma en cuestión.

- “*Contenido*”. Este elemento almacena el texto original del diploma tal y como aparece en la fuente impresa. Carece de nodos sucesores, por lo que preserva totalmente la integridad del texto del diploma.
- “*Datos*”. Este elemento almacena la información relativa a las propiedades atribuidas a un diploma concreto, las cuales nos son útiles para poder asignar una fecha. Siempre que sea viable adscribimos la siguiente información a cada entrada del corpus: 1) el tipo de dato asociado al diploma (asignado por un editor o establecido de forma interna, y si es de forma interna, por año, año de “reinado”, “festividad”, o acontecimiento), 2) el tipo de transacción legal que registra (donaciones, acuerdos, acuerdos finales, confirmaciones, etc.), 3) si la fuente es un documento original o una copia, y si es copia, la naturaleza y fecha de la fuente, 4) el lugar de emisión, 5) la parroquia y el condado correspondientes, 6) la naturaleza general (religiosa o laica) del emisor y del receptor, 7) la naturaleza específica del emisor y del receptor (orden religiosa o institución, o el nombre, título y ocupación de la persona), 8) si se trata de un convento, el nombre y la fecha de su fundación y de su disolución.
- El elemento “*Notas*” almacena información editorial proveniente de fuentes externas relacionadas con el diploma en cuestión y está dividido por temas.
- El elemento “*Marcas*” ejecuta la marca “virtual” mediante el almacenamiento de direcciones de todas las etiquetas de las marcas. La marca virtual fue introducida para establecer un sistema indirecto de marcas textuales por medio de etiquetas XML. Dichas etiquetas trazan un mapa en el texto del diploma en lugar de estar insertadas físicamente en él.

GESTOR DE DOCUMENTOS (DM)⁶

El gestor de documentos (Fig. 5) es un programa de software diseñado por el proyecto DEEDS para facilitar la anexión, modificación y supresión de diplomas en el corpus de DEEDS. Se ejecuta desde Windows y está compuesto por campos de datos agrupados en una serie de pantallas de introducción de datos. La operación “Guardar” se puede iniciar tan pronto como el dato correspondiente haya sido introducido en los campos de forma y resultará en la creación del formato XML del SD. Este formato reemplazaría, o sería añadido, a la base de datos que almacena el corpus de DEEDS. Cualquier SD que haya sido creado o modificado concuerda completamente con la Definición de Tipo de Documento del SD (DTD) (Fig. 6) y con las reglas de validación pertinentes. Este programa elimina todos los errores ortográficos y erratas en las etiquetas y propiedades XML. DM también permite el oportuno acceso de clasificación para el diploma mismo y para datos complementarios.

Además de las funciones anteriormente mencionadas, el programa permite total acceso a datos auxiliares como son nuestras listas de cartularios, instituciones eclesiásticas, y localizadores topográficos (país, condado o parroquia). Todas las listas son documentos XML almacenados en la base de datos y creados y manejados mediante el DM.

MARCA VIRTUAL

La marca virtual representa esa parte de la tecnología exclusivamente desarrollada por el proyecto DEEDS para realizar análisis cuantitativos en función del contenido de los diplomas medievales. El contenido del diploma puede ser descrito y estructurado en distintos términos según el objetivo acometido por la investigación. La clasificación puede hacerse por medio de fórmulas (es decir, partes diplomáticas), de funciones de contenido (es decir, los sujetos de la acción legal, referencias a la información cronológica, nombres de personas, descripciones de propiedades, etc.) o de una frase literal.

El elemento *Marcas* es implementado por medio de un conjunto de “capas”. Cada capa enlaza estrictamente con una categoría específica. La capa misma es una agrupación de ítems, y cada una está provista de la dirección

⁶ NR. DM equivale a *Document manager*.

exacta de las etiquetas de las marcas asociadas a una expresión concreta hallada en el texto. El ítem almacena el lugar en el que comienza y finaliza una frase concreta relativa al principio del texto. No existe límite en cuanto a la cantidad de capas de marcas o ítems que puedan ser introducidas. En la práctica, el gestor de documentos crea, mantiene y visualiza la marca virtual.

Con la marca convencional las etiquetas de las marcas han de estar insertadas dentro del texto. La marca virtual es una alternativa genérica a la misma y sólo requiere que la dirección de las etiquetas de las marcas se recojan y guarden como nodos sucesores del documento “árbol” XML al tiempo que el texto del diploma permanece en su estado original. He aquí algunas de las ventajas del uso de la marca virtual:

- No impone reglas obligatorias en cuanto a la configuración del diploma y, por tanto, puede ser utilizado con cualquier tipo de diploma, incluidos aquellos compuestos por múltiples documentos (como ocurre muchas veces con las “confirmaciones”).
- Proporciona la posibilidad de crear y conservar un número ilimitado de clasificaciones independientes para un diploma concreto.
- Los nodos de las marcas de las clasificaciones independientes pueden solaparse, posibilidad que queda totalmente excluida en el caso de la marcación XML convencional.
- Proporciona acceso rápido y fiable a la información marcada y al texto de un diploma sin necesidad de recurrir a una exhausta navegación por los recursos de todas las ramas del árbol de búsqueda de XML.

La presencia de la marca concluyente, así como de cualquier otra información relativa a los documentos del corpus de DEEDS, nos permite crear, si así lo queremos, versiones filtradas del corpus (llamadas “proyecciones”). Una proyección del corpus podría incluir sólo los documentos que se adecuan a los criterios de algunas propiedades concretas y, respecto de ellas, el texto original podría reemplazarse por el contenido de los segmentos marcados.

Los editores no suelen coincidir en su apreciación sobre los tipos de documentos. Como consecuencia, DEEDS incluye su propia denominación mediante una propuesta estandarizada basada en una clasificación por palabras

(las DONACIONES incluyen siempre formas de dedi y/o concessi, a veces junto con liveravi, tradidi, dimisi, assignavi, contuli, legavi, y donavi; las CONFIRMACIONES incorporan formas de confirmavi y concessi et ratam habui; las RENUNCIAS, formas de quietum clamavi, a veces en conjunción con remisi, concessi, recognovi, resignavi, relaxavi, confirmavi, y dimisi; los CONTRATOS DE ARRENDAMIENTO, formas de dedi, dimisi, concessi, y tradidi; las VENTAS, formas de vendidi, quietum clamavi, concessi, forisaffidavi, y confirmavi; y los ACUERDOS, formas de convenit, o conventio y compositio). Para confirmar la totalidad de palabras y frases positivas relacionadas con los distintos tipos de documentos, se marcan dichas palabras y frases y se establecen listas de estructuras para cada tipo.

Además de para las palabras de disposición, las marcas sirven para distinguir otras divisiones y subdivisiones diplomáticas de un documento. Por tanto, la composición del diploma estándar de PROTOCOLO, CUERPO y ESCATOCOLO se divide más aún, hasta que las partes del PROTOCOLO puedan ser identificadas como invocatio, intitulatio, y salutatio; del CUERPO con las divisiones específicas de los tipos de documento (generalmente promulgatio o notificación, narratio o antecedentes, dispositio o palabras de acción, y las clausulae condicionales); y de las secciones del ESCATOCOLO con la cláusula del sello, la lista de testigos, el lugar de emisión y la fecha.

Para el cálculo final utilizamos proyecciones de los diplomas originales en el análisis del contenido cuantitativo y estimaciones cronológicas para eliminar la influencia negativa de los errores de imprenta, ortográficos, y de inconsistencias en el formato. La proyección del corpus está compuesta por documentos en formato HTML procedentes de los SD formateados en XML mediante el uso de la Hoja de Lenguaje de Estilo Extensible (XSL) y las rutinas de programación patentadas.

ANÁLISIS CUANTITATIVO DEL CONTENIDO Y ESTIMACIONES CRONOLÓGICAS

Los análisis y cálculos cuantitativos basados en el estudio del vocabulario implican extraer del texto del diploma combinaciones de dos o más palabras adyacentes llamadas **estructuras de palabras** en orden consecutivo y examinar las ocurrencias de cada estructura en el corpus. Seguidamente se analizan para generar cálculos sobre las propiedades (atributos) desconocidas de un diploma específico.

Analizamos la distribución de frecuencias en las ocurrencias de las estructuras de palabras (análisis de frecuencias) para realizar un cálculo sobre los atributos no conocidos de un diploma al compararlos con otros diplomas del corpus. El mismo método, llevado a cabo con proyecciones específicas del corpus, nos proporcionaría resultados más definidos y concretos para los parámetros seleccionados.

Los análisis de frecuencia comienzan con la extracción de conjuntos de dos o más palabras consecutivas del texto de un diploma en concreto. El programa comienza desde la primera palabra y va bajando hasta el final del texto extrayendo palabra por palabra hasta llegar a un número fijo de ellas. La generación de conjuntos termina cuando el tamaño de la estructura de palabras alcanza el tamaño del texto del diploma o cuando no concurren más estructuras de dicho conjunto en ninguna otra parte del corpus.

Una vez completada la creación del conjunto, el programa comienza a buscar en el corpus las ocurrencias de cada estructura de palabras extraída. La distribución de ocurrencias resultantes es tratada como un conjunto independiente y se agrupa junto con las características de la estructura de palabras específica. Puede haber una amplia variación en el número y extensión de las estructuras de palabras en cada conjunto, así como en las concurrencias (coincidencias exactas) con otras estructuras del corpus. Tras procesar todas las estructuras de palabras del conjunto, el sistema computa el cálculo y el supuesto margen de error para cada una de ellas. En las etapas siguientes, se prorratean, evalúan y agrupan los cálculos de las estructuras de palabras dentro del cómputo del nivel fijado de las mismas para, una vez más, computar a dicho nivel el **cálculo de la tendencia central** y su margen de error. También se prorratean y evalúan otros cálculos de niveles fijados para fusionarse al nivel del diploma. El cómputo de un cálculo y del margen de error a nivel del diploma concluye el análisis de frecuencias para la **proyección** específica. Los factores utilizados en los cómputos de la evaluación siempre se calculan de forma dinámica y dependen de las características concretas de la **distribución** específica. La meta de la evaluación dinámica es reducir la deformación causada por cálculos naturales en el tamaño de las estructuras de palabras y en las características entre las capas de todos los niveles. Finalmente, se puede computar el cálculo definitivo de las propiedades no conocidas con la fusión de los cálculos realizados por el procesamiento de múltiples proyecciones dentro del mismo diploma.

Nuestra metodología para la estimación cronológica se basa en la

premisa de que cualquiera puede determinar juicios exactos sobre la cronología de un diploma al compararlo con otros diplomas del corpus. Por lo tanto, una coincidencia entre los atributos conocidos de un diploma concreto y los correspondientes atributos en el grupo de diplomas nos permite asignar cálculos de otras propiedades del grupo a dicho diploma. Cualquier singularidad del diploma que pueda ser deducida directa o indirectamente del texto, examinada de forma fiable y con coincidencias de esas mismas características en otros diplomas del corpus puede, asimismo, utilizarse como atributo. Puesto que nuestro corpus está compuesto de documentos independientes, el resultado de coincidencias en los atributos puede no ser sino un conjunto de valores de los mismos. El cálculo del valor del atributo asociado a cada conjunto se obtiene por medio de técnicas estadísticas que miden la distribución de la tendencia central junto con la estimación del supuesto margen de error. En las siguientes etapas del procesamiento se agrupan los resultados obtenidos por cualquier conjunto similar y cada grupo es tratado de nuevo como un conjunto individual. Finalmente, se computa el cálculo definitivo del atributo en concreto mediante el procesamiento del último conjunto generado, el cual, a su vez, representa el resultado de múltiples fusiones de conjuntos producidas durante las etapas intermedias del procesamiento.

La figura 7 muestra la unidad de salida del sistema utilizando todo el corpus (parte superior) y el tipo de contenido basado en la proyección del corpus (parte inferior).

RESUMEN DE INTENCIONES DE LA INVESTIGACIÓN Y RESULTADOS

El principal objetivo del proyecto de investigación DEEDS es desarrollar una metodología para el análisis de la “carta” medieval o diploma. Una vez establecida, dicho modelo puede utilizarse: para realizar estimaciones cronológicas, identificar falsificaciones, distinguir entre autores o identificar autorías, y llevar a cabo distintos temas de investigación histórica basados en análisis del contenido.

Análisis previos confirman que se pueden atribuir cronologías específicas a diplomas no fechados al comparar la distribución de frecuencias de los términos y expresiones extraídos de un diploma no fechado concreto con un conjunto de documentos existente cuyas características tienen valores

reconocidos. La estimación de una cronología correcta depende también de la identificación de las falsificaciones, puesto que el propósito de la falsificación es cambiar la intención del pasado. DEEDS se propone identificar falsificaciones mediante el análisis del contenido, especialmente de palabras dentro del contexto, su ubicación en el texto, y su frecuencia de uso. Señalaremos las secciones que contienen palabras, frases o conceptos que parezcan encontrarse fuera de su contexto cronológico usual, ya que posiblemente hayan sido introducidas por un “improbo” o falsificador. Son pocas las falsificaciones de un documento completo, lo más usual es la falsificación de algunas de las partes que lo componen.

Los investigadores de DEEDS anticipan que será posible mejorar los métodos existentes de certificaciones de autorías al añadir componentes cronológicos, espaciales, léxicos y estructurales a estudios de este tipo. El resultado en el caso de los diplomas será la identificación del trabajo de cada escribano, el abanico de fechas asociado a su actividad y el modelo o modelos que utilizaba.

El proyecto DEEDS utiliza XML para el procesamiento y análisis de diplomas medievales escritos en latín. Creemos que XML es un medio fiable y universal que nos permite combinar y manipular el material textual original y los datos atribuidos al mismo según nos convenga. El progreso actual de la tecnología del procesamiento XML proporciona a los investigadores tecnologías eficaces, como los esquemas XML, como alternativas efectivas al tradicional DTD y las transferencias de datos entre XML y la base de datos en la que es aplicado como característica estándar de todos los sistemas de gestión de base de datos más importantes.

APÉNDICE 1

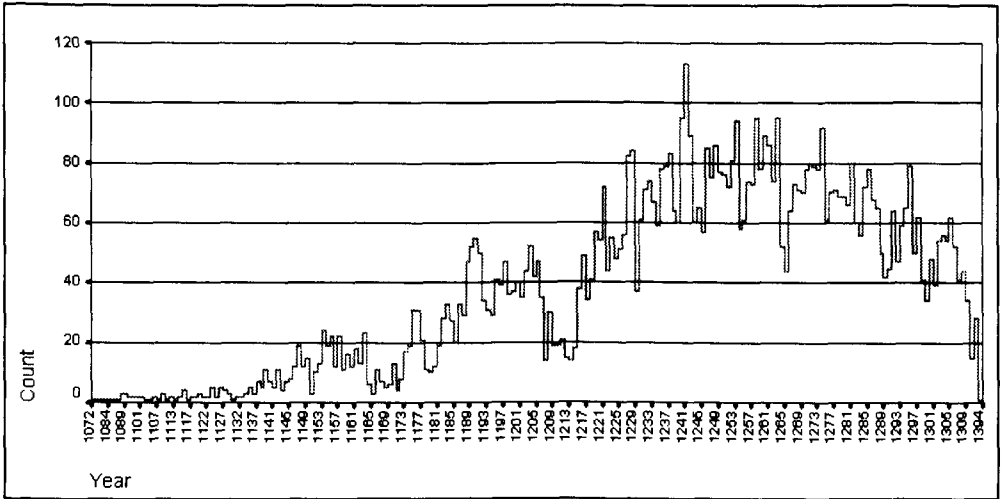


Figura 1. Número de diplomas en el corpus de DEEDS a lo largo del tiempo

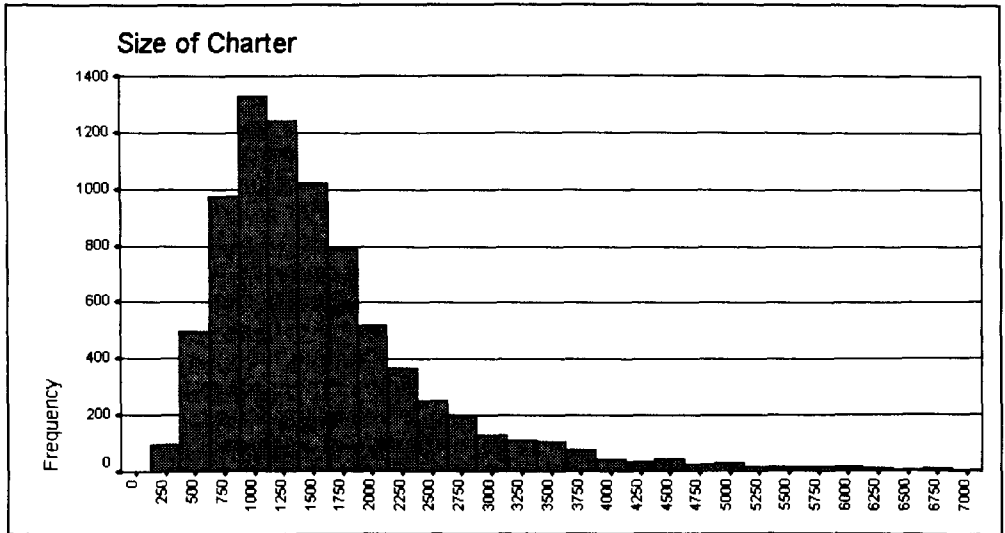


Figura 2. Variación en el tamaño de los diplomas del corpus de DEEDS

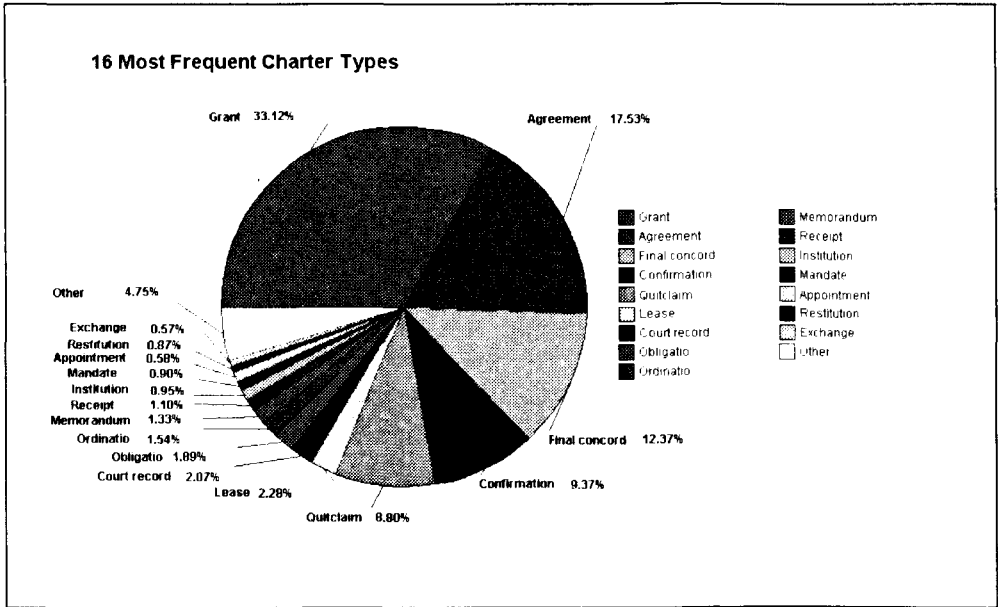


Figura 3. Tipos de diplomas más frecuentes

<sourceDocument dnum="00880049" created="2004-03-08" cartulary="Knights of St John secunda camera">

<content>Sciunt presentes et futuri quod ego Walterus carpentarius filius Henrici molendinarii de Hengham Sibile concessi dedi et hac presenti carta mea confirmavi Rogero de Othulvesho clerico pro servicio suo et pro viginti solidis sterlingorum quos michi dedit in gersumam unam acram terre et tres rodas cum pertinentiis in dicta villa de Hengham Sibile iacentes ex opposito mesuagii Iordani carpentarii inter terram Edithe Smuches et terram que Galfridi Berde fuit et pasturam duarum acrarum terre iacencium propinquius mesuagio Godefridi coci apud le Motstowe ibidem singulis annis cum segetes a mote fuerint scilicet de terra Iohannis filii Alani habend et tenend de me et heredibus meis predicto Rogero et heredibus suis et suis assignatis et quibuscumque vel quandocumque predictam terram et pasturam cum suis pertinentiis dare vendere et assignare voluerint et eorum heredibus libere quiete bene et in pace iure in feodo hereditarie et in perpetuum reddendo inde annuatim michi et heredibus meis vel meis assignatis ille et heredes sui vel sui assignati et eorum heredes 7d ad duos terminos anni videlicet ad Pascha 3d [et] obolum et ad festum sancti Michaelis 3d et obolum pro omnibus serviciis consuetudinibus exactionibus sectis curiarum et universis secularibus demandis salvo tamen servicio domini regis scilicet ad scutagium quando evenerit ad plus et ad minus unum obolum Et ego predictus Walterus et heredes mei vel mei assignati warantizabimus defendemus et acquietabimus totam

predictam terram et pasturam cum omnibus suis pertinentiis predicto Rogero et heredibus suis vel suis assignatis et eorum heredibus per predictum liberum servicium contra omnes gentes in perpetuum In cuius rei testimonium huic presenti scripto sigillum meum apposui Hiis testibus Adam Gravassall Iohanne fratre suo Willelmo filio Willelmi Willelmo le Bloy Martino crok(er) Roberto hostage Roberto Fovel Huberto clerico Huberto crok(er) Martino le Boll Waltero de Haukwode Iohanne de le Hill clerico et aliis Datum apud Hengham Castrum die sancti Dionisii anno regni regis Henrici filii regis Iohannis quadragesimo sexto</content>

<data>

<general>

<origin>Lay</origin>

<transactionType>GRANT</transactionType>

<copySource>CartularyCopy</copySource>

<status ref="">Primary</status>

</general>

<dateInfo>

<singleDate>1262-10-09</singleDate>

<lowDate>0000-00-00</lowDate>

<highDate>0000-00-00</highDate>

<main>Internal</main>

<details>Feast</details>

<details>Regnal</details>

</dateInfo>

<locationInfo>

<location type="issue d">

<locSet>

<place>HEDINGHAM, CASTLE</place>

<co unty>ESSEX</co unty>

<co untry>ENGLAND</co untry>

</locSet>

</location>

<location type="property">

<locSet>

<place>HEDINGHAM, SIBLE</place>

<co unty>ESSEX</co unty>

<co untry>ENGLAND</co untry>

</locSet>

</location>

</locationInfo>

<parties type="givers">

<person type="LAY">

<pname type="Individual" role="" link="">Walter the Carpenter son of Henry the miller of Hedingham, Sible</pname>

</person>

</parties>

<parties type="recievers">

<person type="LAY">

<pname type="Individual" role="" link="">Roger of Othulvesho,

```

clerk</pname>
  </person>
</parties>
</data>
<notes/>
<markup>
  <layer type="numeral">
    <item start="1070" end="1072"/>
    <item start="1115" end="1129"/>
    <item start="1160" end="1172"/>
    <item start="1366" end="1377"/>
  </layer>
  <layer type="name">
    <item start="37" end="104"/>
    <item start="156" end="184"/>
    <item start="378" end="397"/>
    <item start="411" end="425"/>
    <item start="440" end="454"/>
    <item start="524" end="538"/>
    <item start="622" end="642"/>
    <item start="693" end="699"/>
    <item start="1395" end="1403"/>
    <item start="1554" end="1560"/>
    <item start="1766" end="1781"/>
    <item start="1782" end="1800"/>
    <item start="1801" end="1824"/>
    <item start="1825" end="1841"/>
    <item start="1842" end="1858"/>
    <item start="1859" end="1874"/>
    <item start="1875" end="1888"/>
    <item start="1889" end="1904"/>
    <item start="1905" end="1921"/>
    <item start="1922" end="1937"/>
    <item start="1938" end="1958"/>
    <item start="1959" end="1985"/>
  </layer>
  <layer type="location">
    <item start="333" end="347"/>
    <item start="544" end="555"/>
    <item start="2006" end="2021"/>
  </layer>
  <layer type="property">
    <item start="135" end="140"/>
    <item start="189" end="197"/>
    <item start="209" end="237"/>
    <item start="258" end="266"/>
    <item start="267" end="283"/>
    <item start="287" end="297"/>
  </layer>

```

```

<item start="324" end="347"/>
<item start="369" end="377"/>
<item start="369" end="397"/>
<item start="404" end="425"/>
<item start="429" end="454"/>
<item start="463" end="492"/>
<item start="515" end="538"/>
<item start="582" end="589"/>
<item start="1070" end="1072"/>
<item start="1108" end="1114"/>
<item start="1115" end="1129"/>
<item start="1136" end="1159"/>
<item start="1160" end="1172"/>
<item start="1185" end="1230"/>
<item start="1286" end="1307"/>
<item start="1320" end="1329"/>
<item start="1371" end="1377"/>
<item start="1696" end="1707"/>
<item start="1730" end="1738"/>
</layer>
<layer type="date reference">
  <item start="2022" end="2106"/>
</layer>
</markup>
</sourceDocument>

```

Figura 4. Documento fuente (SD)

Document Manager

Documents | Geography | Ecclesiastic Institutions | Cartul. 4 | 00880049

Transaction Type: GRANT

Document Source: Cartulary Copy Manuscript

Origin: Ley

Archive Status: Primary Secondary Embedded Ignore

Reference:

View Markup: property

Reset Text Color Remove Markup

Sciant presentes et futuri quod ego **Walterus carpentarius filius Henrici molendinarii de Hengham Sibille** concessi dedi et hac presenti carta mea confirmavi **Rogero de Othulvesho clerico** pro servitio suo et pro viginti solidis sterlingorum quos michi dedit in **gensumam unam** acram terra et **tres rodas** cum pertinentis in dicta villa de **Hengham Sibille** vacentes ex opposito **mesuagio lurdani carpentarii** inter terram **Edithe Smuches** et terram que **Galfridi Berde** fuit et **pasturam duarum acrarum** terrae facencium propinquius **mesuagio Godetrudi** coeli apud **le Mottowe** ibidem singulis annis cum **segetes** amote fuerint scilicet de terra **Iohannis filii Alani** habent et tenent de me et heredibus meis predicto **Rogero** et heredibus suis et suis assignatis et quibuscumque vel quodcumque predictam terram et pasturam cum suis pertinentis dare vendere et assignare voluerint et eorum heredibus libere quiete bene et in pace iure in feodo hereditarie et in perpetuum reddendo inde annuatim michi et heredibus meis vel meis assignatis ille et heredes sui vel sui assignati et eorum heredes **7d** ad duos terminos anni videelicet ad **Pascha 3d** [et] **obolium** et ad festum **sancti Michaelis 3d** et **obolium** pro omnibus **servitiis consuetudinibus exactionibus sectis** cuniarum et unversis secularibus demandis salvo tamen **servitio domini regis** scilicet ad **scutagium** quando evenent ad plus et ad minus **unum obolum** Et ego predictus **Walterus** et heredes mei vel mei assignati warrantizabimus defendemus et acquietabimus totam predictam terram et pasturam cum omnibus suis pertinentis predicto **Rogero** et heredibus suis vel suis assignatis et eorum heredibus per predictum liberum servitium contra omnes gentes in perpetuum in cuius rei testimonium huic presenti scripto sigillum meum apposui Hic testibus **Adam Gravesall** Iohanne fratre suo **Willelmo filio Willelmi Willelmo le Bloy Martino crok(er) Roberto hostage Roberto Fovel Huberto clerico Huberto crok(er) Martino le Boll Waltero de Heuskwode Iohanne de le Hill** clerico et aliis Datum apud Hengham Castrum

Figura 5. Gestor de documentos

```

<?xml version="1.0" encoding="UTF-8"?>
<!-- edited with XML Spy v3.5 NT (http://www.xmlspy.com) by Michael Margolin
(University of Toronto/DEEDS Project) -->
<!ELEMENT sourceDocument (content, data, notes, markup?)>
<!ATTLIST sourceDocument
  dnum CDATA #REQUIRED
  cartulary CDATA #REQUIRED
  created CDATA #IMPLIED
>
<!-- text section -->
<!ELEMENT content (#PCDATA)>
<!-- data section -->
<!ELEMENT data (general, dateInfo, locationInfo, parties*)>
<!-- general -->
<!ELEMENT general (origin, transactionType+, status, affiliation?, copySource?)>
<!ELEMENT origin (#PCDATA)>
<!ELEMENT transactionType (#PCDATA)>
<!ELEMENT status (#PCDATA)>
<!ATTLIST status
  ref CDATA #IMPLIED
>
<!ELEMENT copySource (#PCDATA)>
<!ELEMENT affiliation (#PCDATA)>
<!-- date -->
<!ELEMENT dateInfo (singleDate, circa?, lowDate, highDate, dateSource?)>
<!ELEMENT singleDate (#PCDATA)>
<!ELEMENT lowDate (#PCDATA)>
<!ELEMENT highDate (#PCDATA)>
<!ELEMENT circa (#PCDATA)>
<!ELEMENT dateSource (main, details*)>
<!ELEMENT main (#PCDATA)>
<!ELEMENT details (#PCDATA)>
<!-- location -->
<!ELEMENT locationInfo (location*)>
<!ELEMENT location (locSet)>
<!ATTLIST location
  type CDATA #REQUIRED
>
<!ELEMENT locSet (place, county, country)>
<!ELEMENT place (#PCDATA)>
<!ELEMENT county (#PCDATA)>
<!ELEMENT country (#PCDATA)>
<!-- participants -->
<!ELEMENT parties (person*)>
<!ATTLIST parties
  type CDATA #REQUIRED
>

```

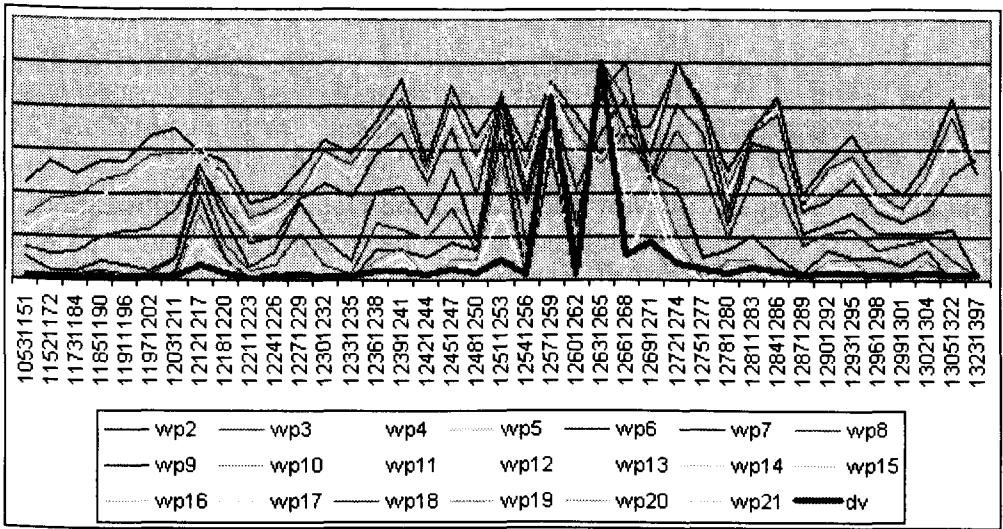
```

<!ELEMENT person (pname, title*)>
<!ATTLIST person
  type CDATA #REQUIRED
>
<!ELEMENT title (#PCDATA)>
<!ATTLIST title
  institution CDATA #REQUIRED
>
<!ELEMENT pname (#PCDATA)>
<!ATTLIST pname
  type CDATA #REQUIRED
  role CDATA #REQUIRED
  link CDATA #REQUIRED
>
<!-- notes -->
<!ELEMENT notes (note*)>
<!ELEMENT note (#PCDATA)>
<!ATTLIST note
  type CDATA #REQUIRED
>
<!-- markup -->
<!ELEMENT markup (layer*)>
<!ELEMENT layer (item*)>
<!ATTLIST layer
  type CDATA #REQUIRED
>
<!ELEMENT item EMPTY>
<!ATTLIST item
  start CDATA #REQUIRED
  end CDATA #REQUIRED
  type CDATA #IMPLIED
  num CDATA #IMPLIED
  date CDATA #IMPLIED
  datee CDATA #IMPLIED
>

```

Figura 6. Fichero de definición del documento (DTD)

St John's Hospital, Oxford, diploma #321, fecha: 1265-66
 Totalidad del corpus



Proyección del corpus basado en tipos de contenido

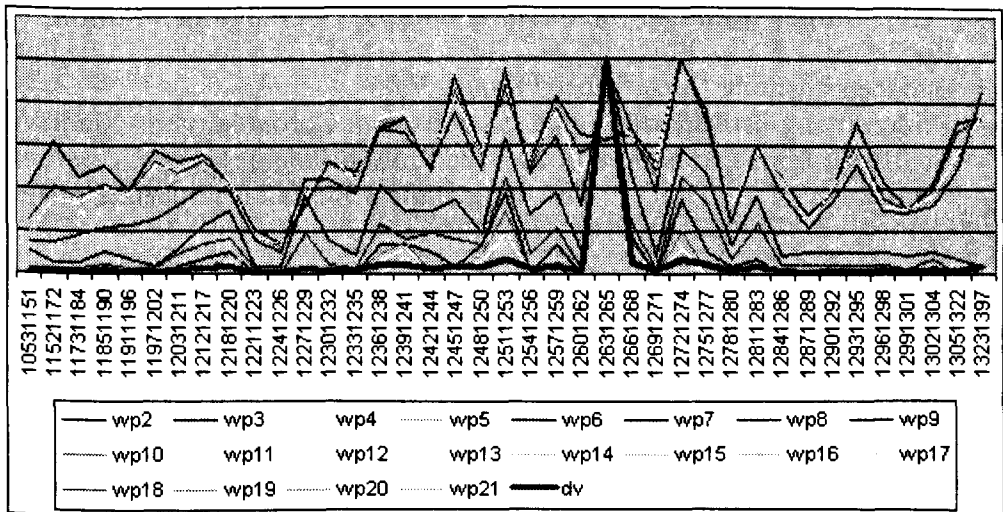


Figura 7. Evaluaciones Cronológicas

REFERENCIAS BIBLIOGRÁFICAS ADICIONALES

- CHIBNALL, Marjorie, "Dating the Charters of the Smaller Religious Houses in Suffolk in the Twelfth and Thirteenth Centuries", in GERVERS, *Dating Undated Medieval Charters*, pp. 51-59.
- DECLERCQ, Georges, "A New Method for the Dating and Identification of Forgeries? The DEEDS Methodology Applied to a Forged Charter of Count Robert I of Flanders for St. Peter's Abbey, Ghent", in GERVERS, *Dating Undated Medieval Charters*, pp. 123-136.
- GERVERS, Michael, "Changing forms of Hospitaller address in English private charters of the twelfth and thirteenth centuries", in *The Crusades and the Military Orders: Expanding the Frontiers of Medieval Latin Christianity*, ed. Zsolt HUNYADI and József LASZLOVSKY, Budapest: CEU, 2001, pp. 395-405
- "The Dating of Medieval English Private Charters of the Twelfth and Thirteenth Centuries", in *A Distinct Voice: Medieval Studies in Honor of Leonard E. Boyle, O.P.*, ed. Jacqueline BROWN & William P. STONEMAN, Notre Dame (Indiana), 1997, pp. 455-504.
- , ed., *Dating Undated Medieval Charters*, Woodbridge, Suffolk & Rochester, NY: Boydell & Brewer, and Budapest: Collegium Budapest/Institute for Advanced Study, 2000.
- , "Identifying Irregularities and Establishing Chronology in Medieval Charters", in K.S.B. KEATS-ROHAN, ed., *Resourcing Sources*, Occasional Publication of the Unit for Prosopographical Research, Linacre College, Oxford, vol. 7, 2002, pp. 164-78.
- GURNEY, Penelope J., and Lyman W. GURNEY "Authorship Attribution: A Computer-Based Approach to a Literary Crux in Late Roman Historiography", in *TEXT Technology*, Volume 10/1 (Winter 2001), pp. 87-105.
- , "Enhanced content analysis of inflected languages through a system of computer-assisted lemmatization", *Conference Abstracts of the 1994 Joint International Conference of the ALL and ACH*, Paris: The Sorbonne, pp. 93-94.
- HUNYADI, Zsolt, "The Identification of a Forgery. Regularities and

Irregularities in the Formulae of the Charters issued by the Székesfehérvár convent of the Knights of St. John of Jerusalem (1243-1353)", in GERVERS, *Dating Undated Medieval Charters*, pp. 137-49.

JUHÁSZ-ORMSBY, Agnes, "Changing Legal Terminology in Dated Private Documents in England in the Twelfth and Thirteenth Centuries. A Case Study: Quitclaims", in K.S.B. KEATS-ROHAN, ed., *Resourcing Sources*, Occasional Publication of the Unit for Prosopographical Research, Linacre College, Oxford, vol. 7, 2002, pp. 195-207.

SHARPE, Richard, "Vocabulary, Word Formation and Lexicography", in: F.A.C. MANTELLO & A.G. RIGG, eds., *Medieval Latin: An Introduction and Bibliographical Guide*, Washington, D.C., 1996, pp. 93-105; "Charters, Deeds and Diplomats", *ibid.*, pp. 230-40.

SPENCER, Amanda, "Dating Charters Using Textual Evidence", in K.S.B. KEATS-ROHAN, ed., *Resourcing Sources*, Occasional Publication of the Unit for Prosopographical Research, Linacre College, Oxford, vol. 7, 2002, pp. 179-94.

XML: W3C, REC-xml-1998-02-10 "Extensible Markup Language (XML) 1.0", <http://www.w3.org/TR/REC-xml>

RESUMEN

El presente artículo trata del funcionamiento del programa informático denominado DEEDS. Su objetivo principal es la datación de diplomas que carecen de elementos cronológicos mediante la comparación de su formulario, o de grupos de palabras, con una base de datos que contiene los textos de innumerables documentos ingleses de los siglos XII y XIII.

ABSTRACT

This article deals with the working of the computer programme called DEEDS. Its main aim is dating diplomas which lack chronological elements by means of comparing its form, or groups of words, using a database that contains the texts of countless 12th and 13th century English documents.

ZUSAMMENFASSUNG

Der vorliegende Artikel handelt über die Funktionsweise des Informatik-

Programms DEEDS. Sein Hauptziel ist die Datierung von Diplomen, die keine chronologischen Elemente aufweisen, über den Vergleich seines Formulars oder von Wortgruppen mit einer Datenbank, die Texte unzähliger englischer Dokumente aus dem 12. und 13. Jahrhundert enthält.

RIASSUNTO

Il presente articolo si occupa del funzionamento del programma informatico denominato DEEDS. L'obbiettivo principale di tale programma è di riuscire a datare diplomi mancanti di elementi cronologici grazie al confronto del loro formulario o tra gruppi di parole, per mezzo di una base di dati che contenga i testi di un numero indefinito di documenti inglesi.

RÉSUMÉ

Le présent article traite du fonctionnement du programme informatique surnommé DEEDS. Son principal objectif est la datation de diplômes manquant d'éléments chronologiques, grâce à la comparaison de leur formulaire, ou des groupes de mots, au moyen d'une banque de données rassemblant les textes d'un nombre considérable de documents anglais des XIIème et XIIIème siècles.