

# Visualización de datos provenientes del análisis de textos técnicos de la antigüedad clásica

Autor: **Sven Bühler Olivé**

[sbuhler@uoc.edu](mailto:sbuhler@uoc.edu)

Director: **Ramon Masià Fornos**

[rmasia@uoc.edu](mailto:rmasia@uoc.edu)

**Universitat Oberta de Catalunya**

**14 de enero de 2015**

## **Resumen**

*Se conoce como Digital Humanities al área de investigación en la que se integran las humanidades y la informática. Uno de los campos de estudio de las Humanidades digitales es el análisis de datos culturales a gran escala, entre los que se encuentra una gran cantidad de textos. Este trabajo se centrará en los textos técnicos escritos en griego antiguo (principalmente matemáticos y astronómicos) que resultan interesantes debido a su gran extensión (unos 2,5 millones de tokens) y léxico extremadamente reducido (menos de 3000 palabras diferentes, o lemas, en total). El objetivo de este proyecto es crear una aplicación web mediante herramientas de software libre que permita representar gráficamente información cuantitativa (tiempo, número de palabras / caracteres, etc.) y cualitativa (género literario, estilo, temática, etc.) correspondiente a las obras y los autores almacenados en un corpus de la antigüedad que abarca obras desde el año 320 a. C. hasta el año 700 d. C.*

## **Palabras clave**

d3.js, digital humanities, matemática griega, software libre, visualización de datos

# Índice de contenidos

|       |  |    |
|-------|--|----|
| 1     | Introducción.....  | 3  |
| 1.1   | Objetivos.....   | 3  |
| 1.2   | Requisitos.....  | 3  |
| 2     | Estado del arte.....                                       | 4  |
| 2.1   | Introducción.....  | 4  |
| 2.2   | Inicios y evolución de la lingüística computacional.....   | 4  |
| 2.3   | Herramientas de visualización.....                         | 6  |
| 3     | Descripción y tratamiento del corpus.....                  | 14 |
| 3.1   | Introducción.....  | 14 |
| 3.2   | Origen del corpus de obras de la antigüedad.....           | 14 |
| 3.3   | Clasificación de las obras del corpus.....                 | 15 |
| 3.4   | Tratamiento de los datos del corpus.....                   | 17 |
| 3.5   | Implementación de las funciones de probabilidad.....       | 18 |
| 4     | Resultados.....  | 21 |
| 4.1   | Timeline.....  | 21 |
| 4.1.1 | Selección del período de tiempo.....                       | 21 |
| 4.1.2 | Selección de las categorías y del valor representado.....  | 22 |
| 4.1.3 | Selección de datos, períodos, autores, obras y líneas..... | 24 |
| 4.1.4 | Barra de botones inferior.....                             | 25 |
| 4.1.5 | Paneles de estadísticas.....                               | 25 |
| 4.1.6 | El gráfico.....  | 26 |
| 4.2   | Zoomable sunburst.....                                     | 30 |
| 4.2.1 | Introducción.....  | 31 |
| 4.2.2 | Exclusión de datos/fechas exactos/aproximados.....         | 32 |
| 4.2.3 | Zoom.....  | 32 |
| 4.2.4 | Modificación del orden.....                                | 33 |
| 5     | Conclusiones.....  | 35 |
| 6     | Referencias.....   | 36 |
|       | ANEXO I.....   | 37 |

# 1 Introducción

## 1.1 Objetivos

Este trabajo presenta una forma novedosa de facilitar el estudio de los textos técnicos que conforman la 'matemática' (en el sentido antiguo del término) griega antigua. Hasta el momento no hay apenas herramientas visuales relacionadas con el estudio de estos textos. El objetivo de esta propuesta es obtener una herramienta que permita visualizar de manera gráfica y dinámica en función de varios parámetros una cronología de la matemática antigua con base cuantitativa, cualitativa y probabilística. Probabilística en cuanto a que la datación de las obras de la antigüedad clásica no es exacta y normalmente se estima un intervalo temporal aplicando una distribución de probabilidad adecuada; cuantitativa porque se miden parámetros estadísticos como, por ejemplo, el número de obras, de palabras o de caracteres escritos en un determinado período; y cualitativa porque se tienen en cuenta las características de las obras como, por ejemplo, el contenido, el tipo de texto o el estilo.

El resultado de este trabajo será una herramienta útil para hacer un estudio visual del *corpus* de textos técnicos que conforman la 'matemática'. Pretende ser una herramienta pedagógica o de apoyo en entornos académicos relacionados con el estudio de esta materia.

El artículo se organiza de la siguiente forma: en primer lugar se introducirá el contexto del trabajo presentando un breve estado del arte de los corpus lingüísticos y analizando las posibles herramientas de software libre (librerías javascript) disponibles para programar la aplicación. A continuación se describirá el corpus utilizado y algunos detalles relevantes de la implementación. Por último, se expondrá la aplicación web resultante.

## 1.2 Requisitos

Los requisitos iniciales del proyecto son:

- La aplicación debe implementarse utilizando herramientas de software libre basadas en javascript y destinadas a la visualización de datos.
- La aplicación obtenida debe funcionar sobre un navegador web.
- Los gráficos obtenidos deben ser dinámicos y versátiles.

## 2 Estado del arte

### 2.1 Introducción

El presente proyecto pretende aplicar las herramientas de visualización de datos modernas a una base de datos con las obras técnicas conservadas de la antigüedad greco-romana en el período que abarca desde el año 350 a. C. hasta el año 700 d. C., permitiendo obtener una visión gráfica, cuantitativa y objetiva de esta información. Aunque el proyecto se podría clasificar dentro de la especialidad de la humanística computacional, habría que destacar que no existen antecedentes sólidos ni bibliografía específica sobre los objetivos propuestos. A continuación se realizará un repaso a los conceptos generales, la historia y la evolución de la humanística computacional y de las técnicas utilizadas en el tratamiento de información lingüística desde sus inicios.

### 2.2 Inicios y evolución de la lingüística computacional

En primer lugar, habría que definir el concepto más importante en este ámbito: corpus lingüístico. Según el lingüista Tony McEnery [McEnery, 2001] se define como corpus lingüístico al estudio del lenguaje basado en ejemplos de uso de la "vida real", por ejemplo textos escritos o discursos hablados, mediante una recopilación sistemática de los mismos. La creación de los primeros corpora se realizó con diferentes objetivos, entre ellos el estudio de la adquisición del lenguaje en niños (1876-1926), el estudio de las distribuciones de frecuencia y secuencias de letras (Kading en 1897), el estudio comparativo de diferentes idiomas (Eaton 1940) o el estudio pedagógico aplicado al aprendizaje de lenguas extranjeras (Fries y Traver en 1940, y Bongers en 1947).

Sin embargo, McEnery recuerda que la metodología de estudio del lenguaje mediante corpus lingüísticos tuvo unos inicios controvertidos, recibiendo duras críticas principalmente del lingüista Noam Chomsky, el cual abogaba por un estudio introspectivo del lenguaje (el conocimiento interiorizado de una lengua) rechazando las metodologías descriptivas como es la de los corpora lingüísticos. Las críticas atacaban el método argumentando que, debido a la naturaleza recursiva de las gramáticas utilizadas, el lenguaje no es finito y en consecuencia es imposible estudiarlo de manera completa tomando únicamente observaciones empíricas. A estas críticas se sumaba que el tratamiento de un corpus formado por varios millones de lemas en aquella época suponía un arduo trabajo y resultaba caro, lento y propenso a errores; inviable en definitiva.

Las opiniones contrarias indicaban que había informaciones imposibles de obtener únicamente mediante la metodología introspectiva, sobre todo cualquier información cuantitativa, como por ejemplo la frecuencia estadística de un lema.

Debido a las críticas de Chomsky, entre otros, hasta el año 1970 los métodos empíricos fueron minoritarios respecto a otros métodos basados en la introspección.

Hoy en día, aunque algunas críticas de Chomsky han sido refutadas (la lingüística de corpus permite extraer información estadística muy útil para el estudio del lenguaje), otras siguen siendo válidas; por ejemplo, no deja de ser cierto que el lenguaje no es finito. A pesar de todo, la metodología basada en los corpora ha evolucionado y es ampliamente utilizada con diversos propósitos. Sobre todo gracias al gran incremento de la potencia de computación de los ordenadores y al ahorro de tiempo respecto de los sistemas manuales de trabajo se ha demostrado su gran utilidad práctica en diversas áreas de aplicación, desde las más teóricas, como el estudio del lenguaje y las lenguas, hasta las más modernas, como la creación de herramientas de traducción automática y de síntesis de voz.

McEnery concluye que la mejor opción es considerar ambos métodos como complementarios y aprovechar las ventajas de ambos supliendo así las carencias de cada uno. Por otra parte, añade que hay casos, como el estudio de textos antiguos, en los que ante la falta de parlantes nativos, la única manera de obtener información es mediante el estudio de los corpora.

Una vez que se dispone de un corpus con una cantidad representativa de lemas se debe realizar un trabajo adicional para poder explotar todas sus posibilidades y convertirlo en una herramienta aún más útil. Este trabajo consiste en enriquecer el texto incorporándole anotaciones con información interpretativa del ámbito lingüístico. Tal como describe el lingüista Geoffrey Leech en **[Leech, 2013]**, la anotación del corpus es una práctica ampliamente aceptada y una contribución crucial para realizar futuros trabajos sobre el corpus. Se define como "corpus annotation" tanto al trabajo de enriquecimiento realizado sobre el texto, como al conjunto de anotaciones resultantes una vez finalizado el trabajo. Leech indica que es paradójico que para poder extraer información útil de un corpus antes haya que añadirle información, pero afirma que una vez completado el trabajo de anotación, el corpus se convierte en una herramienta muy potente y reutilizable con distintos propósitos. Por este motivo es muy importante que las anotaciones se realicen siguiendo los estándares establecidos y que siempre se pueda recuperar el texto en bruto separándolo de las capas

de anotaciones. Aunque el trabajo de anotación es de carácter interpretativo e inherentemente manual, actualmente existen métodos avanzados, basados en algoritmos probabilísticos y en el análisis del contexto (las palabras adyacentes a la que se analiza en cada momento) que producen las anotaciones automáticamente con márgenes de error muy pequeños. Para su correcto funcionamiento se deben alimentar con datos estadísticos preexistentes (por ejemplo, otro corpus anotado manualmente). Existen también programas para la búsqueda de concordancias y la generación de listas de frecuencias.

El tipo de anotación más común es el llamado "Part-of-speech tagging" que trata de anotar la clasificación gramatical de los elementos que conforman el texto, como por ejemplo la forma verbal de un verbo, la distinción entre sustantivo y verbo, etc. Este es el tipo de anotación básico que debería tener cualquier corpus y servirá de base a otros tipos de anotaciones.

## 2.3 Herramientas de visualización

Una vez introducidos los conceptos básicos, ya se puede pensar en incorporar el componente de visualización de datos al proyecto. Para ello se escogerá una herramienta de software, concretamente un librería javascript, que permita representar gráficamente en un entorno web los datos cuantitativos y en algún caso cualitativos relacionados con la base de datos de trabajo. Los únicos requisitos que deberá cumplir la librería escogida serán estar bajo una licencia de software libre y funcionar sobre un entorno web mediante un navegador. Para facilitar la elección de la librería se ha tomado como referencia el análisis reciente de herramientas de visualización realizado por Veit-Dieter Vogt [Vogt, 2014] como parte de su proyecto de máster en software libre, el cual se ajusta exactamente a los requerimientos expuestos. Tras haber realizado un sondeo por internet en busca de librerías adicionales a las propuestas por Vogt se han encontrado bastantes opciones que él no había tenido en cuenta. Sin embargo no se ha encontrado ninguna que se adapte mejor a los requisitos del proyecto. Por ejemplo, las librerías *leaflet.js* y *modest maps.js* se centran en la creación de mapas interactivos; las librerías *three.js* y *babylon.js* se centran en la creación de gráficos 3D y tampoco encajan con las necesidades planteadas; otras muchas se basan en *d3.js*, que ya está incluida en el estudio, o tienen funcionalidades muy limitadas. Por los motivos anteriores y dado que el análisis realizado por Vogt es muy reciente, no se han tenido en cuenta librerías adicionales. En la siguiente tabla se muestran las librerías javascript analizadas por Vogt.

| Librería     | URL                     | Licencia      | Hits 2014  | Hits 2015  |
|--------------|-------------------------|---------------|------------|------------|
| d3.js        | d3js.org                | BSD 3-clauses | 4.400.000  | 16.000.000 |
| graphdracula | graphdracula.net        | MIT           | < 50       | 1.680      |
| jointjs      | jointjs.com             | MPLv2         | 92.900     | 24.400     |
| jsdot        | code.google.com/p/jsdot | MIT           | 805        | 2.170      |
| jsplumb      | jsplumbtoolkit.com      | MIT/GPLv2     | 163.000    | 48.400     |
| Sigma.js     | sigmajs.org             | MIT           | 16.200.000 | 73.800.000 |

*Tabla 1 - Presencia en Internet de varios frameworks js*

Sin pretender repetir el análisis completo realizado por Vogt, se ha realizado un experimento para obtener una primera impresión sobre cada librería en base al número de resultados de búsqueda de cada una en internet. El resultado de este experimento se puede apreciar comparando las dos últimas columnas de la tabla anterior. Se puede observar que jointjs y jsplumb han perdido relevancia y siguen siendo muy poco populares en relación con d3.js y Sigma.js, los cuales han incrementado ostensiblemente su popularidad. A pesar de que los valores absolutos han cambiado, se puede observar que la tendencia sigue siendo la misma que en el análisis de Vogt y esto hace que se puedan tomar en consideración los resultados obtenidos por su análisis.

El análisis general de las librerías realizado por Vogt, sin tener en cuenta las particularidades de su proyecto, consiste en una comparativa según los siguientes factores:

- Funcionalidad.
- Cuota de mercado (de uso).
- Niveles de soporte, mantenimiento y fiabilidad.
- Flexibilidad y posibilidades de personalización.

La comparativa propone como mejor opción la librería d3.js, destacando sobre todo su flexibilidad, ya que permite realizar todo tipo de gráficos (desde la representación de series temporales hasta la representación de jerarquías radiales o gráficos de burbujas). Este último aspecto es esencial para la realización de este proyecto, ya que se pretenden mostrar diferentes maneras de visualizar los datos. También destaca por la gran cantidad de ejemplos y manuales disponibles. Según los datos obtenidos, el framework d3.js parece el idóneo para crear unos gráficos vistosos, dinámicos y funcionales a partir de la base de datos de obras de la antigüedad.

Para comprobar el potencial del framework y confirmar que será una buena elección, a continuación se analizarán sus posibilidades en base a la documentación disponible en la página web del mismo (<http://d3js.org/>) y al libro *D3.js in Action* [Meeks, 2015], un referente en el aprendizaje del framework. Según Elijah Meeks, el autor del libro, las personas que se introducen en la visualización de datos, y concretamente en d3.js, provienen de tres áreas diferentes. Los primeros provienen del desarrollo web tradicional, los segundos del desarrollo de software tradicional (normalmente como parte de una transición hacia HTML5) y los terceros del análisis estadístico utilizando R, Python o aplicaciones de escritorio. Este hecho da una idea de la importancia que están adquiriendo las herramientas de visualización de datos online.

A continuación se muestra una definición sucinta pero muy completa de lo que es el framework, escrita por su creador, el especialista en visualización de datos Mike Bostock:

*"D3.js is a JavaScript library for manipulating documents based on data. D3 helps you bring data to life using HTML, SVG, and CSS. D3's emphasis on web standards gives you the full capabilities of modern browsers without tying yourself to a proprietary framework, combining powerful visualization components and a data-driven approach to DOM manipulation."*

Según la definición anterior, el framework d3.js se basa en la utilización de elementos estándares de la programación web: HTML, CSS y SVG (del inglés Scalable Vector Graphics). SVG es una especificación recomendada por el W3C para describir gráficos vectoriales bidimensionales, tanto estáticos como animados, en formato XML y, por lo tanto, soportada de forma nativa en los navegadores de internet modernos. Esto hace que sea una buena alternativa frente a otros frameworks propietarios.

En la misma definición también se indica que D3 permite la manipulación del DOM (del inglés Document Object Model). Quizás ésta sea una de las características más importantes de D3, ya que se refiere a que D3 permite asociar datos arbitrarios al DOM y realizar transformaciones en función de estos datos, con transiciones suaves entre los diferentes estados. Por ejemplo, permitiría generar una tabla HTML a partir de un array de números o un gráfico de columnas, las cuales se transformarían incrementando o reduciendo su altura en concordancia con los datos asociados.



Para facilitar la creación de diferentes tipos de gráficos, D3 incluye una serie de funciones denominadas *layouts*, que permiten formatear los datos para permitir su representación posterior en forma de gráficos. Las siguientes imágenes ilustrativas son de M. Bostock y se pueden encontrar en <https://github.com/mbostock/d3/wiki/API-Reference#d3layout-layouts>. A continuación se mostrarán algunos ejemplos de gráficos creados mediante los diferentes tipos de *layout* para poder evaluar la potencia del framework.

**Histograma** (Imagen 1) (*d3.layout.histogram*): los datos se preparan para crear una visualización clásica en forma de barras, columnas, etc. Para conseguir un gráfico flexible y dinámico se basa en la definición de escalas y ejes de coordenadas. Una escala consiste en la definición de un dominio de los datos de entrada y de un rango de valores de salida. Un eje se define mediante la aplicación de una escala, un número orientativo de "marcas" (*ticks*) y la posición del eje en el gráfico (arriba, abajo, izquierda o derecha).

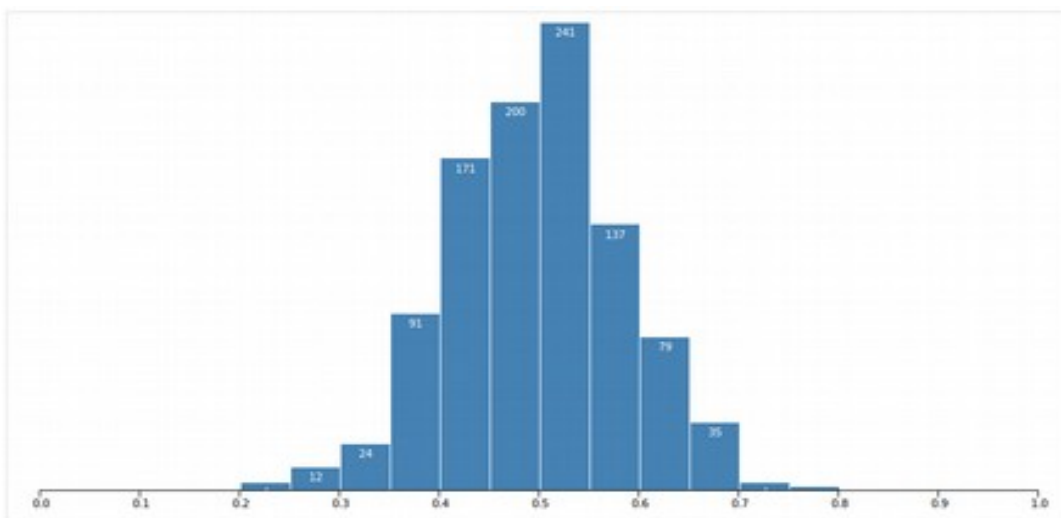


Imagen 1 - Histograma

**Gráfico circular** (*d3.layout.pie*): mediante este *layout* se pueden crear gráficos circulares, como el típico gráfico en forma de quesitos o en forma de anillo. Mediante las transformaciones se pueden conseguir efectos como, por ejemplo, separar una sección de las demás y añadir interactividad. Mediante este *layout* se transforman los datos de entrada obteniendo los ángulos necesarios para la representación gráfica. A partir de estos datos, mediante otra función, llamada *generator*, se generan los elementos SVG (*paths*) que formarán la representación gráfica. Variando el radio inicial y final

de los arcos se pueden generar los gráficos circulares y efectos de transición. En la Imagen 2 se muestra un ejemplo de este tipo de gráfico.

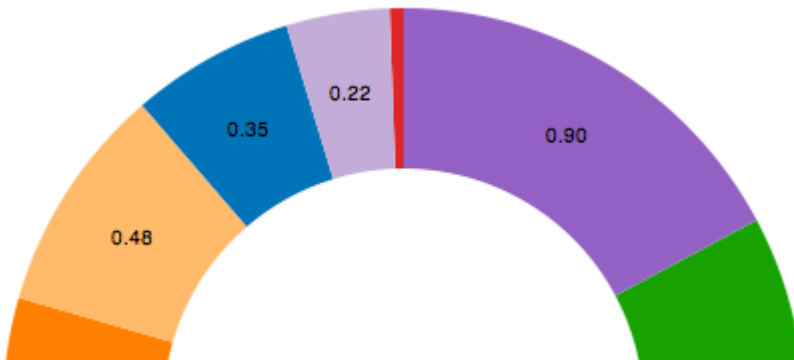


Imagen 2 - Gráfico circular

**Paquete** (*d3.layout.pack*): mediante este *layout* se puede generar toda una familia de gráficos consistentes en elementos anidados entre ellos de forma jerárquica. Si se alimenta con una jerarquía plana se puede obtener un gráfico de burbujas. El *layout* es configurable mediante funciones auxiliares para conseguir la representación deseada (por ejemplo, se puede variar el espaciado entre las burbujas). Variando el orden de la jerarquía dinámicamente se pueden conseguir bonitos efectos de transición. En la Imagen 3 se muestra un ejemplo de este tipo de gráfico.

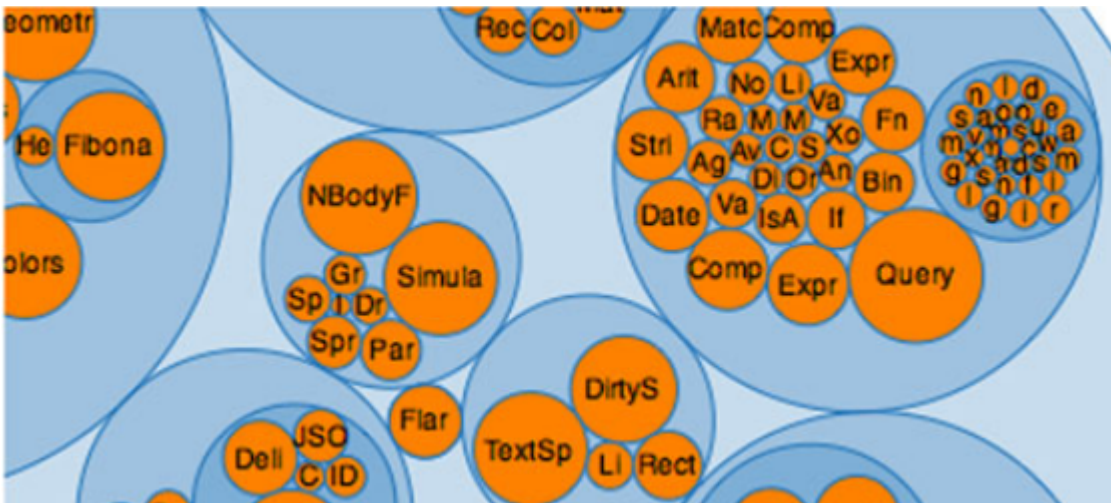


Imagen 3 - Gráfico de burbujas

**Árbol** (*d3.layout.tree*): permite generar diagramas en forma de árbol. Se pueden generar con diferentes orientaciones: vertical, horizontal o radial y realizar transformaciones variando los criterios de selección u ordenación de los datos. En la Imagen 4 se muestra un ejemplo con orientación radial. Se usa de manera análoga al *layout* de tipo paquete.



Imagen 4 - Gráfico en forma de árbol

**Pila** (*d3.layout.stack*): mediante un array formado por otros arrays que representan las diferentes capas, este *layout* calcula las coordenadas y la amplitud de cada capa que servirán para generar la representación gráfica. A menudo se combina con algoritmos heurísticos de ordenación para mejorar la percepción en el resultado final. También permite definir la orientación del gráfico: horizontal, vertical o radial (Imagen 5).

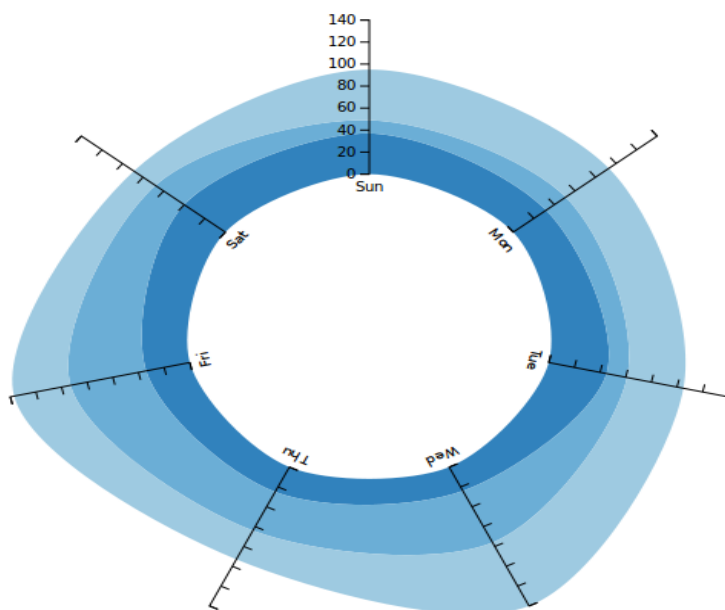


Imagen 5 - Gráfico radial

**Partición** (*d3.layout.partition*): este *layout* permite generar diagramas de manera parecida a los diagramas en forma de árbol, pero rellenando los espacios intermedios entre los nodos. Se pueden generar arcos o rectángulos, creando bonitos gráficos como el de la Imagen 6.



Imagen 6 - Gráfico circular

Los ejemplos anteriores son una muestra de los gráficos básicos que se pueden generar mediante el framework. Hay que tener en cuenta que en las imágenes de ejemplo no se aprecian los efectos de transformación y de transición, que enriquecen mucho la experiencia de usuario. En los ejemplos se han obviado algunos *layouts* que quedan fuera de los requerimientos de este proyecto, como por ejemplo los relacionados con la generación de mapas geográficos. Además de los *layouts* básicos existe un amplio repertorio de plugins que permiten ampliar aún más el catálogo de gráficos posibles.

Tras haber experimentado con los ejemplos disponibles en la página de Bostock, se ha podido comprobar que hay cierta dificultad inicial en comprender el funcionamiento del framework pero una vez superada esta primera barrera, los resultados que se obtienen son muy buenos en relación a

la cantidad de líneas de código necesarias para su programación. En la página <https://github.com/mbostock/d3/wiki/Gallery> se pueden consultar una gran cantidad de ejemplos adicionales que demuestran la potencia y flexibilidad del framework (Imagen 7).

En conclusión, se considera que el framework d3.js se ajustará a los requerimientos del proyecto.

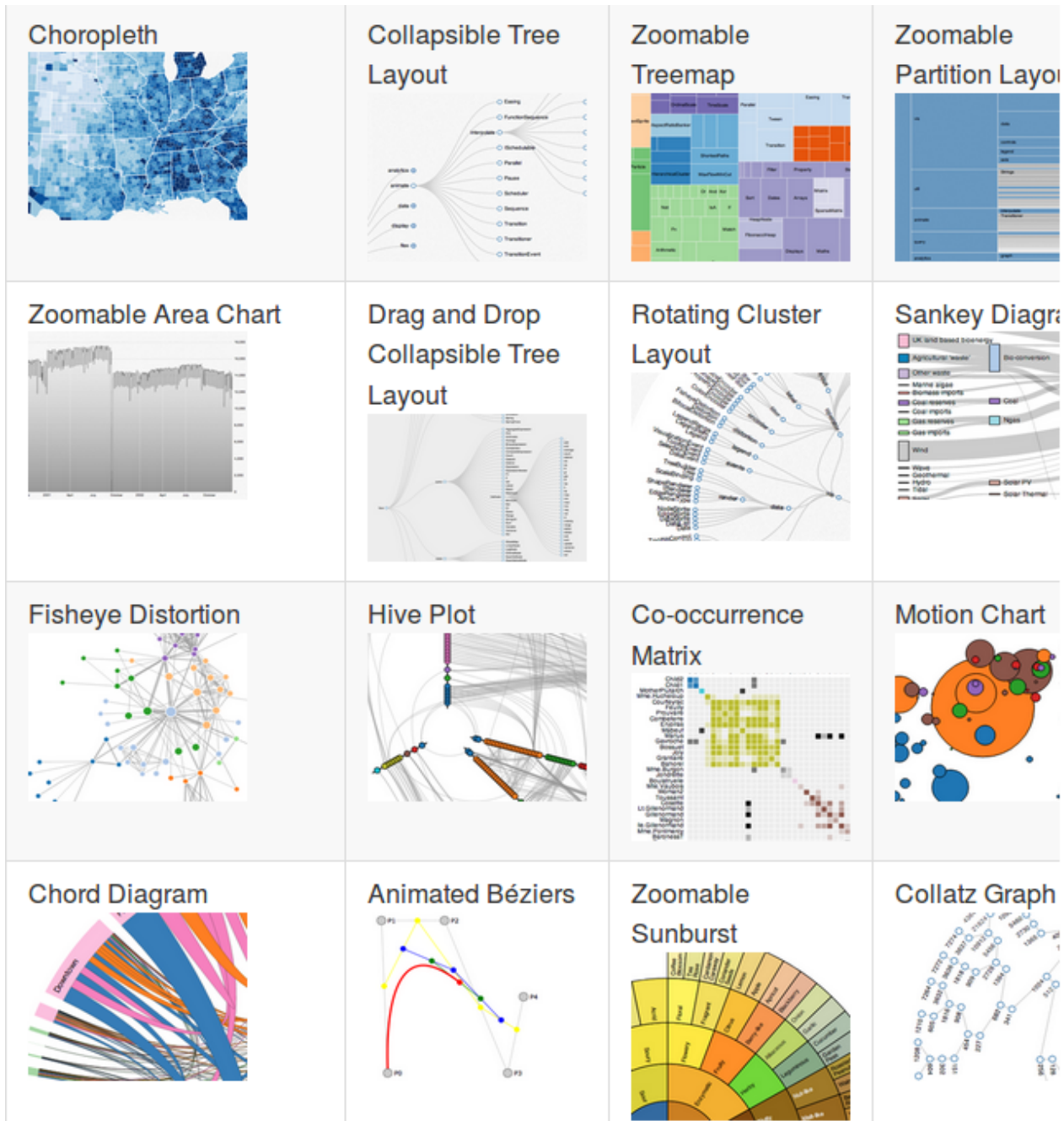


Imagen 7 - Gráficos de ejemplo hechos mediante d3.js

## 3 Descripción y tratamiento del corpus

### 3.1 Introducción

En los capítulos anteriores se ha realizado un breve análisis del estado del arte en el tratamiento automático de *corpora* textuales y se ha descrito el *framework* d3.js escogido para implementar la herramienta web.

Mediante el *framework* d3.js se han implementado dos tipos de visualización diferentes: una línea temporal (*timeline*) y un gráfico de tipo *sunburst*.

El *timeline* situará las obras en una escala temporal representando las funciones de densidad de probabilidad del período estimado de creación de cada obra así como la función de densidad suma de todas las obras. Por otra parte, mostrará las funciones de distribución acumuladas de las distintas categorías y de la suma de todas las obras.

El *sunburst* mostrará las categorías, autores y obras en forma de círculos concéntricos divididos en arcos. Permitirá reordenar y filtrar los resultados y calcular los porcentajes de caracteres y palabras seleccionados.

En los siguientes apartados de este capítulo se describirán la base de datos (corpus) utilizada y algunos detalles relevantes de la implementación.

### 3.2 Origen del corpus de obras de la antigüedad

A efectos de este proyecto se trabajará sobre una base de datos de aproximadamente 150 obras técnicas conservadas de la antigüedad greco-romana. La obtención de esta base de datos queda fuera del ámbito de este proyecto y es fruto de un trabajo manual de consultas en los *corpora* disponibles actualmente sobre esta temática y en otras fuentes relacionadas, destacando sobre todo el corpus del TLG (Thesaurus Linguae Graecae), el cual es consultable tanto en línea a través de internet (<http://stephanus.tlg.uci.edu/>) como mediante copias fuera de línea, como por ejemplo en dvd. La información básica que contendrá la base de datos serán el nombre de la obra, el autor, la

datación y el número de palabras y caracteres de cada obra. Adicionalmente se distinguirán las obras con una datación exacta de las que tengan una datación estimada. Destacar también que el reducido número de obras conservadas que conforman el corpus, permite que se pueda almacenar y tratar fácilmente mediante una hoja de cálculo. Este hecho además resultará ventajoso en cuanto a los requerimientos computacionales necesarios para su tratamiento. El TLG incorpora un apartado de estadísticas en el cual se pueden consultar algunos datos sobre los autores, las obras y los lemas del corpus y son un ejemplo de aplicación de la visualización de datos a los corpus lingüísticos como el que se pretende obtener mediante este proyecto. Por ejemplo, en la Imagen 8 se muestra una gráfica sobre el uso del lema *ἀάατος, -ov* en relación a los autores.



Imagen 8 - Uso del lema *ἀάατος, -ov* en relación a los autores

### 3.3 Clasificación de las obras del corpus

Dado que el proyecto se basará en las obras técnicas conservadas de la antigüedad greco-romana se comentará brevemente el trabajo de recopilación e interpretación de estos textos realizado por Bernard Vitrac [Vitrac, 2015]. Vitrac parte de la idea de que la clasificación de las ciencias matemáticas en la antigüedad greco-romana [Vitrac, 2005], puede considerarse con los siguientes propósitos:

- Como parte de un inventario sistemático de la totalidad del conocimiento humano.
- Como la ordenación de una serie de descubrimientos o invenciones considerados estables.

- Como la base de un programa educativo que debe dominar una parte de la comunidad política para cumplir adecuadamente su rol en la sociedad.
- Como una herramienta para el bibliotecario para almacenar una colección de trabajos matemáticos.

Partiendo de los criterios anteriores, Vitrac intenta obtener las categorías mediante las cuales los autores de esta época clasificaban los escritos matemáticos. Siempre desde un punto de vista histórico comienza por definir las matemáticas y distinguiendo la diferencia entre matemática (en singular) con un sentido filosófico y matemáticas en plural, en el sentido moderno del término.

A partir de los escritos conservados Vitrac concluye que no puede encontrar una definición de matemáticas atribuible a Platón. Sin embargo, sí encuentra suficientes evidencias en los escritos de su discípulo Aristóteles (siglo IV a.C.), en los cuales éste diferencia el rol de "matemático" en oposición al de "físico", (además habla de una tercera forma de saber: la teología). El estudio de la ciencia teórica en oposición al estudio de los objetos.

Posteriormente, analizando los escritos de Proclo (siglo V d. C.), Vitrac evidencia dos maneras de clasificar el conocimiento matemático. El primero, llamado *quadrivium* pitagórico, consiste de cuatro ciencias: aritmética, geometría, música y astronomía. El segundo, atribuido a Geminus of Rhodes (siglo I a. C.), la divide en ocho ramas agrupadas en dos bloques según su naturaleza. Por una parte las de naturaleza teórica (matemática pura): aritmética y geometría, y por otra las relacionadas con objetos materiales (matemática aplicada): mecánica, astronomía, óptica, geodesia, canon y logística.

A partir de estas ideas y de acuerdo a la cantidad y contenido de las obras conservadas, para la realización de este proyecto se considerará una clasificación formada por cuatro categorías, las dos primeras puramente teóricas y las otras dos aplicadas:

- **Geométricas:** relacionadas con la longitudes, áreas y volúmenes.
- **Aritméticas:** estudio de los números y las operaciones hechas con ellos.
- **Astronómicas:** las relacionadas con el comportamiento de los objetos celestes.
- **Técnicas:** las obras relacionadas con la aplicación práctica de las matemáticas, como la óptica, la balística, la geografía, etc.



### 3.4 Tratamiento de los datos del corpus

A partir de la hoja de cálculo que contiene el corpus se generará un fichero separado por comas (tabla.csv) que una vez cargado por el programa permitirá crear un objeto JSON (ver Figura 1), el cual será utilizado para realizar los cálculos y generar los gráficos.

```
childNode = {  "Gauss": gauss,
  >>  "Work": csv[i][0].toLowerCase(),
  >>  "Content": csv[i][4].toLowerCase(),
  >>  "Genre": csv[i][5].toLowerCase(),
  >>  "Style": csv[i][6].toLowerCase(),
  >>  "Author": csv[i][7].toLowerCase(),
  >>  "Works": works,
  >>  "Books": parseFloat(csv[i][9]),
  >>  "Words": words,
  >>  "Chars": chars,
  >>  "Data": data,
  >>  "Date": date,
  >>  "dateBegin": datebegin,
  >>  "dateEnd": dateend,
  >>  "samples": createSamples(mu, sigma, gauss, 1, datebegin, dateend),
  >>  "cutdBegin": cutdBegin,
  >>  "cutdEnd": cutdEnd,
  >>  "worksamples": createSamples(mu, sigma, gauss, works, datebegin, dateend, cutdBegin, cutdEnd),
  >>  "wordssamples": createSamples(mu, sigma, gauss, words, datebegin, dateend, cutdBegin, cutdEnd),
  >>  "charsamples": createSamples(mu, sigma, gauss, chars, datebegin, dateend, cutdBegin, cutdEnd),
  >>  "cdfsamples": createCdfSamples(mu, sigma, gauss, 1, datebegin, dateend),
  >>  "cdfworks": createCdfSamples(mu, sigma, gauss, works, datebegin, dateend),
  >>  "cdfwords": createCdfSamples(mu, sigma, gauss, words, datebegin, dateend),
  >>  "cdfchars": createCdfSamples(mu, sigma, gauss, chars, datebegin, dateend)
}
jsonData.push(childNode);
```

Figura 1 - Objeto JSON generado a partir del corpus

A continuación se describirán detalladamente cada uno de los campos del corpus que se han utilizado:

**Work:** Nombre de la obra o fragmento. Hay algunas obras que se dividen en fragmentos debido a que no se conocen los datos exactos de la obra completa pero sí de una parte de la misma. Por ejemplo: "*Planetary Hypotheses*".

**Mu:** La media del período en el que se clasifica la obra. Por ejemplo, si el período de creación de una obra se estima que es desde el año -320 hasta el año -250, la media será -285.

**Sigma:** Es la desviación típica respecto de la media mu.

**Gauss:** El valor "SI" indicará que la función de probabilidad que define en qué año se escribió la obra es una campana de Gauss (función normal). El valor "NO" indicará que la función de probabilidad es semi-uniforme (uniforme adaptada para que sea continua en los extremos).

**Content:** Indica la clasificación de la obra en cuanto a contenido: *arithmetical, astronomical, geometrical o technical*.

**Genre:** Indica la clasificación de la obra en cuanto a género: *commentary, compilation, popularization o treatise*.

**Style:** Indica la clasificación de la obra en cuanto a estilo: *canonical, informal o mixed*.

**Author:** El nombre del autor de la obra.

**Works:** El nombre de la obra o fragemento.

**Words:** El número de palabras o lemas de la que consta la obra.

**Chars:** El número de caracteres de la obra o fragmento.

**Datebegin:** El año de inicio del período estimado en que se escribió la obra.

**Dateend:** El año final del período estimado en que se escribió la obra.

**Data:** Indica si los datos obtenidos de la obra son exactos (*exact*) o aproximados (*approximate*).

**Date:** Indica si el período estimado en que se escribió la obra es exacto (*exact*) o aproximado (*approximate*).

### 3.5 Implementación de las funciones de probabilidad

En este apartado se mostrará la implementación en javascript de las dos funciones de probabilidad utilizadas en el *timeline*. Los parámetros de entrada utilizados son la media ( $\mu$ ) y la desviación típica ( $\sigma$ ). El tercer parámetro de las funciones,  $x_i$ , se refiere al valor concreto (año) al que se aplicará el cálculo. Para dibujar la función se deberán calcular los valores para cada uno de los años de un período temporal determinado.

En primer lugar se mostrará el código para el cálculo de las funciones de densidad.

- Función de densidad de la distribución normal. Esta función se ha implementado según la definición ampliamente documentada en diversas fuentes, por ejemplo en la página 931 de [Abramowitz, 1972] (Figura 2):

$$f(x | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Figura 2 - Función de densidad

```
function make_gaussian_func_xi(mu, sigma, xi) {
  var sqrt = Math.sqrt, pow = Math.pow, e = Math.E, pi = Math.PI;
  var a = 1 / (sigma * sqrt(2 * pi));
  return a * pow( e, - pow(xi - mu, 2) / (2 * pow(sigma, 2)) );
}
```

- Función de densidad de la distribución semi-uniforme. Los dos extremos de la función se unen con el eje x para que la función sea continua pero de manera que el área siga siendo 1. La definición de esta función se ha extraído de la tabla de obras proporcionada.

```
function make_semiuniform_func_xi(a,b,xi) {
  if ((xi < a) || (xi > b))
    return 0;
  var c = b-a;
  var c2 = c*c;
  if (xi < (a + c/100))
    return 99*xi/c2 - 99*a/c2;

  if (xi > (b - c/100))
    return -99*xi/c2 + 99*b/c2;

  return (100/(99*c));
}
```

En segundo lugar se muestra el código para el cálculo de las funciones de distribución.

- Función de distribución de la distribución normal:

La función de distribución se puede calcular mediante la siguiente fórmula (Figura 3) (pág. 931 de [Abramowitz, 1972]):

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

Figura 3 - Función de distribución

la cual se puede reescribir como (según [Wikipedia, 2015]) (Figura 4):

$$\Phi(x) = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{x}{\sqrt{2}} \right) \right]$$

Figura 4 - Función de distribución

O para una función normal con media  $\mu$  y desviación típica  $\sigma$ :

$$F(x) = \Phi \left( \frac{x - \mu}{\sigma} \right) = \frac{1}{2} \left[ 1 + \operatorname{erf} \left( \frac{x - \mu}{\sigma\sqrt{2}} \right) \right]$$

Figura 5 - F. de distribución con media y desviación típica

Por último, la función de error (erf) se puede calcular mediante una aproximación numérica (pág. 299 de [Abramowitz, 1972]).

La implementación resultante se muestra a continuación:

```
function cdf(x, mean, variance) {
    return 0.5 * (1 + erf((x - mean) / (Math.sqrt(2 * variance))));
}

function erf(x) {
    var sign = (x >= 0) ? 1 : -1;
    x = Math.abs(x);
    // constants
    var a1 = 0.254829592;
    var a2 = -0.284496736;
    var a3 = 1.421413741;
    var a4 = -1.453152027;
    var a5 = 1.061405429;
    var p = 0.3275911;
    // A&S formula 7.1.26
    var t = 1.0 / (1.0 + p * x);
    var y = 1.0 - (((((a5 * t + a4) * t) + a3) * t + a2) * t + a1) * t * Math.exp(-x * x);
    return sign * y; // erf(-x) = -erf(x);
}
```

- Función de distribución de la distribución semi-uniforme. En este caso se utiliza la función de distribución de la función uniforme:

```
function cdf_uniform(x, a, b) {
    if (x <= a) return 0;
    if (x >= b) return 1;
    return (x - a) / (b - a);
}
```

## 4 Resultados

En este apartado se describirán detalladamente el funcionamiento y las características de los dos gráficos implementados.

### 4.1 Timeline

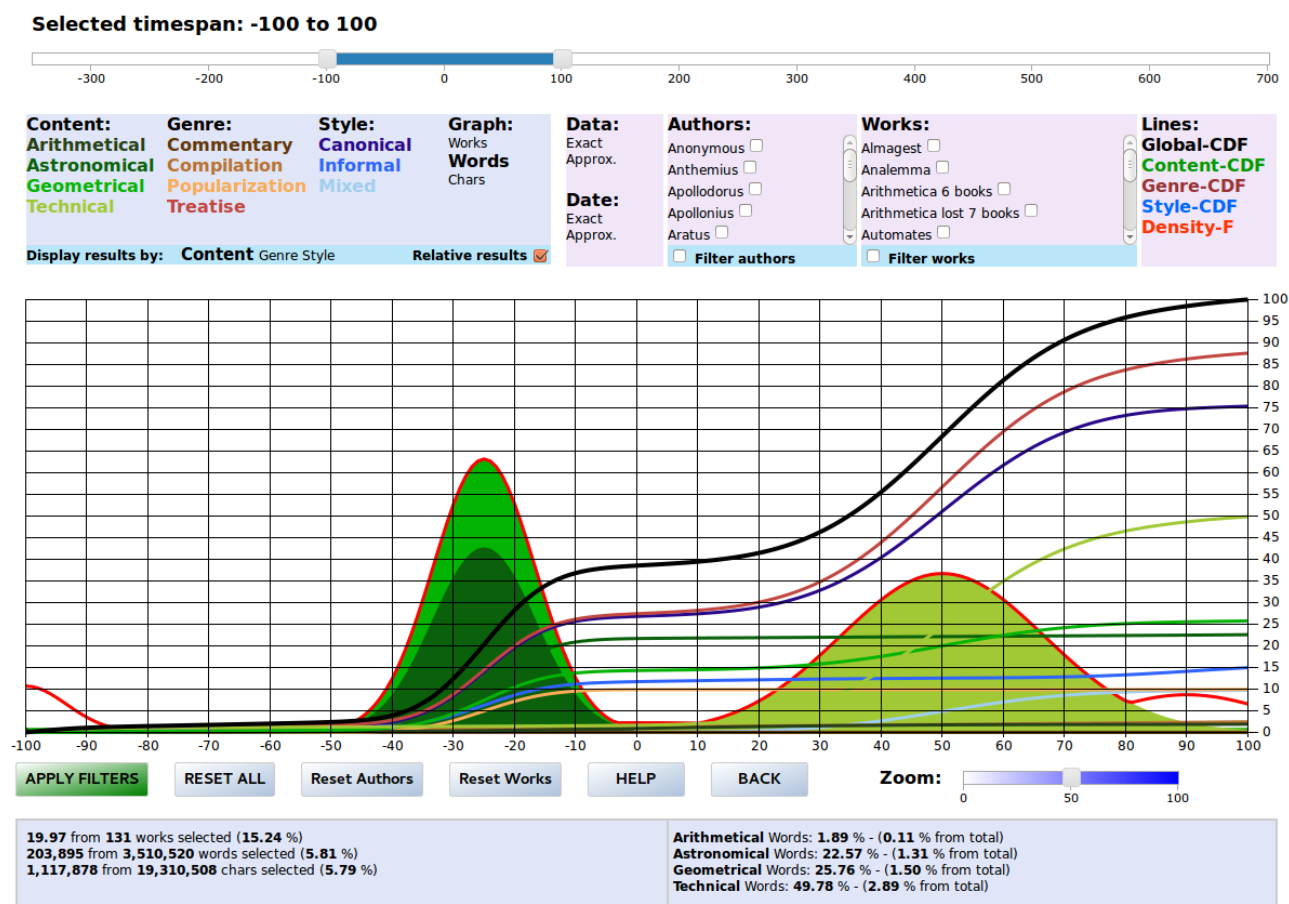


Imagen 9 - Aplicación web con la cronología de obras

En la Imagen 9 se muestra el gráfico con la línea temporal.

#### 4.1.1 Selección del período de tiempo

La barra de la parte superior del gráfico (ver Imagen 9) sirve para seleccionar el período de tiempo a analizar. El período por defecto abarca desde el año 350 a. C. hasta el año 700 d. C.. Para aplicar los cambios en el período seleccionado se debe utilizar el botón "**Apply filters**" de la barra inferior. Una vez aplicados los cambios, en el gráfico se mostrará solamente la sección temporal seleccionada. Si el período seleccionado cortara alguna obra, el fragmento que quede incluido se mostrará sin el

color de relleno (ver Imagen 10 e Imagen 11).

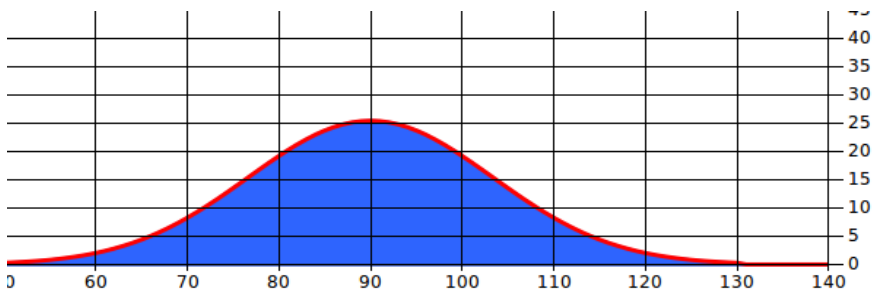


Imagen 10 - Representación completa de la función de densidad

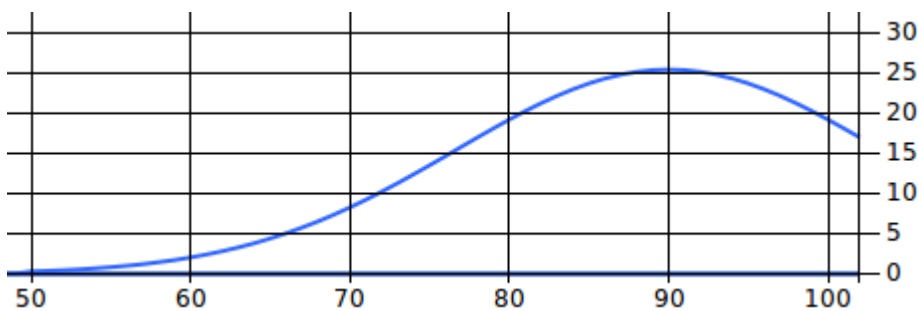


Imagen 11 - Representación parcial de la función de densidad

En este caso las estadísticas (número de obras, lemas o caracteres) se calcularán proporcionalmente a la parte que quede incluida. Esto provocará que se muestren decimales en el panel de estadísticas. (ver Imagen 12).

**13.37** from **131** works selected (**10.21** %)  
**173,85** from **3,510,520** words selected (**4.95** %)  
**951,038** from **19,310,508** chars selected (**4.92** %)

Imagen 12 - Estadísticas de la selección actual

#### 4.1.2 Selección de las categorías y del valor representado

| Content:            | Genre:         | Style:    | Graph:       |
|---------------------|----------------|-----------|--------------|
| <b>Arithmetical</b> | Commentary     | Canonical | <b>Works</b> |
| Astronomical        | Compilation    | Informal  | Words        |
| Geometrical         | Popularization | Mixed     | Chars        |
| Technical           | Treatise       |           |              |

Display results by: Content Genre **Style** Relative results

Imagen 13 - Menú de selección de categorías y del valor representado

#### **a) Filtros de selección de categorías**

El menú de la parte superior izquierda (ver Imagen 13) permitirá seleccionar las obras en función de su categoría (**Content**, **Genre** o **Style**). Haciendo clic sobre los posibles valores de cada categoría éstos se activarán o desactivarán. Haciendo clic sobre el nombre de la categoría, por ejemplo "Content", se invertirá la selección de todos los valores de esta categoría, de esta forma podemos seleccionar todos los valores mediante un sólo clic. El color de cada valor será el mismo que se muestre en las funciones de densidad y en las líneas que representan las funciones de distribución acumulada.

#### **b) Selección de la información a mostrar**

El cuarto menú de este cuadro, "**Graph**", sirve para escoger el tipo de información de la obra que se quiere comparar: número de obras (**works**), número de lemas (**words**) o número de caracteres (**chars**).

#### **c) Selección de la categoría a mostrar**

El menú "**Display results by**" de la parte inferior del cuadro permite seleccionar la categoría que se visualizará en el gráfico y de la cual se calcularán las estadísticas. Las funciones de densidad del gráfico se colorearán en función de los valores de la categoría seleccionada en este menú (ver Imagen 17 de la página 27).

#### **d) Selección de la escala (relativa / global)**

Mediante la casilla de selección "**Relative results**" se puede modificar la escala respecto de la cual se calcularán los porcentajes para la representación de las funciones de distribución acumuladas. Si se selecciona esta casilla el porcentaje se calculará respecto de las obras seleccionadas, en caso contrario el porcentaje se calculará en función del total de obras de la base de datos.

### 4.1.3 Selección de datos, períodos, autores, obras y líneas

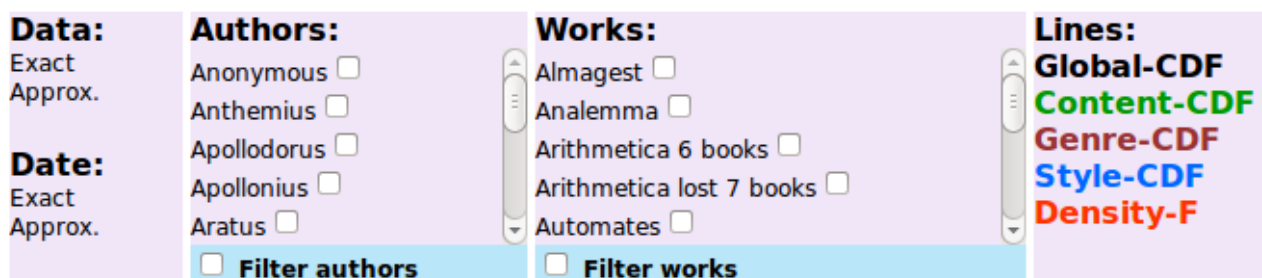


Imagen 14 - Menú de selección de datos, períodos, autores, obras y líneas

#### a) Exact / approximate data / date

Dado a que los datos y la datación de algunas obras son aproximados, mediante los menús "**Data**" y "**Date**" de la primera columna se podrán filtrar las obras con datación/datos exactos/aproximados.

#### b) Authors / Works

La segunda y tercera columnas, "**Authors**" y "**Works**", sirven para filtrar sólo los autores o las obras que se seleccionen. Estos menús funcionan de la siguiente manera:

- Al seleccionar un autor se seleccionan automáticamente todas las obras del mismo.
- Al deseleccionar un autor se deseleccionan automáticamente todas las obras del mismo.
- Al seleccionar una obra se selecciona automáticamente el autor de la misma.
- Al deseleccionar una obra sólo se deselecciona el autor si no hay más obras del mismo seleccionadas.

Para que se tengan en cuenta estos filtros hay que activarlos seleccionando la casilla "**Filter authors**" o "**Filter works**" correspondientes. Si se modifica la selección de autores o de obras cuando el filtro correspondiente ya está activado, los cambios se pueden aplicar deseleccionando y volviendo a seleccionar el filtro o mediante el botón "**Apply filters**" de la barra de botones inferior.



### c) **Mostrar / ocultar líneas**

Por último, el menú de la cuarta columna de este panel, "***Lines***", sirve para visualizar u ocultar las líneas correspondientes a la función de distribución acumulada global, a las funciones de distribución acumuladas de cada categoría y a la función de densidad global. Haciendo clic sobre el título ("***Lines***") se invierte la selección actual de líneas visualizadas y ocultas.

En caso de querer mostrar únicamente las líneas y ocultar las funciones de densidad bastará con situar el control de zoom en el valor cero.

## 4.1.4 **Barra de botones inferior**

La barra de botones inferior permite realizar las siguientes acciones:

- "***Apply filters***": aplica todos los cambios pendientes realizados en la selección del período y en las selecciones de autores y obras.
- "***Reset all***": devuelve el gráfico a su estado inicial, limpiando todas las selecciones realizadas.
- "***Reset authors***": deselecciona todos los autores sin afectar a la selección de obras.
- "***Reset works***": deselecciona todas las obras sin afectar a la selección de autores.
- "***Help***": muestra una ventana con las instrucciones de uso del gráfico.
- **Barra de zoom**: dado que las funciones de densidad no representan valores absolutos, sino únicamente la relación entre ellas, mediante este control es posible aumentar o disminuir el tamaño de las mismas para mejorar su visualización.

## 4.1.5 **Paneles de estadísticas**

Los dos paneles de la parte inferior muestran las estadísticas correspondientes a la selección actual (Imágenes 15 y 16).

1.00 from 131 works selected (0.76 %)  
3,771 from 3,510,520 words selected (0.11 %)  
22,502 from 19,310,508 chars selected (0.12 %)

Imagen 15

Canonical Works: 0 % - (0.00 % from total)  
Informal Works: 100 % - (0.80 % from total)  
Mixed Works: 0 % - (0.00 % from total)

Imagen 16

El panel izquierdo muestra el número de obras, lemas y caracteres de la selección actual y su porcentaje respecto del total de la base de datos.

El panel de la derecha, dependiendo de la selección realizada en el menú "**Graph**" (ver apartado 4.1.2.b), mostrará el porcentaje de obras, lemas o caracteres de cada valor de la categoría seleccionada en "**Display results by**" (ver apartado 4.1.2.c). Se muestra tanto el porcentaje en relación a la selección actual como en relación al total de la base de datos (entre paréntesis).

#### **4.1.6 El gráfico**

El eje horizontal representa el tiempo en años. Inicialmente abarcará desde el año 350 a. C. hasta el año 700 d. C. pero se puede modificar mediante el selector descrito en el apartado 4.1. El eje vertical representa el porcentaje de obras, lemas o caracteres; y relativo o absoluto, dependiendo del estado de la casilla "**Relative results**" (ver apartado 4.1.2.d), representado por las funciones de distribución acumulada.

A continuación se describirán los distintos elementos representados en el gráfico.

##### **a) Funciones de densidad de probabilidad de las obras individuales.**

En el gráfico se muestran las funciones de densidad de probabilidad correspondientes a todas las obras seleccionadas mediante los filtros y opciones mostradas en los apartados anteriores. Hay que tener en cuenta que algunas representaciones se solaparán y si la selección es amplia algunas incluso podrían quedar totalmente ocultas; en estos casos será útil el control de selección de autores y obras individuales.. Por ejemplo, en la Imagen 19 de la página 28 se aprecia que hay hasta 9 obras bajo un gráfico que aparenta muchas menos. A menudo esto ocurre con las obras de un mismo autor.

Por otra parte, tal como ya se ha comentado anteriormente, el color de relleno y del contorno de cada función se corresponde con su clasificación dentro de la categoría seleccionada actualmente en el menú "**Display results by**" (ver apartado 4.1.2.c). En la Imagen 17 se puede ver el mismo gráfico coloreado según las tres posibles categorías.

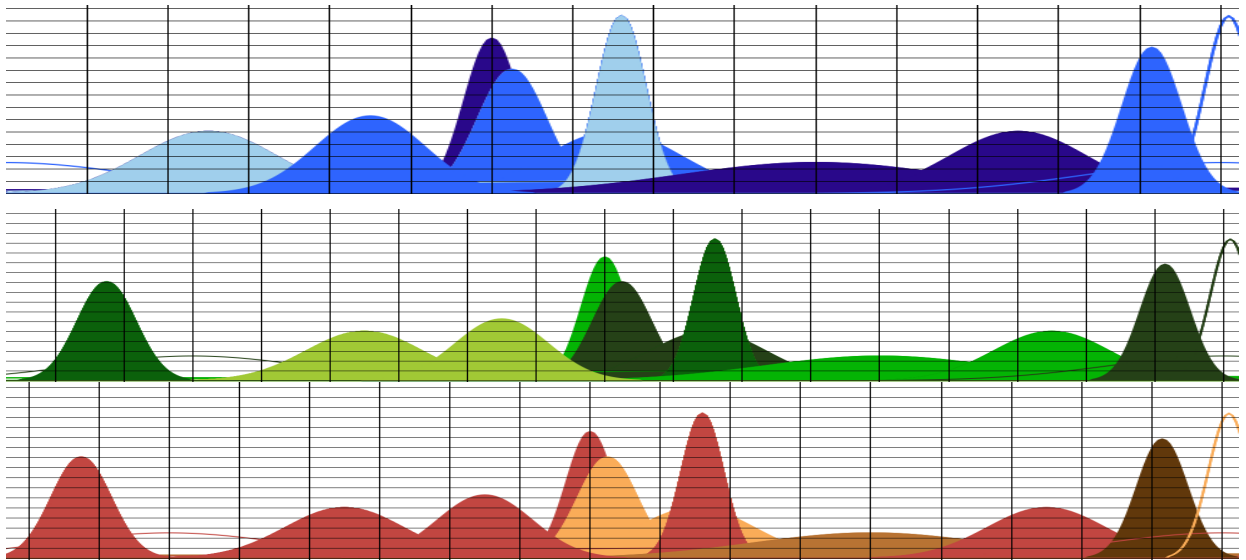


Imagen 17

Haciendo click sobre una función de densidad se muestra un recuadro con la información sobre la obra que representa. Si se hace clic sobre el recuadro los detalles se fijan a modo de *post-it* en la parte superior-izquierda del gráfico (ver Imagen 18).

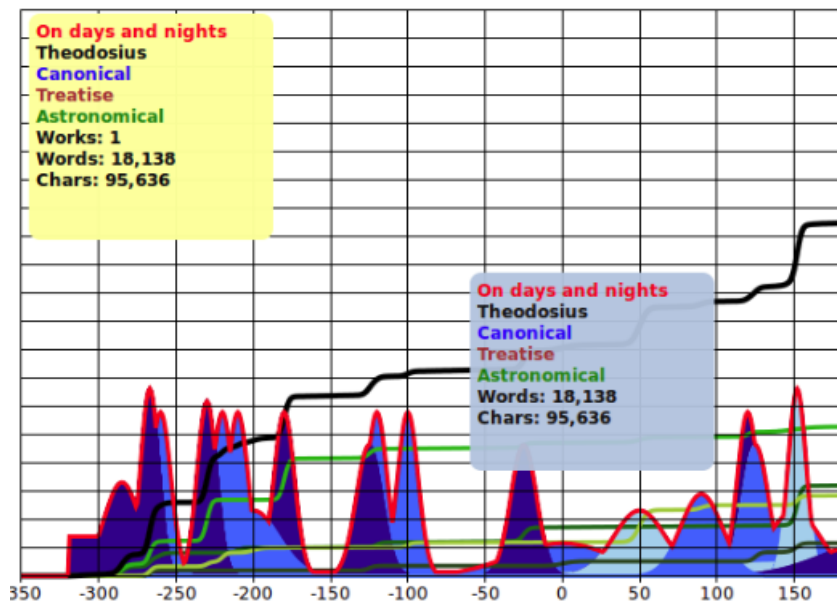


Imagen 18

## b) Función de distribución acumulada global.

La línea negra representa la función de distribución resultante de la selección actual completa de obras. Haciendo clic sobre esta línea en una posición determinada, se mostrará un recuadro con todas las obras cuyo período estimado de creación contenga el año correspondiente a la posición equivalente en el eje de abscisas (ver Imagen 19).

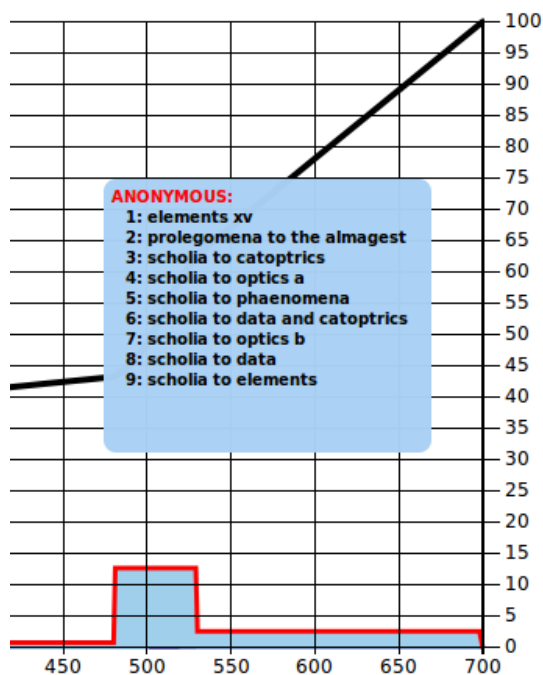


Imagen 19 - Porcentaje relativo

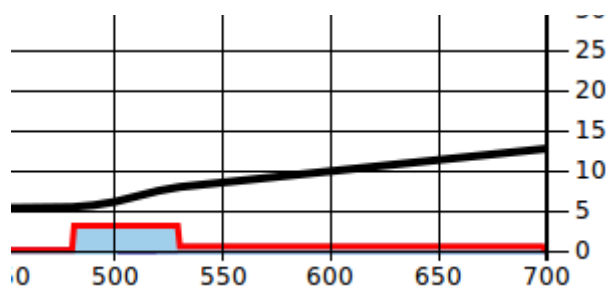


Imagen 20 - Porcentaje absoluto

En el ejemplo de la Imagen 19 el porcentaje es relativo a las obras seleccionadas actualmente, por lo que llega al 100%. El porcentaje absoluto para la misma selección apenas llegaría al 15% (ver Imagen 20).

## c) Funciones de distribución acumulada de las categorías seleccionadas.

Las funciones de distribución acumulada correspondientes a las categorías seleccionadas se dibujan según los colores mostrados en el menú de selección de categorías descrito en el apartado 4.1.2. Para que se muestren las funciones de una categoría debe haberse seleccionado al menos un valor de la misma y debe estar seleccionada en el menú "**Lines**" (ver apartado 4.1.3.c). En la Imagen 21 se muestran todas las funciones de distribución acumulada simultáneamente.

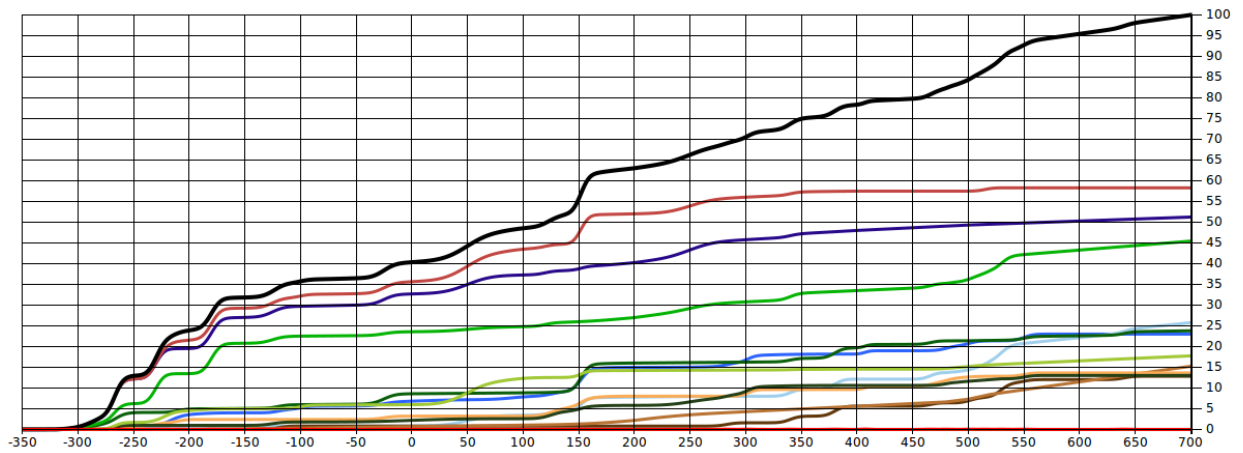


Imagen 21 - Funciones de distribución acumulada (zoom al mínimo)

Al hacer clic sobre cualquiera de estas líneas se mostrará un recuadro con todas las obras cuyo período estimado de creación contenga el año correspondiente a la posición equivalente en el eje de abscisas y cuya clasificación se corresponda con la categoría y valor de la línea escogida. Por ejemplo, en la Imagen 22 hemos hecho click sobre la línea correspondiente al valor "**Geometrical**" de la categoría "**Content**".

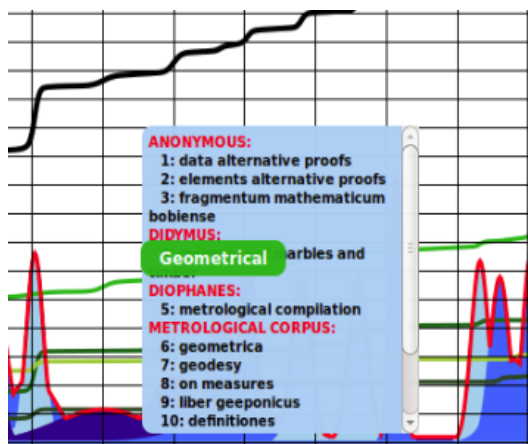


Imagen 22 - Recuadro de obras

#### d) Función de densidad global.

La línea roja (ver Imagen 23) representa la suma de todas las funciones de densidad individuales.

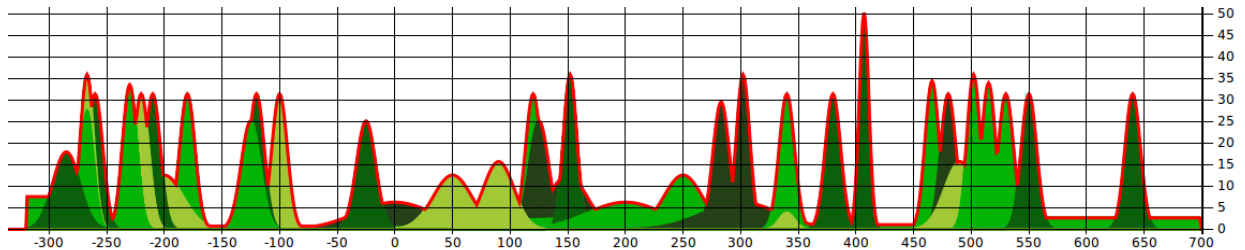


Imagen 23 - Función de densidad

## 4.2 Zoomable sunburst

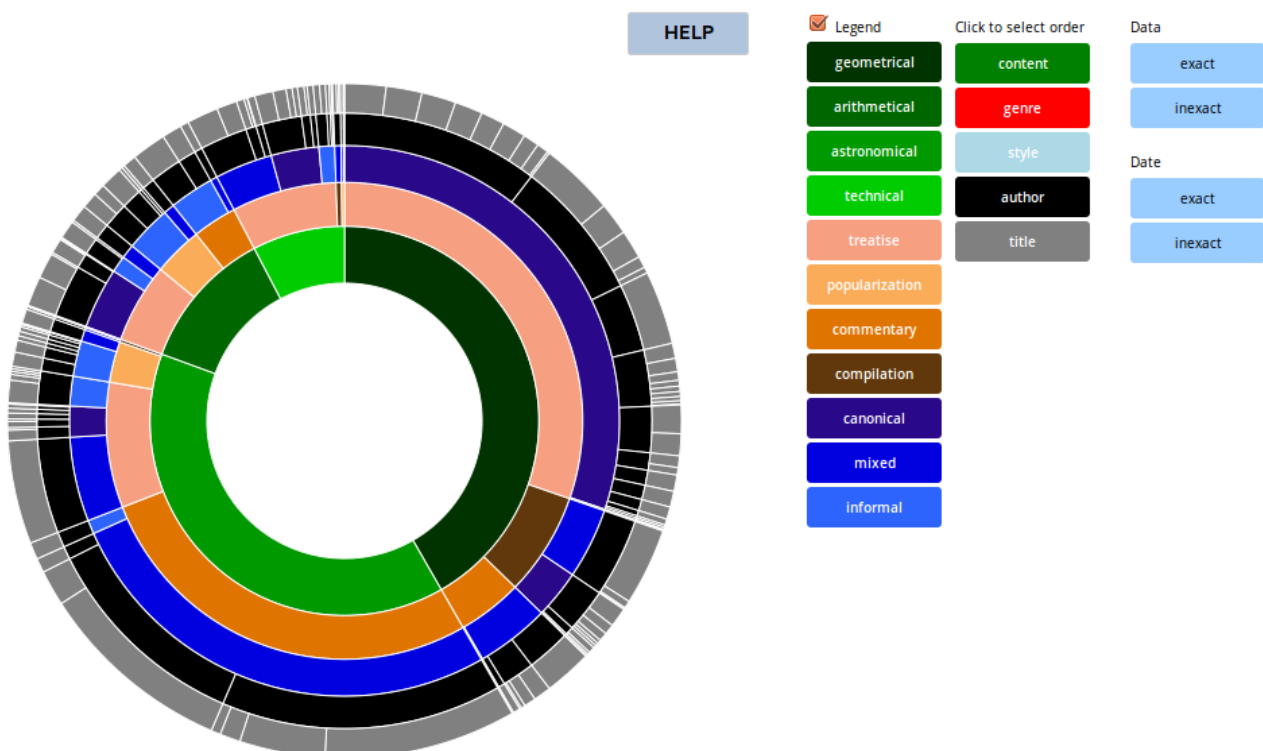


Imagen 24 - Zoomable sunburst

Mediante el botón "**HELP**" se puede acceder a una nueva ventana con las instrucciones de uso del gráfico.

### 4.2.1 Introducción

En los siguientes apartados se explicarán las diferentes partes del *zoomable sunburst* (Imagen 24). En este gráfico cada categoría está representada por un anillo del sunburst y cada anillo se divide en arcos dependiendo de la clasificación de la obra en la categoría representada. Además se añaden dos anillos adicionales que representan a los autores y a las obras respectivamente. La correspondencia entre colores y valores/categorías se puede consultar en la leyenda a la derecha del *sunburst*. Al mover el ratón sobre el gráfico se resalta la parte del gráfico recorrida (ver Imagen 26). Al mismo tiempo en la parte superior se muestra la secuencia de valores seguida (ver Imagen 25) y en la parte inferior derecha se muestran los porcentajes de lemas y caracteres correspondientes (ver Imagen 27). Los dos primeros se calculan en relación a todas las obras y los dos segundos en relación a la parte de obras seleccionada.

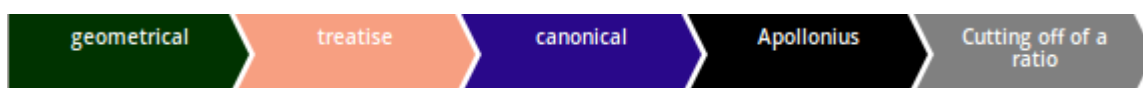


Imagen 25 - Secuencia seleccionada en el sunburst

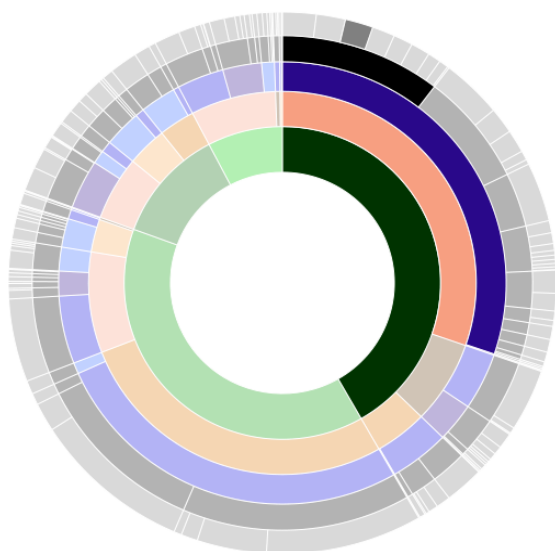


Imagen 26 - Realización de una selección

Words: 1.65%  
Chars: 1.68%  
Words (relative): 1.65%  
Chars (relative): 1.68%

Imagen 27 - Estadísticas

## 4.2.2 Exclusión de datos/fechas exactos/aproximados

Como ya se ha comentado en el caso del *timeline* los datos y las fechas de las obras del corpora pueden ser exactos o aproximados. Mediante los botones azules (Imagen 28) a la derecha de la pantalla se pueden filtrar los datos en función de estos parámetros. Al pulsar sobre alguno de estos botones éste se volverá gris y el gráfico se actualizará automáticamente con los nuevos datos.

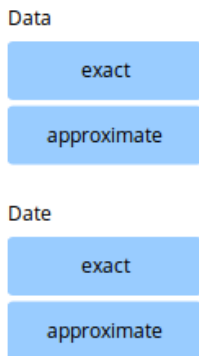


Imagen 28

## 4.2.3 Zoom

Pulsando sobre cualquiera de los arcos del *sunburst*, éste se convertirá en el nuevo centro del gráfico y automáticamente se ampliarán todos los arcos hijos de este formando nuevos círculos. De esta forma se pueden analizar más detalladamente las secciones deseadas. Los porcentajes mostrados en la parte inferior derecha del gráfico se adaptaran en consonancia.

Por ejemplo, en la Imagen 29 se ha hecho zoom sobre el valor "**canonical**" del círculo correspondiente a la categoría "**style**" facilitando la visualización de los autores y obras de este tipo.



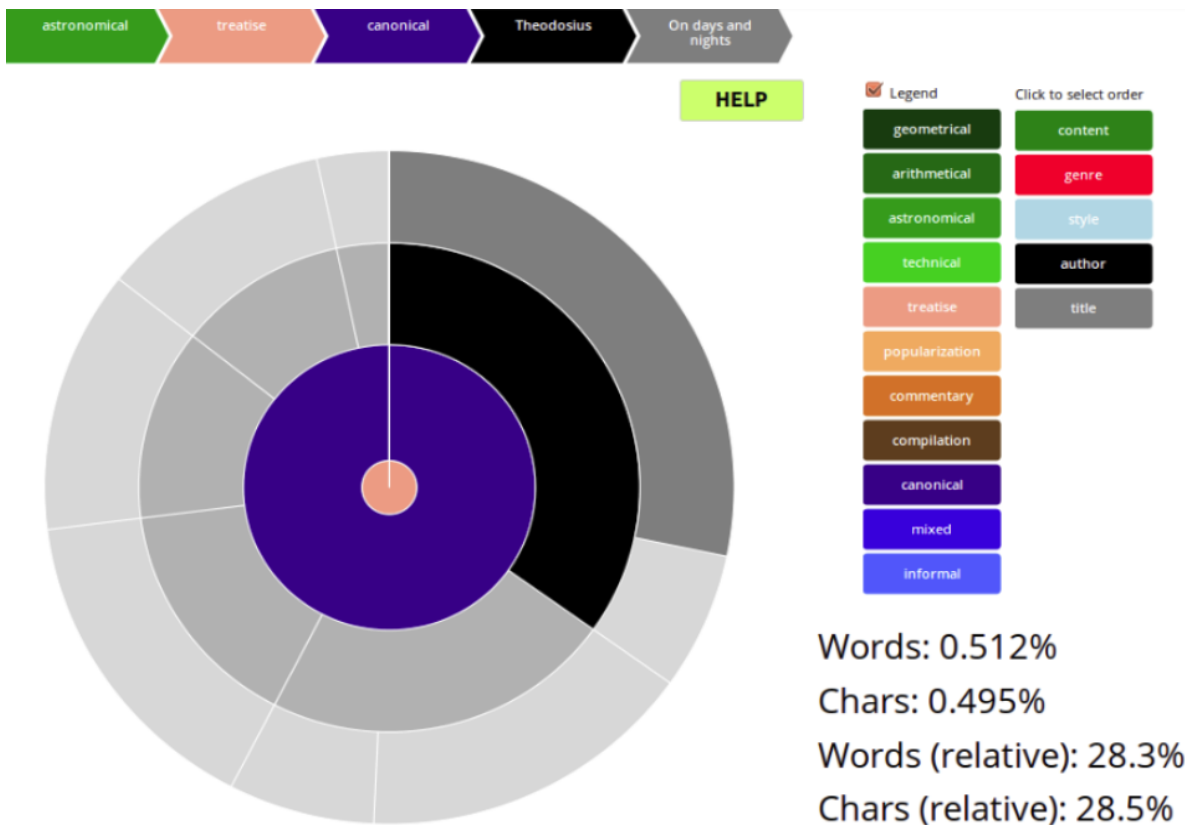


Imagen 29 - Ejemplo de zoom

#### 4.2.4 Modificación del orden

El orden por defecto de los anillos se puede modificar mediante los botones de reordenación que se encuentran entre la leyenda y los botones azules. El primer botón se corresponde con el anillo central del *sunburst* y así en adelante. Al pulsar sobre alguno de estos botones éste subirá una posición intercambiándose con el botón situado debajo (ver imágenes 30 y 31).



Imagen 30



Imagen 31



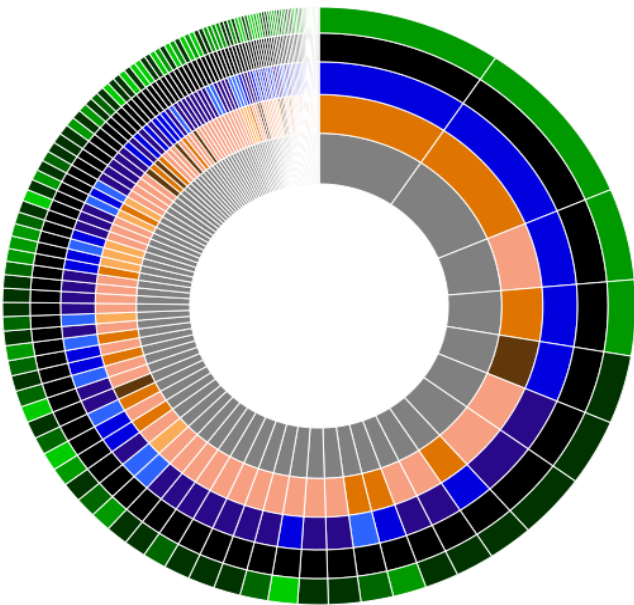
Imagen 32



Imagen 33

En caso de que se pulse sobre el primer botón, éste se intercambiará con el último (ver imágenes 32 y 33). En resumen, al pulsar sobre un botón éste subirá de posición. Siguiendo esta pauta los botones se pueden reordenar como se desee.

El gráfico se va adaptando automáticamente al nuevo orden. Por ejemplo, en la Imagen 34 - Modificación del orden de los anillos se muestran en primer lugar las obras, lo cual hace que se repitan mucho los diferentes valores de las categorías.



*Imagen 34 - Modificación del orden de los anillos*

## 5 Conclusiones

Con la implementación de los dos gráficos mediante el framework d3.js se considera cumplido el objetivo principal del proyecto, que era la obtención de una herramienta web para estudiar visualmente el *corpora* de textos técnicos de la antigüedad. Este objetivo se ha cumplido respetando los requisitos iniciales:

- El *framework* d3.js se publica con la Nueva Licencia BSD o BSD modificada (de 3 cláusulas) por lo que se trata de una herramienta de software libre.
- La aplicación se puede ejecutar sobre cualquier navegador que soporte javascript.
- Se han implementado dos gráficos diferentes y la visualización se puede ajustar dinámicamente en función de los principales parámetros disponibles en el *corpora* de obras técnicas proporcionado.

Por otra parte, se ha satisfecho el objetivo secundario de realizar un breve repaso del estado del arte y de las herramientas de visualización en el contexto del trabajo.

Respecto al framework d3.js, se puede concluir que ha cumplido las expectativas y se adecuado bien a este proyecto; sin embargo, hay que mencionar que el rendimiento decae rápidamente a medida que se incrementa la cantidad de datos a procesar y el número de muestras a calcular, pudiendo convertirse en una barrera insalvable dependiendo del caso.

Se espera que el resultado de este trabajo, además de ser útil en entornos académicos, sirva de inspiración o de base para futuras mejoras (por ejemplo, incorporando más parámetros del *corpora*, adaptándose a otros *corpora*, etc.) y también como demostración de las ventajas e inconvenientes de las herramientas de visualización utilizadas.

## 6 Referencias

[**Abramowitz, 1972**] Abramowitz, M., Stegun, I.A. (1972). *Handbook of mathematical functions*. New York: Dover Publications

[**Leech, 2013**] Leech, G. (2013). Introducing corpus annotation. En R. Garside, G. Leech, T. McEnery (Eds.), *Corpus Annotation. Linguistic Information from Computer Text Corpora* (pp. 1-18). New York: Routledge.

[**McEnery, 2001**] McEnery, T., Wilson, A. (2001). Early corpus linguistics and the Chomskyan revolution. En McEnery, T., Wilson, A. (Eds.), *Corpus Linguistics An Introduction* (pp. 1-23). Second Edition. Edinburgh: Edinburgh University Press.

[**Meeks, 2015**] Meeks, E. (2015). *D3.js in Action*. Shelter island, New York: Manning Publications Co.

[**Vitrac, 2005**] Vitrac, B. (2005). Les classifications des sciences mathématiques en Grèce ancienne. *Archives de Philosophie*, 68 (2), 269-301.

[**Vitrac, 2015**] Vitrac, B. (2015). *La transmission des textes mathématiques grecs*. Consultada el 2 de diciembre de 2015, en [https://www.academia.edu/16162595/La\\_transmission\\_des\\_textes\\_math%C3%A9matiques\\_grecs](https://www.academia.edu/16162595/La_transmission_des_textes_math%C3%A9matiques_grecs)

[**Vogt, 2014**] Vogt, V.D. (2014). *An open source browser-based software tool for graph drawing and visualisation*. Tesina de Master. Barcelona: Universitat Oberta de Catalunya.

[**Wikipedia, 2015**] Wikipedia contributors. (2015). *Normal distribution*. Consultada el 18 de noviembre de 2015, en [https://en.wikipedia.org/w/index.php?title=Normal\\_distribution](https://en.wikipedia.org/w/index.php?title=Normal_distribution)

## **ANEXO I**

El artículo encajaría en la temática de las siguientes revistas:

Revista Caracteres - <http://revistacaracteres.net>

Journal of Digital Humanities - <http://journalofdigitalhumanities.org/>