



UNIVERSIDAD NACIONAL DE COLOMBIA

Sistema de visualización para asistir la navegación dentro de un foro de discusión en línea

Andrés Mauricio Velandia Moya

Universidad Nacional de Colombia
Facultad de Ingeniería, Departamento de Ingeniería de sistemas e industrial
Bogotá D.C., Colombia
2012

Sistema de visualización para asistir la navegación dentro de un foro de discusión en línea

Andrés Mauricio Velandia Moya

Tesis o trabajo de grado presentada(o) como requisito parcial para optar al título de:
Magíster en ingeniería - ingeniería de sistemas y computación

Director(a):

PhD. Jean Pierre Charalambos Hernandez

Línea de Investigación:

Computación Gráfica

Grupo de Investigación:

RemixLab

Universidad Nacional de Colombia

Facultad de Ingeniería, Departamento de Ingeniería de sistemas e industrial

Bogotá D.C., Colombia

2012

Este trabajo esta dedicado a mi familia.

Agradecimientos

Quiero agradecer a mi familia por su comprensión y apoyo incondicional, a mi director Jean Pierre por sus valiosos consejos, a mis amigos Jaime y Alejandro por sus grandes aportes en el proyecto, y a la comunidad de *processing*, especialmente a Andrés Colubri por su asesoría y a Casey Reas por permitirnos usar los datos del foro.

Resumen

Hemos desarrollado una herramienta de visualización que permite a los usuarios de un foro de discusión navegar gráficamente a través del mismo. El proyecto incluye el diseño de una metáfora de visualización que destaca las características de los elementos del foro (categorías, sub-categorías, hilos y mensajes) y la implementación de la misma. También, se crearon filtros que permiten la interacción con el usuario y se incluyó un motor de búsqueda textual. El prototipo fue validado con datos reales tomados del foro de discusión perteneciente al proyecto open source “Processing”.

Palabras clave: Visualización de datos; Foros de discusión en línea; Interfaces de usuario.

Abstract

We have developed a visualization tool that enables the users of a discussion forum to navigate graphically through it. The project includes the design of a visualization metaphor that highlights some characteristics of the elements in the forum such as categories, subcategories, threads, and the implementation thereof. Also some filters that allow interaction with the user were created, included a textual search engine. The prototype was tested with actual data taken from the discussion forum that belongs to the open source project of the Processing language.

Keywords: Online discussion forums; Data visualization; User Interfaces.

Índice general

| | |
|--|-----------|
| 1. Introducción | 1 |
| 2. Trabajos relacionados | 5 |
| 2.1. Enfocados a la visualización de los usuarios | 5 |
| 2.1.1. People Garden | 5 |
| 2.1.2. Anthropomorphic visualization | 6 |
| 2.2. Enfocados en visualización del hilo | 7 |
| 2.2.1. BuIB - Visualizing bulletin board activity | 7 |
| 2.2.2. Astrograph for Exploring discussion Forums | 8 |
| 2.2.3. DifVis - Discussion Forum Visualizer | 9 |
| 2.2.4. Treemaps to visualize threaded Discussions Forums | 10 |
| 2.2.5. Loom | 10 |
| 3. Metodología y diseño | 12 |
| 3.1. Selección del foro a visualizar | 12 |
| 3.1.0.1. El proyecto Processing | 12 |
| 3.2. Metodología aplicada al proyecto | 13 |
| 3.2.1. Acquire | 13 |
| 3.2.2. Parse | 14 |
| 3.2.3. Filter | 16 |
| 3.2.4. Mine | 18 |
| 3.2.5. Represent | 19 |
| 3.2.6. Refine | 21 |
| 3.2.7. Interact | 23 |
| 4. Implementación y resultados | 25 |
| 4.1. Aplicación para el filtrado de los datos | 25 |

| | |
|--|-----------|
| 4.2. Herramienta de visualización | 26 |
| 4.2.1. Cálculo del color | 27 |
| 4.2.2. Cálculo de la posición por una sola espiral (diseño 1) | 28 |
| 4.2.3. Cálculo de la posición para varias espirales (diseño 2) | 28 |
| 4.2.4. Interacción del usuario | 29 |
| 4.3. Resultados | 31 |
| 5. Conclusiones y trabajo futuro | 34 |
| Bibliografía | 36 |

Índice de figuras

| | |
|---|----|
| 2.1. Representación de una comunidad en PeopleGarden [21]. Los usuarios con más cantidad de pétalos son los más importantes, debido a la cantidad de mensajes colocados por ellos. | 6 |
| 2.2. Anthropomorphic visualization [19]. Método usado para visualizar la información histórica de un individuo. | 7 |
| 2.3. Diferentes visualizaciones del sistema BuIB [18, 17] a) corresponde a la vista Temporal Thread Development, b) corresponde a la vista User Thread Participation y c) corresponde a la vista Timeline | 8 |
| 2.4. DifVis[16]. | 10 |
| 2.5. Treemaps para visualizar foros de discusión en PDAs [11]. | 11 |
| 2.6. Proyecto Loom [10]. La representación en círculos emula las conversaciones del mundo real. | 11 |
| 3.1. El archivo <i>Bugs.cat</i> contiene los nombres de las sub-categorías pertenecientes a la categoría Bugs | 14 |
| 3.2. Fragmento del archivo <i>BugFixes.txt</i> . Contiene la información de los hilos publicados en dicha sub-categoría. Las direcciones de correo electrónico fueron cambiadas por seguridad. | 15 |
| 3.3. Fragmento del archivo de mensaje <i>1028325051.txt</i> . Los campos de correo electrónico y de la dirección IP fueron cambiados para seguridad del usuario. | 15 |
| 3.4. Ejemplo archivo xml con los datos del foro. | 16 |
| 3.5. Fragmento del documento xml usado para indexar los mensajes del foro. La etiqueta doc indica el inicio del documento. | 18 |
| 3.6. Diseño 1. Se representan los hilos a través de arboles y las categorías como un bosque. | 20 |
| 3.7. Diseño 2. Se representan las categorías a través de círculos. | 20 |
| 3.8. Diseño inicial de la metáfora de visualización. | 20 |

| | | |
|-------|--|----|
| 3.9. | Segundo diseño de la metáfora de visualización. En este caso se cambió el filtro por fecha y el significado de los colores. | 21 |
| 3.10. | Imagen de la implementación inicial de la metáfora. Cada uno de los puntos representa un hilo en el foro. Se puede observar que el espacio entre los hilos no es suficiente para representar el tamaño usando el diámetro del círculo. | 22 |
| 3.11. | Tercer diseño de la metáfora. Se muestra la longitud del hilo y el número de mensajes usando la coordenada z | 22 |
| 3.12. | Diseño de la metáfora que muestra los mensajes dentro del hilo. | 23 |
| 3.13. | Cuarto diseño de la metáfora. Se modificó el diseño inicial para poder diferenciar las categorías principales del foro. | 24 |
| 4.1. | Diagrama de clase del programa usado para extraer los datos. | 25 |
| 4.2. | Arquitectura de la herramienta de visualización. | 26 |
| 4.3. | Diagrama de clases del prototipo. | 27 |
| 4.4. | Implementación del primer diseño. Los datos corresponden al foro “ <i>processing forum alpha</i> ”. Se representa el foro a través de una sola espiral. | 29 |
| 4.5. | Implementación del segundo diseño. Los datos corresponden al foro “ <i>processing forum alpha</i> ”. El foro se representa como una galaxia en espiral, donde los brazos representan las categorías principales. | 30 |
| 4.6. | Ejemplo de interacción con el usuario. En este caso se buscaron los hilos que contienen la palabra <i>bug</i> entre el 27 de junio de 2003 al 12 de mayo del 2004 en todas las categorías del foro. | 30 |
| 4.7. | Implementación navegación dentro del hilo. | 32 |

1 Introducción

Las comunidades online actuales han facilitado superar las barreras geográficas, permitiendo que personas de todos los lugares del mundo puedan comunicarse entre si, compartir sus ideas, puntos de vista y problemas entre otros. Los foros de discusión han sido parte fundamental para fomentar este intercambio de ideas.

En la actualidad, la participación en los foros de discusión se basa en su totalidad en operaciones de texto. La gran cantidad de datos almacenados en ellos podrían ser aprovechados de forma más eficiente si se guía este proceso a través de una interfaz gráfica que describa visualmente tanto el contenido almacenado, como las características del foro. Además de la búsqueda, existen otras necesidades dependiendo del rol que el usuario desempeñe en el foro, que se pueden solucionar usando una herramienta de visualización. En este trabajo diseñamos y desarrollamos esta herramienta de visualización que permite al usuario interactuar en un foro. Esta se inspiró en una galaxia espiral, donde se ubican los hilos del foro a partir de la fecha de actualización. Cada brazo de la galaxia corresponde a una categoría principal del foro.

A continuación describiremos de forma breve los foros de discusión y discutiremos las necesidades de los usuarios dependiendo del rol que desempeñan. También se definirá que es la visualización de información, así como su uso en foros de discusión. Finalmente mostraremos la organización del documento.

Foros de discusión

Los foros de discusión (también conocidos como message boards, o bulletin boards) son sitios en internet ampliamente utilizados en el mundo actual, que permiten el intercambio de conocimientos. Este proceso se basan en la publicación y consulta de mensajes de texto por parte de los usuarios.

Los foros se encuentran organizados por categorías y sub-categorías las cuales contienen

hilos de discusión (o threads). Un hilo, básicamente consiste en un conjunto de mensajes que están relacionados directamente con un tema.

Tipos de usuario de un foro

Para guiar este proyecto, definiremos claramente los usuarios y sus necesidades básicas; así podremos diseñar la metáfora que las solucione. A continuación se describen los usuarios:

Usuarios administradores

Son los creadores del foro, su principal tarea es mantenerlo ordenado y validar que la participación se realice de acuerdo a las reglas indicadas en el mismo. Entre sus necesidades básicas encontramos:

- El administrador debe validar qué hilos se pueden cerrar y cuales deben permanecer abiertos; para realizar esta acción el administrador debe identificar los hilos que no tienen mensajes publicados recientemente.
- Verificar que los hilos se encuentren en la categoría correcta.
- Identificar los hilos con mayor participación por parte de los usuarios del foro.
- Conocer las categorías que tienen una mayor o menor participación por parte de los usuarios y de esta manera validar si la organización es correcta.

Usuarios frecuentes

Son los que participan activamente y por lo tanto han sido autores de una gran cantidad de mensajes e hilos. Este usuario puede estar interesado en realizar las siguientes tareas:

- Encontrar hilos nuevos del tema de su interés.
- Identificar los hilos con mayor participación por parte de los usuarios del foro.
- Reconocer a los usuarios que comparten sus mismos intereses.

Usuarios temporales

Este tipo de usuarios no tiene una participación significativa en el foro, sino por el contrario participan en él para buscar sobre un tema particular. En este caso se identificaron las siguientes tareas:

- Encontrar rápidamente los hilos que contienen la información de interés.
- Poder ver el número de participaciones y la antigüedad en el hilo de su interés.

Foros de discusión y visualización de información

Es natural para el ser humano comunicar sus ideas a través de imágenes, y resulta útil aprovechar esa capacidad para entender conceptos e ideas difíciles de explicar con palabras.

La visualización de información es una nueva disciplina y tiene como objetivo encontrar una representación gráfica de datos abstractos y sus relaciones, de forma intuitiva, con el fin de ayudar a las personas a entender y analizar los datos. Por otro lado, también se encarga del estudio de representaciones visuales de información a gran escala, por lo que se adaptan perfectamente a nuestro objeto de estudio.

Con respecto a los foros de discusión, podemos observar claramente como la visualización de información sería de gran utilidad en ellos y como ésta podría mejorar el acceso a la información, si se encuentra una representación gráfica que permita entender mejor su contenido. En este caso, lo que se desea es mejorar la exploración a partir de una interfaz gráfica que permita acceder a la información y dar solución a las necesidades de cada usuario descritas en la sección anterior.

Con el objetivo de guiar nuestro trabajo se definieron las siguientes características ideales que debería tener el sistema para dar solución a la problemática de los foros:

- Debe tener la capacidad de conectarse a cualquier foro indicado por el usuario. En la actualidad, existe un gran número de sistemas administradores de foros. Para mencionar unos pocos se encuentra (PhpBB, YaBB, Bulletin Board, Zoho... etc). En este caso, el sistema debería permitir seleccionar el foro independiente del sistema administrador.
- Responder a las necesidades de cada uno de los usuarios descritos en la sección anterior.

- Destacar visualmente hilos que por sus características y por su contenido puedan ser más importantes para el usuario.

Organización del documento

En este documento presentamos una alternativa que intenta dar solución a las problemáticas de los usuarios, cumpliendo con algunas características descritas en la sección anterior. Debido al tiempo disponible para realizar el proyecto, este se centró en la visualización y se plantea como trabajo futuro la conexión a diferentes sistemas administradores de foros y el análisis semántico de los datos; esto debido a que es un trabajo arduo y llevaría más tiempo del que se dispone para realizar esta tesis.

El documento esta organizado de la siguiente forma:

- El capítulo 2 presentará los trabajos relacionados en el área de visualización de información aplicados a foros de discusión.
- El capítulo 3 describirá la metodología usada en el proyecto, así como el diseño de la metáfora de visualización.
- El capítulo 4 se enfocará en la implementación y en los resultados obtenidos.
- Finalmente, se discutirán las conclusiones y el trabajo futuro.

2 Trabajos relacionados

Existen varios trabajos que han querido dar solución a los problemas descritos. Hemos realizado una clasificación en dos grupos dependiendo de la información mostrada. El primero consiste en los trabajos que se enfocan en representar las características de los participantes en el foro, mientras que el segundo en como visualizar los mensajes y los hilos publicados en él. A continuación describimos algunos de ellos.

2.1. Enfocados a la visualización de los usuarios

En esta categoría encontramos los siguientes trabajos:

2.1.1. People Garden

En la investigación realizada por Xiong, R. y Donath[21] se diseñó el sistema conocido como PeopleGarden. Este usa flores como una metáfora para representar a los individuos y un jardín para el ambiente online. El objetivo de este proyecto consiste en mostrar la actividad de un individuo en la comunidad, y sus relaciones con otros usuarios.

People-Flower Representa un usuario en el foro con las siguientes características:

- Los pétalos de la flor corresponden a los mensajes publicados por el usuario.
- Para destacar los mensajes más recientes de los más antiguos se usa un color más fuerte.
- A medida que transcurre el tiempo el color de los pétalos se tornará más claro.
- Para indicar que otro usuario del foro a dado respuesta a uno de los mensajes, se usa círculos sobre el pétalo.
- El tallo de la flor indica la antigüedad del usuario en el foro.



Figura 2.1: Representación de una comunidad en PeopleGarden [21]. Los usuarios con más cantidad de pétalos son los más importantes, debido a la cantidad de mensajes colocados por ellos.

People-Garden Se encarga de visualizar todos los usuarios de la comunidad. Los usuarios más importantes se destacan por número de pétalos de la flor que los representa. La Figura 2.1 muestra este componente con más detalle.

2.1.2. Anthropomorphic visualization

En este trabajo realizado por Perry, E. y Donath, J. [19] se propone una forma de representar la información histórica de un individuo en línea a través de la figura humana. Con base en esto se usan las diferentes partes del cuerpo para representar emociones, porcentajes de respuestas, número de mensajes escritos etc. La figura 2.2 muestra como se realiza la visualización.

Cada cuadro en el cuerpo representa los mensajes que han sido colocados por el usuario, los ojos abiertos indican que porcentaje de mensajes escritos por él son respuestas a mensajes de otros y la separación de sus piernas indica que ha recibido varias respuestas de otros usuarios. Otras características como el color y la expresión del rostro indican el estado emocional del usuario, el cual es generado por el tono de sus mensajes. El método hace énfasis en la visualización de la información histórica de los usuarios, y permite identificar más fácilmente las personas que tienen más importancia en la comunidad, pero no es adecuado para la búsqueda de mensajes.

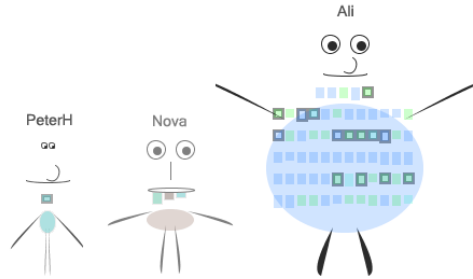


Figura 2.2: Anthropomorphic visualization [19]. Método usado para visualizar la información histórica de un individuo.

2.2. Enfocados en visualización del hilo

2.2.1. BulB - Visualizing bulletin board activity

BulB es un desarrollo que permite mejorar la visualización de la comunidad a través de analogías, con ello se pretende facilitar la participación de los usuarios en el foro [17, 18]. Este sistema representa los hilos de discusión, por medio de tallos, en el que su altura indica el tiempo total en que el hilo ha permanecido activo. Para representar la información del usuario, se usa un color que lo identifica dentro del foro. A pesar de visualizar la información de los usuarios, este trabajo se encuentra enfocado a la visualización de la información histórica de cada hilo.

Este sistema posee tres tipos de visualizaciones, las cuales se describirán a continuación.

Temporal Thread Development Permite ver los mensajes colocados en el thread por los usuarios en la parte superior del tallo. Los mensajes son mostrados teniendo en cuenta el tiempo en el cual fueron publicados.

User Thread Participation Muestra el porcentaje de participación de cada usuario en el thread, por medio de una gráfica circular (o de pastel).

Timeline Cada mensaje es ubicado en el tallo, dependiendo de la fecha en la cual se publicó.

Estos tipos de visualización se pueden apreciar en la figura 2.3.

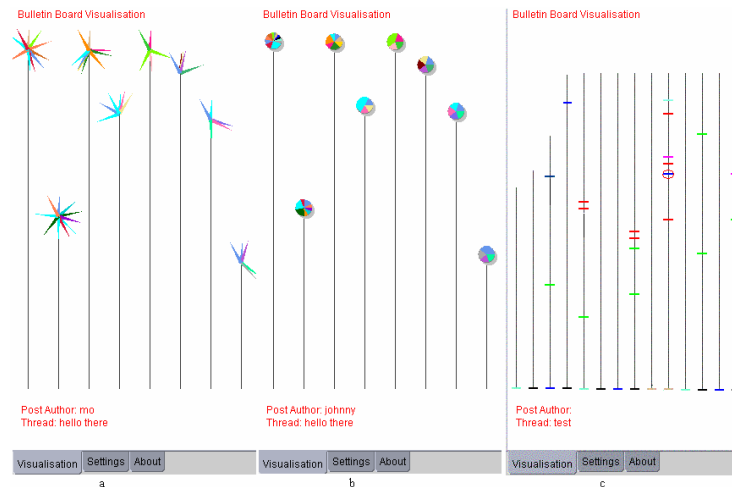


Figura 2.3: Diferentes visualizaciones del sistema BuIB [18, 17] a) corresponde a la vista Temporal Thread Development, b) corresponde a la vista User Thread Participation y c) corresponde a la vista Timeline

2.2.2. Astrograph for Exploring discussion Forums

Existe otra investigación desarrollada por Kim [15] en la cual se describe el sistema astrograph for discussion forum o (ADF), este intenta solucionar el problema de la visualización de información en foros de discusión online. Para ello, ADF transforma el texto contenido en una abstracción visual, usando gran cantidad de atributos como color, intensidad y posición. Los autores proponen un método efectivo para mejorar las búsquedas en los foros de discusión haciendo uso de Filtros colaborativos, en donde los mismos usuarios ayudan a otros a encontrar los artículos, para ello se asume que existen usuarios con las mismas preferencias [15]. Sin embargo el uso de este tipo de búsqueda solo es eficiente si muchos usuarios ingresan sus opiniones sobre un artículo importante. Por esto, en el sistema ADF se decidió filtrar la información haciendo uso de filtros colaborativos, pero en lugar de apoyarse en los comentarios de los usuarios, se basa en el historial de los mismos. Adicional al filtrado de mensajes, ADF propone un método de visualización. Este método se divide en tres partes que serán tratadas a continuación.

Visualizando la dimensión de los threads Cada uno de los hilos (conversaciones o threads) puede contener un gran número de mensajes, dependiendo de que tan popular sea el tema que se está comentando, por ello se representa cada threads por medio de un círculo y el número de mensajes a través de el radio.

Visualizando la frecuencia de las referencias Al contrario de los filtros colaborativos, los cuales se basan en opiniones dadas por los lectores sobre un mensaje en particular, el autor sugiere realizar este filtro a través del número de referencias hacia un hilo en particular. Los hilos más referenciados por los usuarios son más importantes y por esto son los mas brillantes en la visualización (tienen más intensidad de brillo).

Visualización del tiempo de publicación Para visualizar el tiempo de vida de una discusión los autores usaron el atributo color, esto se logró clasificado cada hilo en un grupo. Existen tres tipos de grupos disponibles, cada uno de los cuales esta asignado a un color primario, y el tiempo es representado introduciendo una mezcla de colores.

Filtro basado en contenido El filtrado por contenido es de vital importancia para el usuario, por ello este sistema analiza los mensajes, obteniendo las palabras más relevantes en el hilo y luego mostrándolas al usuario. Este proceso es computacionalmente costoso, por ello utilizaron dos procedimientos de búsqueda de información, el primero conocido como stop-word removal, cuyo objetivo principal es eliminar las palabras que sirven para establecer el contexto, y el algoritmo suffix stemming, este agrupa varias palabras relacionadas en una sola. Finalmente el ADF analiza la frecuencia de dichas palabras en el documento, y con base a eso se muestran sobre el correspondiente objeto. El propósito es permitir al usuario un método por el cual pueda navegar a través de sus temas de interés con facilidad.

Múltiples Vistas Escaladas Este tipo de vistas permite a los usuarios, lograr entender los datos presentados, y sus relaciones. ADF provee tres tipos de vistas. La primera es una vista general del foro, en donde el usuario puede ver todos los threads del foro. En esta vista el usuario puede seleccionar el área de su interés. La segunda vista corresponde al área de interés seleccionada por el usuario, en ella se muestra con más detalle la información contenida en el thread (incluyendo palabras claves, título, e información adicional del hilo).

2.2.3. **DifVis - Discussion Forum Visualizer**

DifVis es otra propuesta realizada por Kim y Johnson[16] para visualizar foros de discusión, que continúa con el trabajo realizado en Astrograph Discussion Forum [15]. En ese trabajo el autor propone el uso de cuadrados para representar cada hilo. El tamaño de

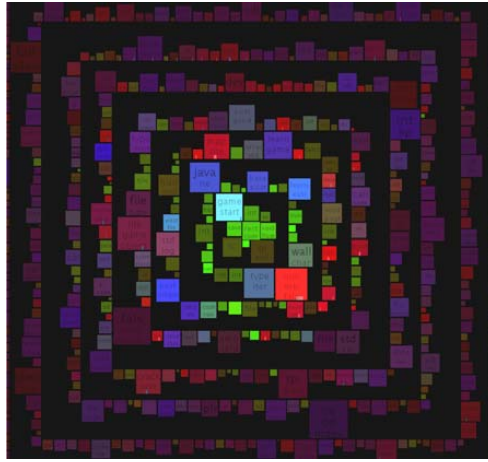


Figura 2.4: DifVis[16].

dichos cuadrados representa el tamaño relativo del hilo medido en número de palabras (sin incluir stop-words). Los colores son usados para representar la popularidad (color verde), la actividad (color azul) y el tiempo (color rojo) de cada hilo. Por último la ubicación del hilo indica su importancia relativa dentro del foro. Ver figura 2.4.

2.2.4. Treemaps to visualize threaded Discussions Forums

En este trabajo los autores proponen el uso de treemaps con el propósito de facilitar el uso de foros de discusión en dispositivos móviles PDAs [11]. El tamaño de cada recuadro del treemap corresponde al número de mensajes que contiene el hilo, mientras que los colores son usados para representar los resultados de una búsqueda (indicando el hilo más relevante) o para indicar la actividad en los hilos. Ver figura 2.5.

2.2.5. Loom

Loom fue un proyecto desarrollado por [10, 7] para visualizar la información contenida en usenet (un sistema para compartir noticias muy similar a los foros de discusión actuales). Para ello, loom organizaba los usuarios a través de círculos por los temas que discutían (es decir que cada círculo correspondía a un hilo). Los usuarios son representados por cuadrados cuyo tamaño representa sus características. Al mostrar los hilos en forma circular se pueden identificar los temas con mayor relevancia en la comunidad o los más discutidos en ese momento. Los mensajes publicados, se muestran desde el usuario que los publicó

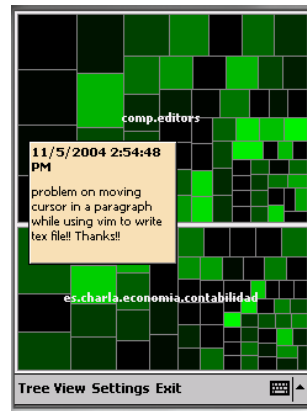


Figura 2.5: Treemaps para visualizar foros de discusión en PDAs [11].

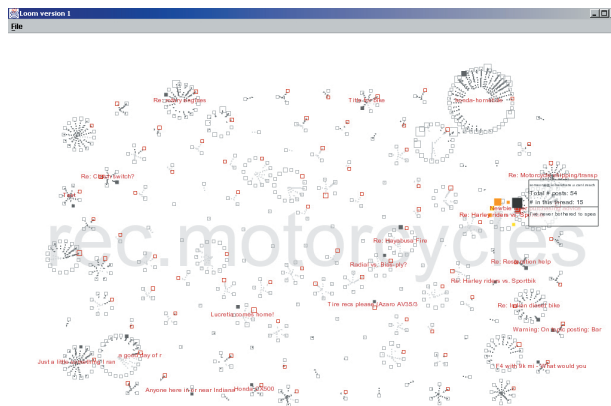


Figura 2.6: Proyecto Loom [10]. La representación en círculos emula las conversaciones del mundo real.

hacia el centro del círculo. Se pueden ver todas estas características en la figura 2.6.

3 Metodología y Diseño de la metáfora de visualización

3.1. Selección del foro a visualizar

Debido al gran número de sistemas administradores de foros existentes, se simplificó el desarrollo del prototipo para que se enfocara en uno solo. En nuestro caso, se seleccionó el primer foro del proyecto open source “Processing” (llamado processing forum alpha [13]) el cual se encuentra implementado en el sistema administrador de foros YaBB [6].

3.1.0.1. El proyecto Processing

El proyecto open source “Processing” surgió en el año 2001 y su objetivo principal consistía en hacer sketches (o bocetos) que sean veloces e interactivos. Otro objetivo era crear un lenguaje de programación que sea fácil de entender por los estudiantes de artes y que facilitara el trabajo con gráficos para los estudiantes técnicos [20, 14]. Sin embargo, actualmente processing se ha convertido en una herramienta que permite obtener gráficos con un acabado profesional.

Processing se puede dividir en dos partes principales. La primera consiste en un lenguaje de programación basado en Java, que permite realizar animaciones e imágenes de forma sencilla. Gracias a esto, es posible integrar processing a cualquier proyecto java existente. La segunda parte consta de una herramienta de software que ayuda a copilar el código realizado en ese lenguaje.

En nuestro proyecto se usó processing con dos objetivos. El primero como conjunto de datos para probar el prototipo desarrollado y segundo como soporte para implementar la metáfora de visualización diseñada.

3.2. Metodología aplicada al proyecto

Para el desarrollo de este proyecto se decidió seguir la metodología planteada por Ben Fry en [12] para la construcción de visualizaciones interactivas. Esta se encuentra dividida en 7 pasos que se describirán a continuación:

Acquire Consiste en adquirir los datos que se desean visualizar.

Parse Una vez adquirido los datos se debe reconocer la estructura de los mismos.

Filter En este paso se eliminan los datos que no son necesarios para la visualización.

Mine Consiste en realizar estadísticas básicas o minería de datos para resumir y/o encontrar relaciones entre ellos.

Represent Seleccionar la representación visual para los datos.

Refine Hacer ajustes a la representación seleccionada en el paso anterior.

Interact Agregar métodos que permitan al usuario controlar la forma en que se muestran los datos.

En esta sección describiremos cada uno de los pasos aplicados a nuestro caso de estudio.

3.2.1. Acquire

Para obtener los datos del foro, se contemplaron las siguientes 3 opciones:

- Obtener los datos a través de un plugin diseñado para dicho propósito. Este sería el caso ideal, sin embargo en este momento no es posible debido a que no ha sido implementado por parte de la comunidad que soporta el sistema administrador de foros (en el caso de processing es YaBB [6]) y como se mencionó en la introducción implementar esta funcionalidad se encuentra fuera del alcance del proyecto.
- Obtener los datos a través del uso del proceso web scraping. Este permite extraer información de un sitio web, simulando la exploración humana a partir del uso de

```
Bugs
Proce55ing_software_bugs
Proce55ing_website_bugs
BugFixes
```

Figura 3.1: El archivo *Bugs.cat* contiene los nombres de las sub-categorías pertenecientes a la categoría Bugs

implementación a bajo nivel del protocolo http [5]. Web scraping se enfoca en transformar el código html en estructuras que sean fáciles de entender y de almacenar localmente. Sin embargo tiene el inconveniente que en caso que se cambie la estructura de la página se debe actualizar el proceso de extracción.

- Obtener los datos directamente de la fuente. Esto es adquirirlos de la comunidad ya sea a través de archivos de texto o accediendo directamente a la base de datos del foro. Esta fue la alternativa que seleccionamos, ya que como se mencionó en la introducción nuestro enfoque no era la extracción, sino la metáfora de visualización.

Los datos fueron obtenidos directamente de los administradores del foro de la comunidad de Processing. Por razones de privacidad, se nos indicó no usar los datos personales de los usuarios. Los datos se recibieron en archivos de texto codificados en formato iso-8859-1 y se describirán en la siguiente sección.

3.2.2. Parse

Los archivos del foro a analizar se encuentran organizados en dos carpetas principales: Boards y Messages.

Carpeta *Boards* En esta carpeta encontramos la información de las categorías y sub-categorías del foro, organizada en archivos de texto.

Los archivos que tienen la extensión *.cat* contienen la información de la categoría principal y de las sub-categorías que la componen.

El nombre de la categoría se encuentra en la parte superior de archivo, mientras que las sub-categorías en la parte final, separados por un espacio (ver ejemplo en la figura 3.1). En total se recibieron 6 archivos de este tipo correspondientes a las 6 categorías principales.

```
1102885084|problem with jikes|artkast@mail@gmail.com|12/12/04 at 21:58:04|0|artkast|xx|0
1097363967|editor initial gap|flight404@mail1@flight404.com|10/10/04 at 01:19:15|3|flight404|xx|0
```

Figura 3.2: Fragmento del archivo *BugFixes.txt*. Contiene la información de los hilos publicados en dicha sub-categoría. Las direcciones de correo electrónico fueron cambiadas por seguridad.

```
code exchange|adrien@mail@boring.ch|08/02/02 at 23:50:51|adrien|xx|0|192.168.0.247|Well, I would
suggest that the message board needs a &quot;code area&quot;, where people can UL the stuff they
make, bug others for comments, critiques, debugging, etc.<br><br>Don't know if P5 will run in the
browser like DBN did, but it would be great to be able to preview stuff.<br><br>It could get wild if
one could give write access to files – collaborative editing!<br><br>Yes, I will only suggest things
which require lots of work ;-)||
Re: code exchange|adrien@mail@boring.ch|08/02/02 at 23:53:52|adrien|xx|0|192.168.0.247|uh, maybe
that's what the &quot;examples&quot; and &quot;tools&quot; areas are...<br><br>Gee, I'm really
making the impression of an intelligent and observant participant! I promise to keep up the good work!||
```

Figura 3.3: Fragmento del archivo de mensaje *1028325051.txt*. Los campos de correo electrónico y de la dirección IP fueron cambiados para seguridad del usuario.

La información de los hilos se encuentra almacenada en los archivos con la extensión *.txt* separada por el caracter “/”. Entre la información contenida encontramos el id y el nombre del hilo publicado, el nombre de usuario y su correo electrónico, y la fecha de la última publicación. (ver ejemplo en la figura 3.2). Se recibieron 22 archivos de este tipo, correspondientes a las sub-categorías del foro.

Carpeta Messages Como su nombre lo indica, la carpeta Messages contiene la información de los mensajes publicados en el hilo. Esta es almacenada en archivos de texto con el nombre *<id_hilo>.txt*. Al igual que en el caso anterior, los datos se encuentra separados por el caracter “/”. Entre ellos encontramos el título del mensaje, la fecha en que se publicó, el nombre y el correo electrónico del usuario, la dirección IP (pública) desde donde se publicó y el texto codificado en html (Ver ejemplo figura 3.3). En este caso se encontraron 3305 archivos de texto, correspondientes a los hilos del foro.


```

<forum>
  <category>
    <categoryName>Programming Questions & Help</categoryName>
    <categoryId>100</categoryId>
    <categoryMessageNumber>4585</categoryMessageNumber>
    <subcategories>
      <category>
        <categoryName>Syntax</categoryName>
        <categoryId>101</categoryId>
        <categoryMessageNumber>2780</categoryMessageNumber>
        <threads>
          <thread>
            <threadId>1030969228</threadId>
            <threadTitle>field</threadTitle>
            <threadUser>martin</threadUser>
            <threadDate>09-01-02 at 22:19:30</threadDate>
            <messages>
              <message>
                <messageId>1</messageId>
                <messageUser>Martin</messageUser>
                <messageDate>08-27-02 at 15:03:22</messageDate>
                <messageText>hi. </messageText>
              </message>
            </messages>
          </thread>
        </threads>
      </category>
    </subcategories>
  </category>
</forum>

```

Figura 3.4: Ejemplo archivo xml con los datos del foro.

3.2.3. Filter

Debido a las restricciones dadas por la comunidad de processing referente a los datos y a que no todos ellos eran necesarios en el proyecto, fue necesario realizar un proceso de filtrado. Por lo que se decidió almacenar dichos datos filtrados en un archivo con el formato *xml* para facilitar las operaciones de lectura y también para no depender de la estructura de archivos del sistema YaBB, de esa forma si se va a usar nuestra aplicación con otro sistema administrador de foros, lo único que se debe hacer es generar un archivo xml con la estructura que definimos.

El archivo xml mantiene el orden del foro (categorías, sub-categorías, hilos y mensajes) adicionando los siguientes campos nuevos (ver en la figura 3.4):

- Número de mensajes por categoría.
- Id del mensaje para poder identificarlo en el hilo.
- Id de la categoría. En este caso el id además de identificar la categoría se usa para diferenciar las categorías principales de las sub-categorías.

Proceso de filtrado

Antes de realizar el proceso de filtrado, se modificó la codificación de los archivos de iso-8859-1 a UTF-8 para no tener inconvenientes con los caracteres especiales. Esta operación se realizó por medio de un script en linux que realizó la conversión usando el programa *iconv* [3].

Luego de codificar los archivos, se seleccionó la información necesaria para cada elemento del foro la cual describiremos a continuación:

Categoría

- Nombre de la categoría
- Sub-categorías que la componen

Hilo

- Id
- Fecha de publicación
- Sub-categoría a la que pertenece
- Nickname del usuario que lo publicó

Mensaje

- Fecha de publicación
- Mensaje de texto
- Nickname del usuario que lo publicó

Usuario

- Nickname del usuario

Por último, se construyó una aplicación en el lenguaje Java encargada de crear el archivo xml con los datos filtrados. El diseño de la aplicación se describirá con mayor detalle en la sección 4.1.

```

<add>
  <doc>
    <field name="id">1028325051.1</field>
    <field name="username">adrien</field>
    <field name="date">Fri Aug 02 23:50:51 COT 2002</field>
    <field name="message">Well, I would suggest that the message board needs a \&quot;code area\&quot;, where people can UL the
stuff they make, bug others for comments, critiques, debugging, etc. Don&apos;t know if P5 will run in the browser like DBN did, but it
would be great to be able to preview stuff. It could get wild if one could give write access to files -- collaborative editing! Yes, I will
only suggest things which require lots of work ;-)</field>
  </doc>
  <doc>
    <field name="id">1028325051.2</field>
    <field name="username">adrien</field>
    <field name="date">Fri Aug 02 23:53:52 COT 2002</field>
    <field name="message">uh, maybe that&apos;s what the \&quot;examples\&quot; and \&quot;tools\&quot; areas are... Gee,
I&apos;m really making the impression of an intelligent and observant participant! I promise to keep up the good work!</field>
  </doc>
</add>

```

Figura 3.5: Fragmento del documento xml usado para indexar los mensajes del foro. La etiqueta doc indica el inicio del documento.

3.2.4. Mine

Debido a que en el proyecto se trabajó con todos los datos almacenados en el foro, no fue necesario la elaboración de estadísticas sobre estos. Sin embargo, se realizó una reorganización de los hilos por la fecha de su última publicación.

Otra tarea importante consistió en aplicar técnicas de recuperación de información para permitir las búsquedas textuales dentro del foro. Para ello se utilizó el proyecto open source *apache solr* [1], que contiene una gran cantidad de algoritmos de búsqueda ya implementados.

Para que apache solr pueda realizar las búsquedas, primero se deben agregar e indexar los documentos que se retornarán al momento de realizarlas. En nuestro caso los documentos sobre los cuales queremos buscar son los mensajes de texto ingresados por los usuarios. Esta tarea se realiza a través de un documento xml formado con la estructura mostrada en la figura 3.5, en donde cada mensaje se encuentra encapsulado entre las etiquetas `<doc>` y `</doc>` y contiene los siguientes campos:

- *id*: Corresponde al identificador único del documento. En nuestro caso se usó el id del hilo concatenado con el número del mensaje (por ejemplo “1028325051.1” corresponde al primer mensaje del hilo identificado con el id 1028325051).
- *username*: Indica el usuario que publicó el mensaje.
- *date*: La fecha en que se publicó el mensaje.
- *message*: El mensaje de texto con el cual se va a indexar el documento.

Para crear el documento xml se adaptó el software utilizado en la sección anterior (ver sección 4.1 para más detalles).

Finalmente antes de indexar los datos, se debe modificar el archivo de configuración de apache solr (llamado schema.xml) para indicar los campos que se desean indexar y el tipo de índice que se quiere crear. Para el proyecto se indexaron los campos *message* y *username*. También, se agregó un filtro para eliminar los *stop-words* y el código html que contienen los mensajes.

3.2.5. Represent

En esta etapa se planteó un diseño que cumpliera con las necesidades de los usuarios del foro, para ello consideramos incluir las siguientes características:

- Permitir ver los hilos con mayor participación por los usuarios.
- Mostrar la historia de los hilos del foro.
- Identificar los hilos más antiguos, así como los más actuales.
- Filtrar los hilos por un rango de fechas.
- Observar las categorías con mayor participación por parte de los usuarios.

Basados en estas características se diseñaron varias versiones de la metáfora de visualización como las de la figura 3.6 y 3.7. Estas fueron descartadas por no cumplir con varias de las características descritas. La metáfora seleccionada se muestra en la la figura 3.8, la cual consiste en una espiral, en donde se ubican los hilos dependiendo de la fecha de su última actualización. Los hilos se representan a través de círculos, cuyo diámetro indica su tamaño y el color la actividad dentro del foro. Sin embargo, este diseño no cumplía con varias de las características señaladas por las condiciones iniciales (por ejemplo, no se podía identificar la categoría a la que pertenece el hilo), por lo cual se realizaron los siguientes ajustes: Primero, se descartó el filtro por años y en su lugar se agregó un control que permitiera seleccionar una ventana de tiempo. Segundo, se cambió el uso del color para identificar la categoría a la que pertenece el hilo. Estos detalles se pueden apreciar en la figura 3.9.

Con base en este segundo diseño se inició la construcción de la herramienta de visualización en el lenguaje java. (Los detalles del prototipo se tratarán en la sección 4.2).

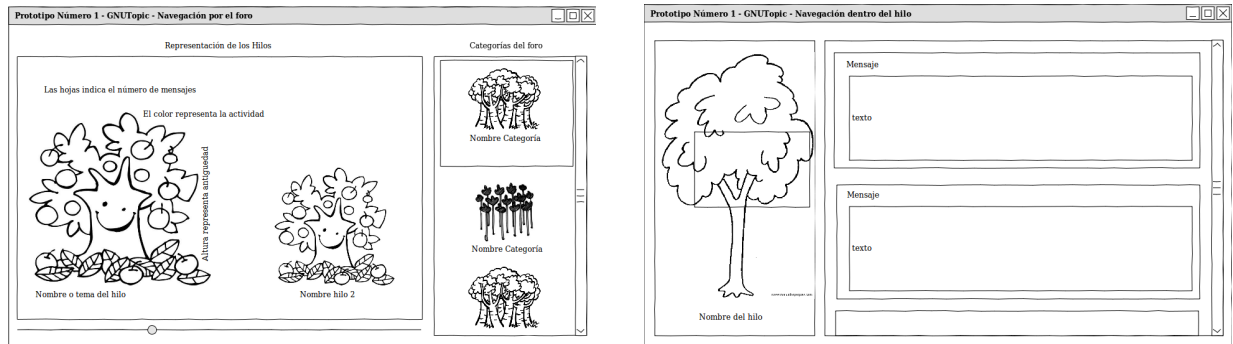


Figura 3.6: Diseño 1. Se representan los hilos a través de arboles y las categorías como un bosque.

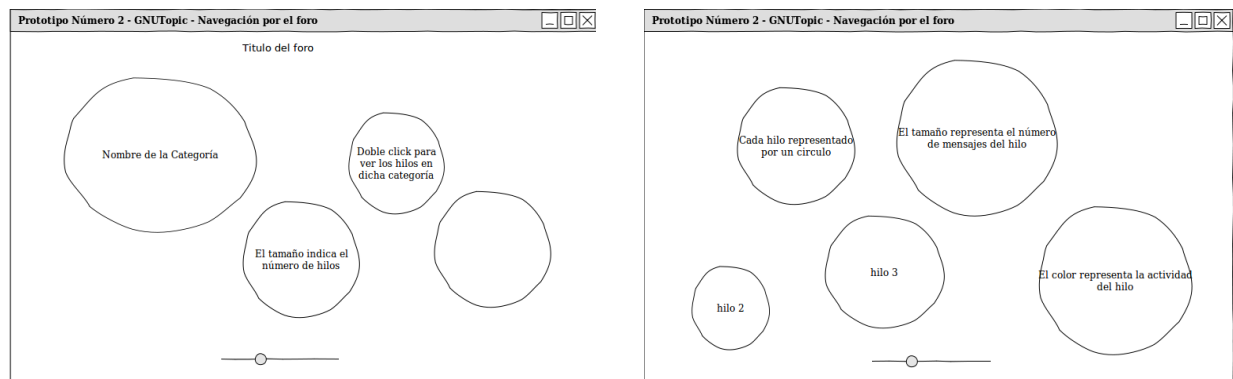


Figura 3.7: Diseño 2. Se representan las categorías a través de círculos.

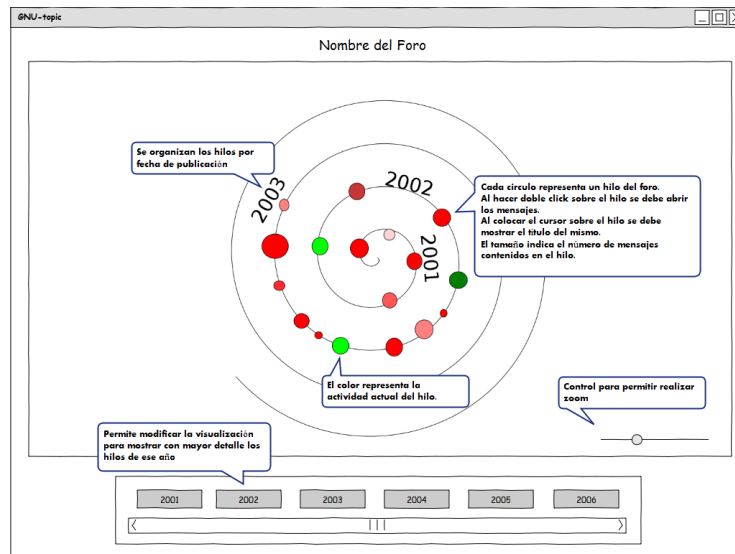


Figura 3.8: Diseño inicial de la metáfora de visualización.

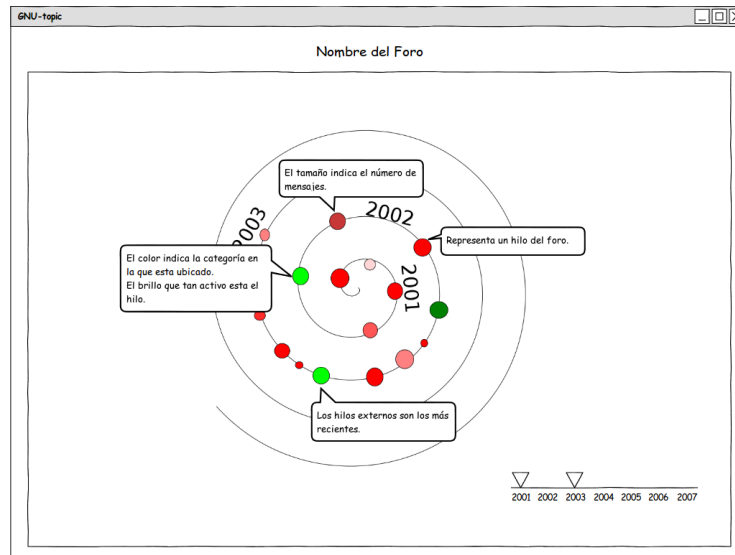


Figura 3.9: Segundo diseño de la metáfora de visualización. En este caso se cambió el filtro por fecha y el significado de los colores.

La forma en espiral se seleccionó debido a que indica al usuario que todos los hilos giran al rededor de un tema principal y permite destacar el tiempo en que transcurren las conversaciones. También, permite aprovechar mucho mejor el espacio en pantalla.

En la primera versión se encontró que el uso del diámetro para representar el tamaño del hilo no era adecuado, ya que los hilos se encontraban muy juntos (un ejemplo se puede ver en la figura 3.10), por lo que se modificó la metáfora para que los hilos se representaran a través de un conjunto de puntos que cuelgan de la espiral (similar a la visualización de mensajes mostrada en [18, 17]) y que representan los mensajes, de esta forma se pueden ver tanto el historial del hilo como el número de mensajes que lo componen. Esta modificación se puede ver en la figura 3.11.

En cuanto a la visualización de los mensajes en el hilo, estos se muestran usando la metáfora de la figura 3.12, donde cada círculo representa un mensaje publicado en el hilo y los colores identifican al usuario que publicó el mensaje dentro del hilo.

3.2.6. Refine

Al finalizar la primera etapa de implementación, se descubrió que la espiral generada era demasiado grande y por lo tanto era confuso identificar las categorías y sub-categorías del foro. También, se encontró que los hilos se sobreponen y por lo tanto se hacen difíciles

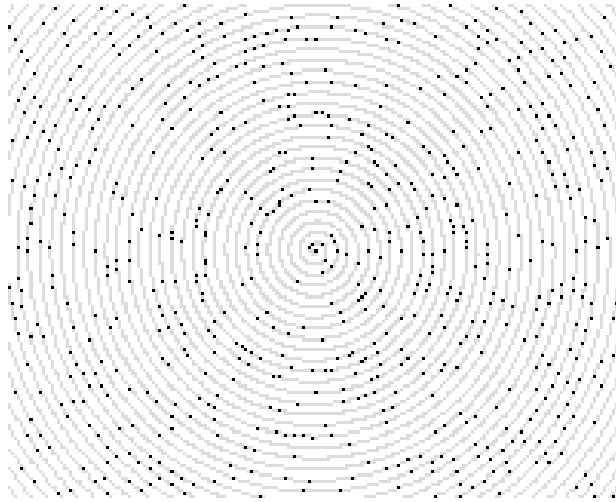
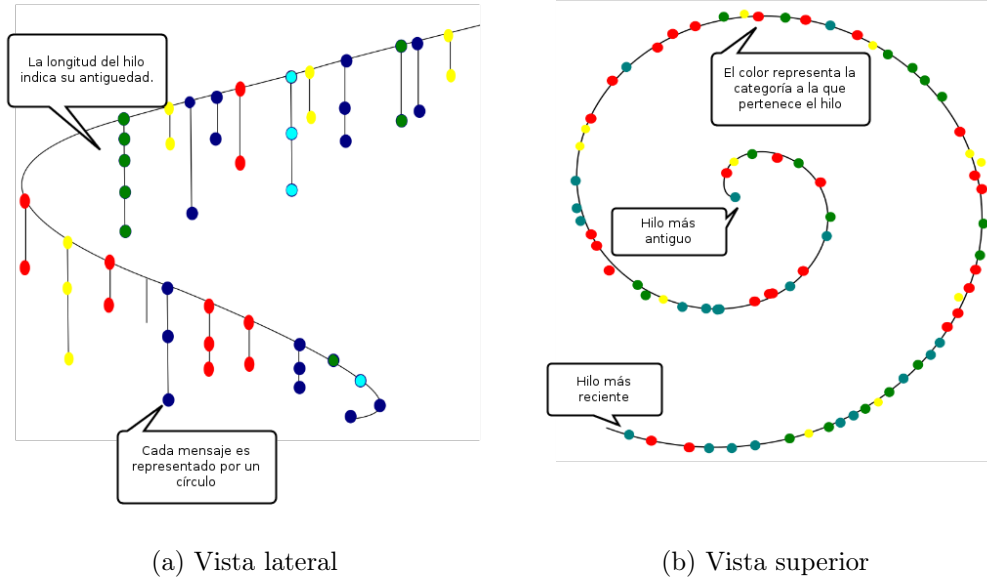


Figura 3.10: Imagen de la implementación inicial de la metáfora. Cada uno de los puntos representa un hilo en el foro. Se puede observar que el espacio entre los hilos no es suficiente para representar el tamaño usando el diámetro del círculo.



(a) Vista lateral

(b) Vista superior

Figura 3.11: Tercer diseño de la metáfora. Se muestra la longitud del hilo y el número de mensajes usando la coordenada z .

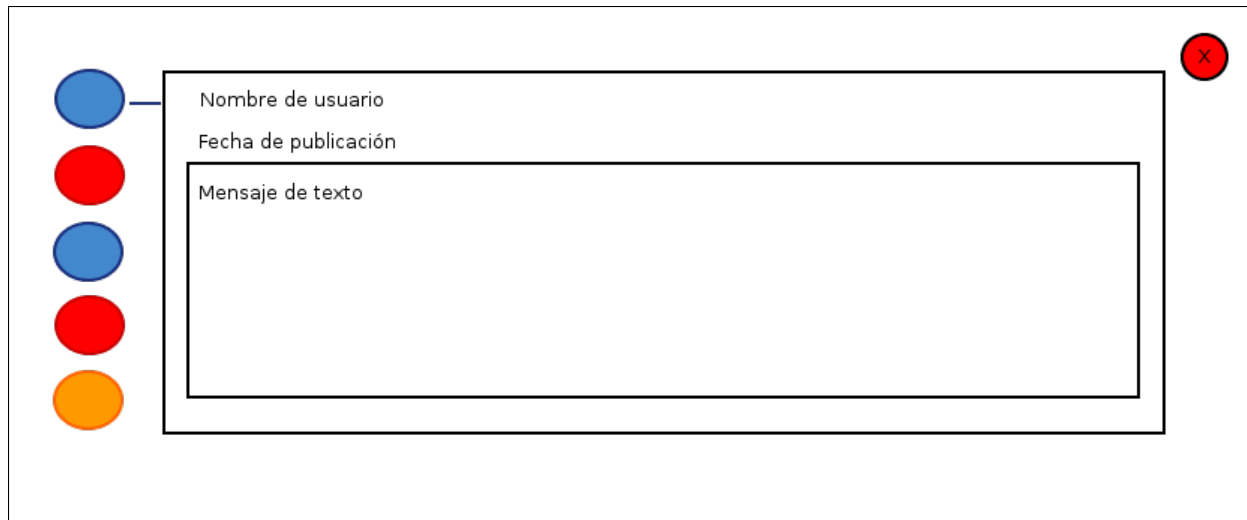


Figura 3.12: Diseño de la metáfora que muestra los mensajes dentro del hilo.

de seleccionar por los usuarios. Por esta razón se modificó la metáfora para que simulara una galaxia en espiral (ver definición en [4]), en donde cada brazo representa una categoría principal (ver figura 3.13).

3.2.7. Interact

La visualización por si sola, no basta para resolver las necesidades de los usuarios. Por este motivo se diseñaron varias funcionalidades para que el usuario pueda interactuar con la herramienta. A continuación describiremos cada una de ellas:

Manejo de la cámara El usuario puede manejar la cámara de la escena acercando o alejando la visualización. Además, puede rotar la visualización verticalmente para poder ver la longitud de los hilos y girar la espiral horizontalmente para ver los hilos que se encuentran en el fondo de la espiral.

Selección de la visualización El usuario puede seleccionar la visualización que desee (una sola espiral o varias espirales dependiendo del número de categorías).

Cambiar el nivel de detalle de la visualización Esta tarea consiste en aumentar o decrementar el espacio entre los hilos.

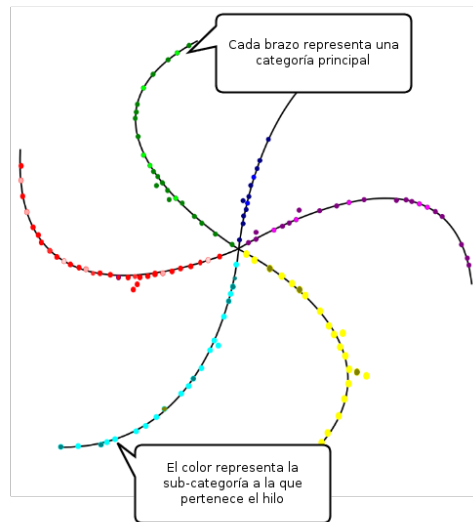


Figura 3.13: Cuarto diseño de la metáfora. Se modificó el diseño inicial para poder diferenciar las categorías principales del foro.

Filtrar los hilos por categoría y sub-categoría El usuario debe seleccionar la categoría o sub-categoría que desea visualizar. Los hilos que no pertenecen a la categoría seleccionada por el usuario se tornarán transparentes.

Filtrar los hilos por fecha de la última publicación Seleccionar la fecha mínima y máxima de publicación. Al igual que en caso anterior los hilos que no cumplan con esta condición se volverán transparentes.

Filtrar los hilos por búsqueda de texto El usuario puede realizar búsquedas textuales dentro del foro.

Selección del hilo El usuario tiene la posibilidad de seleccionar los hilos deseados y ver su contenido.

Selección de los mensajes Por limitaciones en el espacio, el usuario podrá ver los mensajes solo cuando ubique el cursor sobre el círculo que los representa.

4 Implementación y resultados

En este capítulo describiremos de forma breve el diseño de la herramienta de visualización y mostraremos los resultados obtenidos. Como se mencionó en la sección anterior, fue necesario crear una aplicación adicional encargada de construir el archivo con los datos filtrados y de la construcción del archivo para indexar los datos en apache solr. A continuación describiremos con mayor detalle las dos aplicaciones.

4.1. Aplicación para el filtrado de los datos

Basándonos en la organización del foro, se creó un programa en el lenguaje Java que lee cada una de las categorías y las almacena en memoria, manteniendo el orden de los hilos y mensajes. Esto se logró, creando una estructura de datos similar a la del foro estudiado. El diseño se puede observar con mayor detalle en el diagrama de clases de la figura 4.1. La clase *ForumCategory* representa la categoría del foro, la clase *ForumThread* el hilo, la clase *ForumMessage* los mensajes y por último la clase *CSVDataManager*, encargada de leer la información contenida en los archivos y almacenarla en la estructura.

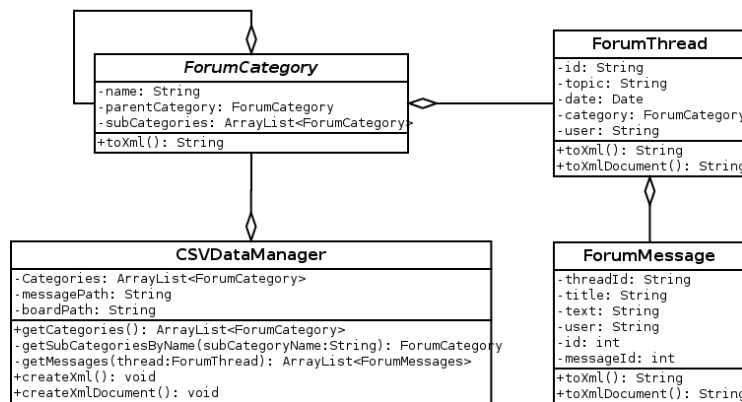


Figura 4.1: Diagrama de clase del programa usado para extraer los datos.

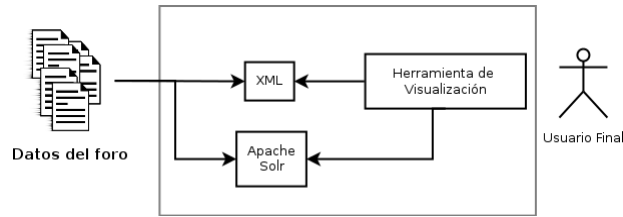


Figura 4.2: Arquitectura de la herramienta de visualización.

El proceso de lectura de datos se realizó a través del método *getCategories()* de la clase *CSVDataManager*, mientras que el de filtrado de datos a través del método *createXml()*. Por último la construcción del archivo para indexar los documentos en apache solr se realizó a través del método *createXmlDocument()*.

4.2. Herramienta de visualización

El desarrollo de la herramienta de visualización se realizó en el lenguaje de programación Java, usando las librerías de “processing” para la elaboración de los gráficos y de librerías que mejoran la interacción y el desempeño de escenas 3d como lo son *GLGraphics*[9] encargada de mejorar el rendimiento de los objetos 3d y *Proscene*[8] que permite el control de la cámara y la interacción con la escena 3d. La arquitectura desarrollada puede observarse en la figura 4.2.

El diseño se puede ver en el diagrama de clases de la figura 4.3. En él se mantienen las clases *ForumCategory*, *ForumThread* y *ForumMessage* usadas en la aplicación anterior. La clase *SpiralVisualization* se encarga de inicializar y dibujar los gráficos a través del método *setup()* y *draw()* respectivamente. La clase *SpiralPainter* calcula las coordenadas para cada uno de los mensajes así como su color a través de los métodos *getCoords()* y *getColors()* e inicializar los modelos 3d usando el método *initDrawCategories()*. La clase *XmlParser* se encarga de obtener los datos desde el archivo xml a través del método *parseXmlFile()* y también de realizar las búsquedas en apache solr a través del método *search()*.

La herramienta muestra las dos metáforas diseñadas (una espiral o varias espirales) vistas en la sección anterior. El cálculo de la posición y del color para cada hilo se muestran a continuación.

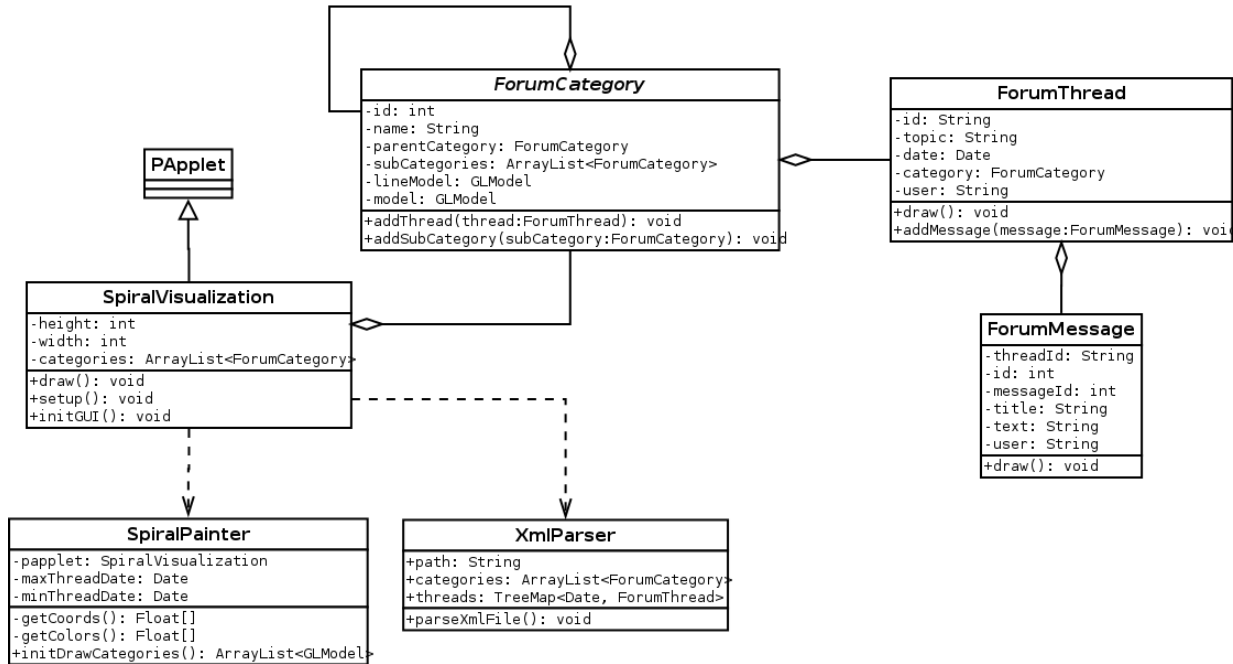


Figura 4.3: Diagrama de clases del prototipo.

4.2.1. Cálculo del color

El color de cada uno de los hilos fue calculado usando el modelo de color HSV [2] (También conocido como HSB). La propiedad *Hue* indica la categoría principal a la que pertenece, mientras que *Sat* (saturation) la sub-categoría en la que se encuentra publicado. La propiedad *B* (brightness) es usada para definir la visibilidad del hilo con un valor constante.

La ecuación 4.1 y 4.2 muestra el cálculo de los colores; donde n_i indica la categoría principal del hilo, N el número de categorías principales, s_i la sub-categoría donde se encuentra el hilo y S el número de sub-categorías que pertenecen a n_i .

$$Hue_i = \frac{n_i}{N} \quad (4.1)$$

$$Sat_i = \frac{s_i}{S} \quad (4.2)$$

4.2.2. Cálculo de la posición por una sola espiral (diseño 1)

El cálculo de la posición se realizó aplicando la función descrita en la ecuación 4.3. Donde a y b son constantes con valores entre 0 y 1 usadas para dar una apariencia elíptica a la espiral y de esta forma ajustarla mejor a la pantalla, r corresponde a la distancia desde el centro de la espiral hasta el i -ésimo hilo, y t corresponde al ángulo en que debe ubicarse el hilo calculado a partir del tiempo del último mensaje publicado como se indica en la ecuación 4.4.

Para calcular t primero se restó el tiempo en que se publicó el último mensaje del i -ésimo hilo (representado en la variable T_i) menos el mismo tiempo calculado para el hilo más antiguo (Variable T_{min}). Este tiempo luego es multiplicado por la constante R que indica el número de radianes asignados para representar cada hora.

Las coordenadas x , y y z se calculan como indica la ecuación 4.5. Para dibujar los mensajes solo es necesario calcular la coordenada z , y se mantienen las coordenadas x y y usadas para dibujar el hilo.

$$r_i = \frac{(ab)}{\sqrt{(bcost_i)^2 + (asint_i)^2}} \quad (4.3)$$

$$t_i = (T_i - T_{min}) R \quad (4.4)$$

$$x_i = rcost_i, y = rsint_i, z = t_i \quad (4.5)$$

En caso que existan dos hilos en la misma posición (fueron publicados al mismo tiempo), se modifican los valores de x y y para que los hilos no se superpongan. Esto se logra sumando un valor que dependerá del número de hilos que comparten dicha posición. Esta visualización se puede ver en la figura 4.4.

4.2.3. Calculo de la posición para varias espirales (diseño 2)

Para lograr separar las categorías se modificó la formula en que se calcula la variable t (el ángulo que le corresponde al hilo). En este caso dicha variable dependerá del número

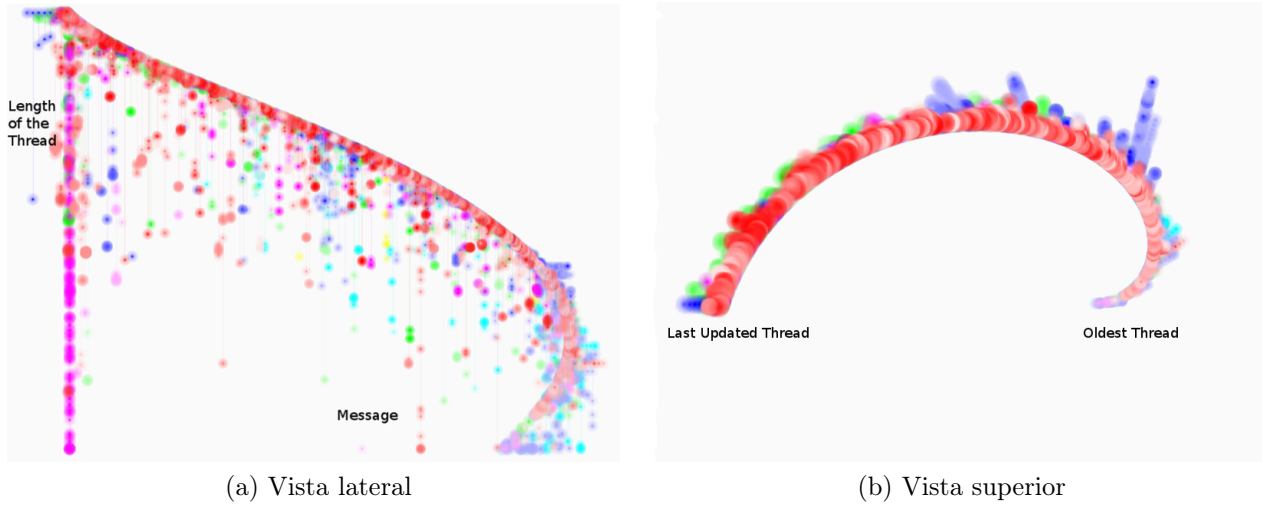


Figura 4.4: Implementación del primer diseño. Los datos corresponden al foro “*processing forum alpha*”. Se representa el foro a través de una sola espiral.

de categorías y de la categoría principal a la que pertenece el hilo. El calculo se puede ver en las ecuaciones 4.6 y 4.7, donde N corresponde al número de categorías principales, d es una constante que genera la curvatura de la espiral y n_i indica la categoría principal del hilo. La visualización se puede ver en la figura 4.5.

$$p = \frac{2\pi}{N} - d \quad (4.6)$$

$$t_i = (N (T_i - T_{min}) + n_i) p \quad (4.7)$$

4.2.4. Interacción del usuario

El usuario puede interactuar con la herramienta a través del mouse y del teclado. A continuación describiremos los métodos de interacción desarrollados. Se pueden ver todas estas características en la figura 4.6.

Acercar o alejar la escena Para realizar esta tarea se usa el botón derecho del mouse.

Desplazar la escena Verticalmente y horizontalmente Esta tarea se puede realizar usando el botón izquierdo del mouse. Para cambiar el eje de rotación, se usa la tecla “x”

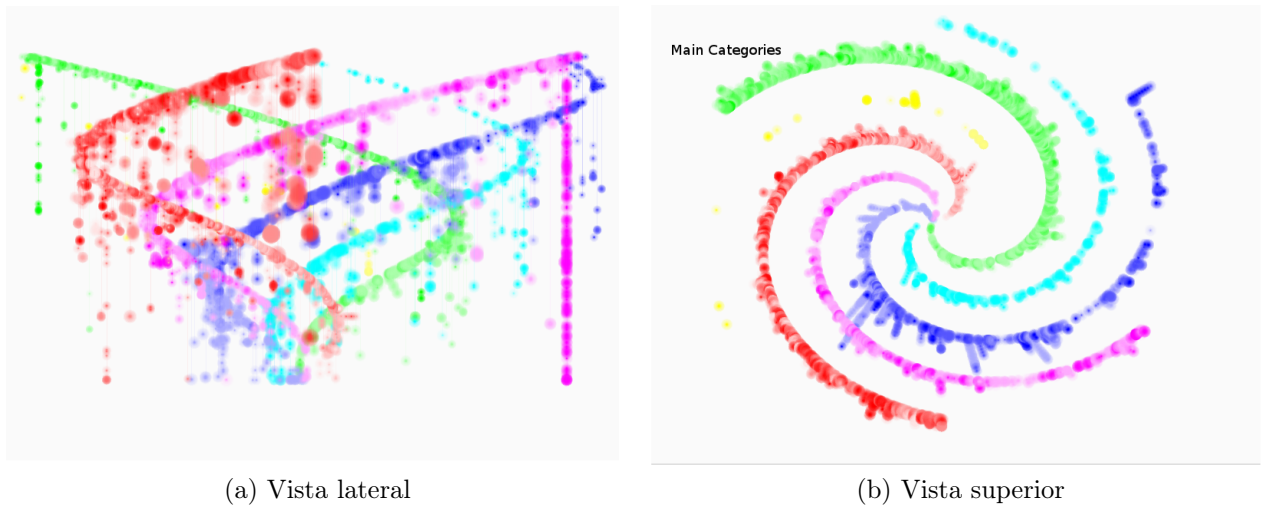


Figura 4.5: Implementación del segundo diseño. Los datos corresponden al foro “*processing forum alpha*”. El foro se representa como una galaxia en espiral, donde los brazos representan las categorías principales.

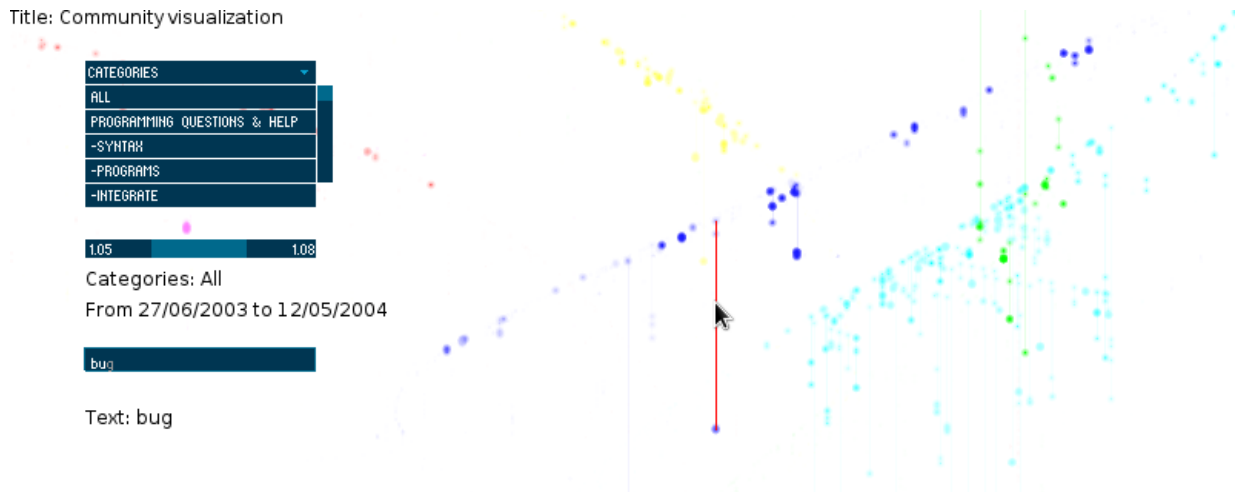


Figura 4.6: Ejemplo de interacción con el usuario. En este caso se buscaron los hilos que contienen la palabra *bug* entre el 27 de junio de 2003 al 12 de mayo del 2004 en todas las categorías del foro.

del teclado.

Selección de la visualización Esta operación la puede realizar presionando la tecla “b” del teclado.

Cambiar el nivel de detalle de la visualización Para ello se usan las teclas “v” y “c” del teclado.

Filtrar los hilos por categoría y sub-categoría En este caso se creó un listado con las categorías y sub-categorías del foro. Para filtrar se debe seleccionar la categoría deseada en la lista.

Filtrar los hilos por fecha de la última publicación Se creó un control que permite al usuario seleccionar la fecha mínima y máxima.

Filtrar los hilos por búsqueda de texto Se agregó un campo de texto donde el usuario puede ingresar sus búsquedas y luego presionar la tecla “Enter” para realizar la consulta. Como se indicó en la sección anterior, la búsqueda se realiza usando la herramienta apache Solr, a través de la url <http://direccionIP:8983/solr/select/?q=texto> pasando en el parámetro *q* la cadena a buscar. Los documentos que cumplen con los parámetros de búsqueda son retornados a través del formato xml, y contienen el id del hilo donde se encuentra el texto consultado.

Selección del hilo Para ello debe ubicar el cursor del mouse sobre el hilo; Esta operación mostrará el título del hilo. Para ver los mensajes el usuario debe realizar click dos veces sobre el hilo seleccionado. Esto mostrará la pantalla con los mensajes del hilo (ver figura 4.7).

4.3. Resultados

En esta sección con base en las necesidades planteadas para cada tipo de usuario, hacemos una revisión de cuales son resueltas por nuestra herramienta.



Figura 4.7: Implementación navegación dentro del hilo.

Usuarios administradores

- *Validar qué hilos se pueden cerrar y cuales deben permanecer abiertos:* La visualización permite identificar de forma rápida los hilos más antiguos en el foro. Sin embargo, debido al alcance del proyecto el usuario administrador no puede cambiar el estado del hilo. Se plantea esta funcionalidad en el trabajo futuro cuando el sistema cuente con la conexión al sistema administrador de foro.
- *Verificar que los hilos se encuentren en la categoría correcta:* El buscador textual puede ayudar a identificar estos hilos, sin embargo, para un mejor resultado se requiere de un análisis más avanzado en cuanto al contenido del hilo (análisis semántico de la información contenida en él) que por cuestiones de tiempo esta fuera del alcance del proyecto.
- *Identificar los hilos con mayor participación por parte de los usuarios del foro:* En la visualización se identifican fácilmente los hilos con una mayor participación por parte de los usuarios.

- *Reconocer las categorías que tienen una mayor o menor participación por parte de los usuarios:* Esto se logra fácilmente en el segundo diseño de la metáfora.

Usuarios Frecuentes

- *Encontrar hilos nuevos del tema de su interés:* Vemos que la herramienta resuelve esta necesidad a través de los filtros por categorías, fechas y el buscador textual.
- *Identificar los hilos con mayor participación por parte de los usuarios del foro:* Ver usuario administrador.
- *Reconocer a los usuarios que comparten sus mismos intereses:* Esta necesidad no está cubierta en su totalidad por la herramienta. Por ahora el usuario puede buscar los hilos en los que un usuario participa usando el buscador textual. Se plantea como trabajo futuro identificar los hilos en los que un usuario ha participado a través de la ventana de mensajes.

Usuarios temporales

- *Encontrar rápidamente los hilos que contienen la información de interés:* Este objetivo se logra fácilmente a través del buscador de texto.
- *Poder ver el número de participaciones y la antigüedad en el hilo de su interés:* También se encuentra implementado en la metáfora.

5 Conclusiones y trabajo futuro

La herramienta de visualización implementada pretende guiar el proceso de búsqueda y contribuir a la solución de las necesidades propias de los usuarios del foro. Hasta el momento, se diseñó e implementó con éxito una metáfora de visualización totalmente interactiva que permite destacar varias características de las conversaciones, y con ella se han logrado resolver varias de las necesidades de los usuarios. Sin embargo, todavía hay características que nos gustaría agregar al proyecto, las cuales describiremos a continuación:

- Como se indicó en la introducción deseamos que nuestra herramienta pueda conectarse de forma más directa con varios sistemas administradores de foro, a través de un plugin.
- Queremos migrar esta herramienta a javascript, ya que es soportado por la mayoría de navegadores, lo que permitiría los usuarios interactuar de forma más directa con la herramienta. Además, los navegadores de última generación permiten la interacción con modelos 3d (por medio del WebGL) por lo que se podría migrar nuestro proyecto sin problemas.
- Otra funcionalidad que deseamos implementar es la posibilidad de hacer búsquedas por usuarios (destacar en la metáfora los hilos en donde un usuario ha participado). En este momento es posible realizar esta tarea haciendo uso de la búsqueda por texto (a través del nombre de usuario). Sin embargo, esta búsqueda debería ser gráfica, seleccionando el usuario directamente en la visualización de los mensajes.
- Agregar un filtro por número de mensajes. Aunque la interfaz muestra de forma eficiente los mensajes de los hilos, en la práctica se hace necesario agregar un filtro por número de mensajes para facilitar la navegación. Muchas veces el usuario puede no estar interesado en los hilos que tengan pocos mensajes publicados por lo que este filtro podría facilitar este tipo de búsquedas.

- Funcionalidad de publicar nuevos mensajes e hilos. En este momento el proyecto se enfocó a la visualización de la información existente en el foro, sin embargo una vez se cuente con el plugin que permita interactuar con el sistema administrador de foro, se pretende permitir el ingreso de nuevos mensajes de texto por parte de los usuarios.
- Editar la información del hilo y del mensaje por parte de los administradores del foro.

Bibliografía

- [1] Apache solr project. Internet: <http://lucene.apache.org/solr/>, 2012. (2012 Feb.) <http://lucene.apache.org/solr/>.
- [2] Modelo hsv. Internet: http://en.wikipedia.org/wiki/HSL_and_HSV, 2012. Revisado en Febrero de 2012 en http://en.wikipedia.org/wiki/HSL_and_HSV.
- [3] Programa iconv. Internet: <http://en.wikipedia.org/wiki/Iconv>, 2012. Revisado en Febrero de 2012 en <http://en.wikipedia.org/wiki/Iconv>.
- [4] Spiral galaxy def. Available: http://en.wikipedia.org/wiki/Spiral_galaxy, 2012. (2012 Feb.) http://en.wikipedia.org/wiki/Spiral_galaxy.
- [5] Web scraping. Available: http://en.wikipedia.org/wiki/Web_scraping, 2012. (2012 Feb.) http://en.wikipedia.org/wiki/Web_scraping.
- [6] Yabb project. Available: <http://www.yabbforum.com>, 2012. (2012 Feb.) <http://www.yabbforum.com>.
- [7] D. Boyd, H.-Y. Lee, D. Ramage, and J. Donath. Developing legible visualizations for online social spaces. *Hawaii International Conference on System Sciences*, 4:115, 2002.
- [8] Jean Pierre Charalambos. Proscene project. Internet: <http://code.google.com/p/proscene/>, 2012. Revisado en Febrero de 2012 en <http://code.google.com/p/proscene/>.
- [9] Andrés Colubri. Glgraphics project. Internet: <http://glgraphics.sourceforge.net/>, 2012. (2012 Feb.) <http://glgraphics.sourceforge.net/>.
- [10] Judith Donath. A semantic approach to visualizing online conversations. *Commun. ACM*, 45:45–49, 2002.

- [11] Björn Engdahl, Malin Köksal, and Gary Marsden. Using treemaps to visualize threaded discussion forums on pdas. In *CHI '05 extended abstracts on Human factors in computing systems*, CHI EA '05, pages 1355–1358, New York, NY, USA, 2005. ACM.
- [12] Ben Fry. *Visualizing Data: Exploring and Explaining Data with the Processing Environment*. O'Reilly Media, 1 edition, January 2008.
- [13] Ben Fry and Casey Reas. Processing forum alpha. Available: [//www.processing.org/discourse/yabb/YaBB.cgi](http://www.processing.org/discourse/yabb/YaBB.cgi), 2012. (2012 Feb.) [//www.processing.org/discourse/yabb/YaBB.cgi](http://www.processing.org/discourse/yabb/YaBB.cgi).
- [14] Ben Fry and Casey Reas. Processing project. Available: <http://www.processing.org>, 2012. (2012 Feb.) <http://www.processing.org>.
- [15] Beomjin Kim. Astrograph for exploring discussion forums. In *IKE'03: Proceedings of the International Conference on Information and Knowledge Engineering.*, pages 321–330, 2003.
- [16] Beomjin Kim and Philip Johnson. Graphical interface for visual exploration of online discussion forums. *Journal of Systemics, Cybernetics and Informatics*, 4(4), January 2006.
- [17] R. Mohamed, J. D. Ferguson, D. Elswailer, A. MacCormick, J. Wilson, and G. R. S. Weir. Bulb-visualizing bulletin board activity. *Proceedings of the 2003 International Conference on Information and Knowledge Engineering*, pages 321–327, 2004.
- [18] R. Mohamed, J. D. Ferguson, A. MacCormick, and D. Elswailer. Developing bulletin board visualizations. *Proceedings of the IADIS International Conference on Web Based Communities*, pages 11–18, 2004.
- [19] Ethan Perry and Judith Donath. Anthropomorphic visualization: a new approach for depicting participants in online spaces. pages 1115–1118, Vienna, Austria, 2004. ACM.
- [20] C. Reas and B. Fry. *Getting Started with Processing*. Make Series. O'Reilly Media, 2010.

-
- [21] Rebecca Xiong and Judith Donath. Peoplegarden: creating data portraits for users. In *UIST '99: Proceedings of the 12th annual ACM symposium on User interface software and technology*, pages 37–44, New York, NY, USA, 1999. ACM.