

SOFTWARE VISUALIZACIÓN DE INFORMACIÓN

Visualización de Datos, Semántica y Medios Online de Prensa Chilenos

////////////////////////////////////

ACADÉMICO RESPONSABLE:

Diego Gómez

VINCULACIÓN:

Departamento de Diseño, Facultad de Arquitectura y Urbanismo, Universidad de Chile

CONTACTO:

e: diegogomez@uchilefau.cl

t: 2978-3043

PLAZOS DESARROLLO:

Julio y Agosto de 2013

FINANCIAMIENTO:

Vicerrectoría Investigación y Desarrollo (VID) U. de Chile, Creación U-Apoya Línea 4

////////////////////////////////////

1. Descripción General

Este proyecto de creación apoyado por la VID de la Universidad de Chile, busca visualizar contenido noticioso publicado por diversos *medios de prensa online chilenos*. Se busca que las visualizaciones logradas, *den pistas* de los distintos *modos* en que un mismo tema es o puede ser tratado por los diversos medios.

Técnicamente lo que se busca es trabajar visualizando texto. A partir de feeds RSS de los medios de prensa online seleccionados, se harán análisis cuantitativos y otros de carácter cualitativo respecto al texto de las noticias y artículos disponibles.

Por ejemplo, en lo cuantitativo, se querrá saber cuantas noticias o artículos publica el

medio A dentro de las categorías **política, economía, espectáculos, deportes, internacional, cultura, otros** . Es decir, cuantificar.

En lo referido a lo cualitativo, lo que se busca es poder determinar el carácter de una noticia o artículo. Esto es, poder determinar si el texto tiene un carácter **beligerante, complaciente, ambiguo u optimista**. Utilizando *regular expressions* (regEx), se compararán las palabras de la noticia o artículo en cuestión con palabras que previamente se han categorizado (en un bases de datos probablemente) como propias de cada uno de los cuatro caracteres establecidos. Luego, y por ejemplo, mediante formulas de cuantificación se podrá establecer si la **noticia 1** tiene **X o +** cantidad de palabras propias del perfil beligerante, y por lo tanto, es más o menos beligerante.

2. La Visualización

Hay que señalar primeramente que este software es parte de un proyecto de creación, por lo que la visualización no se ajustará necesariamente a estándares convencionales de visualización, ni pretende privilegiar la objetividad de la información y su comunicación de modo eficiente. Antes de eso, se busca hacer foco en la experimentación de nuevas formas (lenguajes visuales) para el diseño de visualizaciones, la valorización de la expresión y la subjetividad.

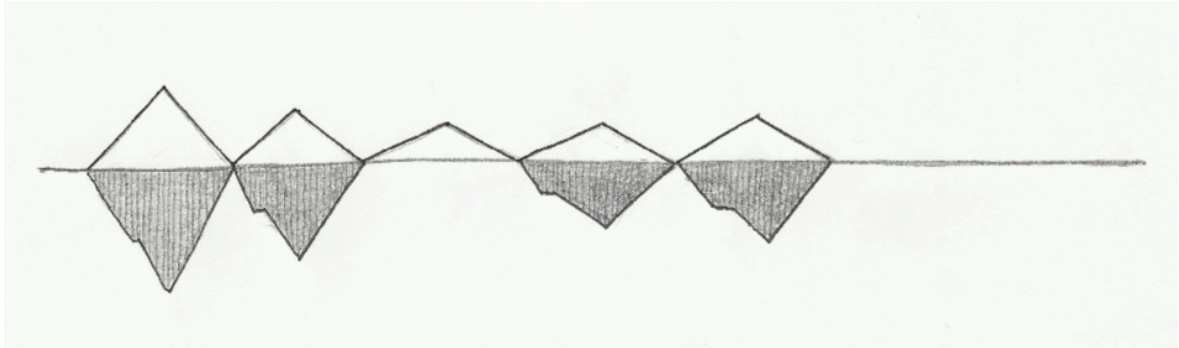
La visualización usará la analogía del paisaje, de modo abstracto, para mostrar visualizar la información a la que hace referencia. Cuando digo abstracto, me refiero a que no será literalmente paisaje identificable totalmente como tal. Por ejemplo, hay formas, que en cierto modo, hacen referencia o parecen montañas, pero no son eso exactamente.

Dicho lo anterior, el software tiene dos secciones de visualización propiamente tales. Habrá otras secciones pero netamente informativas (explicaciones, créditos, colofón, etc.)

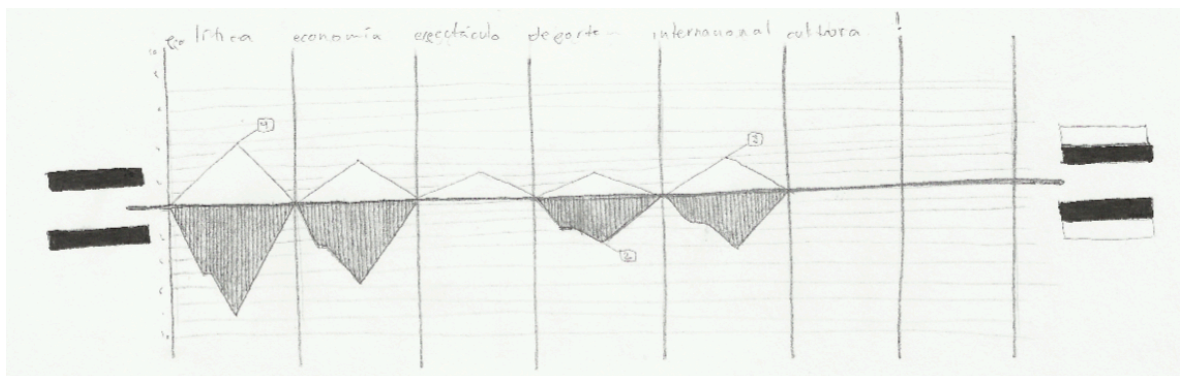
a. Primera Sección

Es lo primero que se ve al hacer correr el software. Hay un eje horizontal en el centro de la pantalla. Sobre él, se ubican 7 formas triangulares (que llamaré

montañas) y bajo él otras 7 formas triangulares (que llamaré icebergs). La sección superior (montañas) es de carácter cuantitativo y cada montaña da cuenta de la cantidad de noticias o artículos en una de las 7 categorías determinadas (política, economía, espectáculos, etc). Por otro lado, la parte inferior (icebergs) es cualitativa y señala que tan *beligerante*, *complaciente*, *ambiguo* u *optimista*, es el grupo de noticias o artículos (justo arriba de ella) de una categoría .



En un primer momento el software presenta sólo las formas señaladas, sin ninguna explicación o leyenda de texto. Un pequeño botón permite desplegar un menú general. Con eso, es posible ver e interactuar con el botón que me permite selección el medio de prensa de interés (si inicia con uno por defecto) en la parte superior, y el perfil de análisis cualitativo en la parte inferior (también se inicia con uno por defecto).



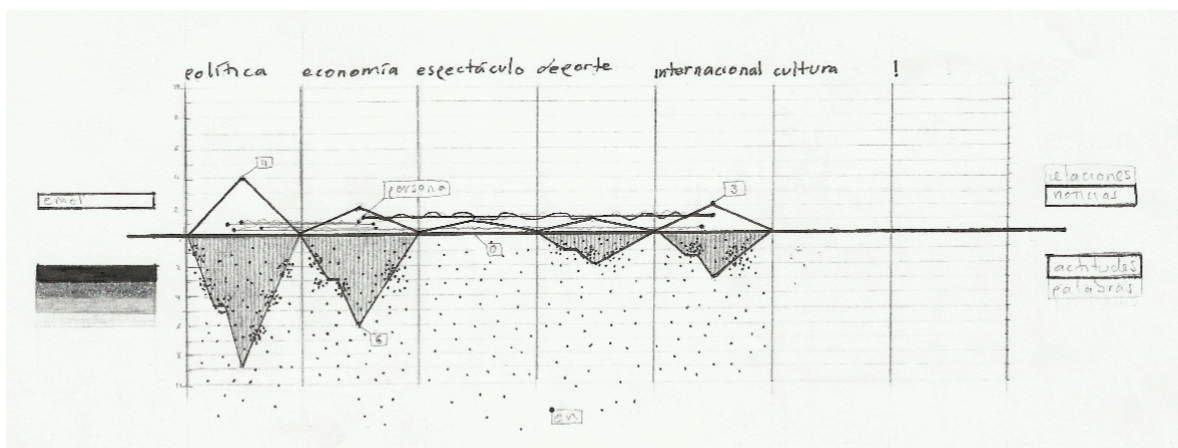
Además, otro botón me permite dejar a la vista la grilla que señala las columnas para cada una de las categorías de contenidos antes nombradas y facilita apreciar

las relaciones de cambio de altura de cada una de las formas (montañas y icebergs). Cabe decir que las “cumbres” de las montañas se modifican en relación a cuantos artículos hay en la categoría y, el vértice inferior de los icebergs también varía en tanto la intensidad del carácter en cuestión.

Por último, una ultima capa de relaciones puede ser desplegada a través del uso de botones ubicados a la derecha de la visualización.

El de la parte superior permite ver relaciones de personajes entre las categorías. Es decir, por ejemplo, si el nombre *Bachelet* aparece en la categoría *política* y *economía*, aparecerá un punto en cada sección, ambos unidos por una línea. Además, este línea, tendrá sobre ella otra línea, la contendrá más o menos curvas dependiendo de cuantas veces aparezca ese nombre en cada categoría. Así por ejemplo, si el nombre *Bachelet* aparece sólo 2 veces en la categoría *política*, y 6 en la categoría *economía*, el lado de la línea cercano al primer punto tendrá menos curvas que el sector de la línea cercano al segundo punto.

En la parte inferior el botón ver pequeños círculos que corresponden a las palabras presentes en los distintos artículos. Sobre los icebergs, se irán agrupando a través de “clusters” los círculos (palabras) que permiten señalar que una sección es más o menos beligerante, complaciente, etc.



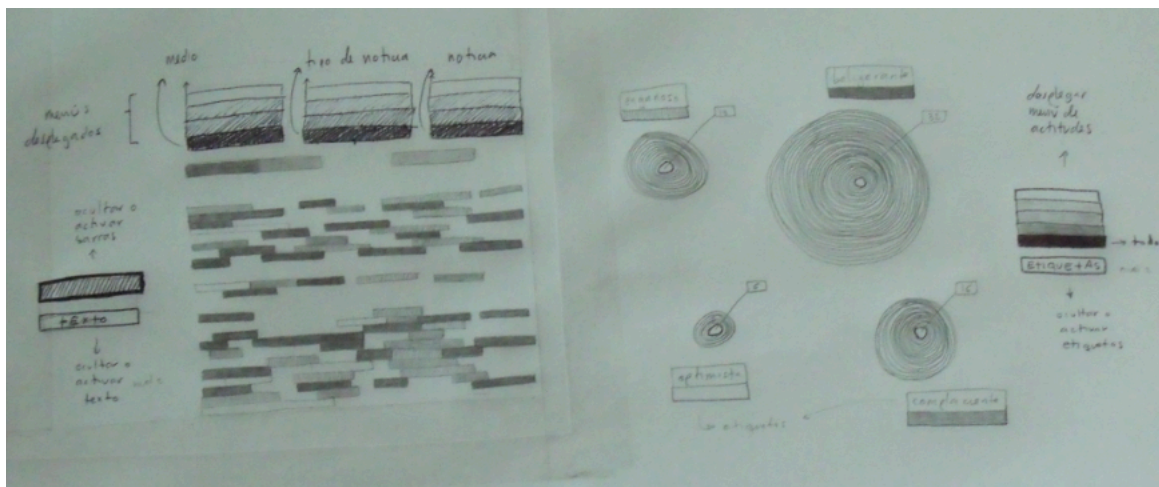
Al posar el cursor del mouse sobre las “cumbres” de las montañas aparecerá el número de noticias/artículos en cada categoría. Del mismo modo, al posar el

cursor sobre puntos dentro de las “montañas” se verá el nombre del personaje en cuestión. Y, al posar el cursor sobre los círculos en la parte inferior se verá la palabra a la cual hace referencia. Por otro lado, al hacer clic en alguna de las montañas, el software llevará a la segunda sección donde se puede profundizar en los artículos de la categoría en cuestión. Por último, es deseable que exista un botón que permita imprimir y descargar como PDF la visualización desplegada (sin leyendas).

b. Segunda Sección

Aquí el usuario podrá profundizar sobre los artículos de una categoría (ej. política). Lo que se observa en primer lugar es la pantalla dividida en dos.

En el costado izquierdo el texto de una noticia/artículo desplegado y navegable con scroll si corresponde. Pero además, en una segunda capa, sobre el texto, se pueden ver bloques sobre palabras (como si estuvieran censuradas) que señalan (por su tono) si la palabra en cuestión cabe dentro de las palabras de perfil beligerante, complaciente, etc. Estas capas se pueden, mediante botones, encender o apagar indistintamente; así, puedo tener desplegado sólo el texto, sólo los bloques, o ambos a la vez. También por botones, puedo seleccionar un noticia/artículo en base a los disponibles para la categoría en cuestión. Se inicia por defecto con el más reciente.



En el costado derecho, una visualización más abstracta, muestra a través de

grupos de anillos concéntricos, la cantidad de palabras que componen cada uno de los perfiles bajos los cuales se puede “enjuiciar” la noticia/artículo desplegado (beligerante, complaciente, ambiguo, optimista). Al posar el cursor del mouse sobre los anillas es posible también ver la palabra en cuestión.

3. Empaquetamiento

Debido a políticas de planificación académicas y de asociación internacional del Departamento de Diseño, es altamente deseable que el software se desarrolle con Processing.

El despliegue y distribución del software debe ser vía web, directamente en browser vía HTML5+JS, precisamente utilizando el modo JavaScript presente en Processing. Sin embargo dada la utilización de Librerías (por ejemplo para GUI) sabemos que pueden presentarse dificultades para alcanzar este objetivo, debido a potenciales incompatibilidades de las librerías con el modo JavaScript. Si situación es insalvable, la aplicación deberá distribuirse vía descarga como “stand-alone” multiplataforma (Mac, Linux, Windows). También se considera la distribución abierta del código Processing. Se pueden establecer un licenciamiento Creative Commons para esto último.

4. Referentes

Nuestro referente más directo, en lo técnico y de algún modo para la segunda sección de visualización, es el proyecto “ReConstitution2012” del estudio estadounidense SOSO Limited. Este proyecto, analizó en vivo los debates presidenciales de los candidatos a presidente de ese país:

<http://www.recon12.com/index.html>

Recientemente el mismo estudio ha liberado el código fuente para el análisis de texto para sus proyectos. Nosotros no lo hemos estudiado, pero quizá les sea de utilidad, aunque tengo entendido ellos no han trabajado con Processing sino con OpenFrameworks:

<http://sabotage.sosolimited.com>

Por ultimo, un referente indirecto es el proyecto *They Rule*, realizado por el diseñador y artista Josh On entre el 2001 y 2004:

<http://www.theyrule.net>

Este proyecto es interesante para nosotros porque constituye una visualización que en lo técnico se presenta directamente en browser, y el lo visual busca analogías interesantes para mostrar la temática que cubre, alejándose de los formatos estándar y clásicos para las visualizaciones de información. Al mismo tiempo, la interfaz es sencilla y queda en un segundo plano, y los códigos cromáticos (color) de la visualización son muy acotados (grises), dándole mayor relevancia a las formas y la conexiones entre éstas.