WorldLens: Exploring World Events Through Media

by

Jonathan Eliezer Speiser

B.S., University of Maryland (2011)

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
in partial fulfillment of the requirements for the degree of

Master of Science in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2014

Author ... Signature redacted ....
Program in Media Arts and Sciences,
School of Architecture and Planning,
May 9, 2014

Certified by . Signature redacted ....
Andrew B. Lippman
Associate Director & Senior Research Scientist, MIT Media Lab
Thesis Supervisor

Accepted by ..... Signature redacted ....
Professor Patricia Maes
Associate Academic Head, Program in Media Arts and Sciences

WorldLens: Exploring World Events Through Media

by

Jonathan Eliezer Speiser

Submitted to the Program in Media Arts and Sciences,
School of Architecture and Planning,
on May 9, 2014, in partial fulfillment of the
requirements for the degree of
Master of Science in Media Arts and Sciences

Abstract

The goal of WorldLens is to provide a visual answer to the question: "what is happening in the world?" This thesis entails the design and development of a system that provides an exploratory view into world events across varied media types. The fundamental ideas are to take a data-driven approach by crawling a broad swath of content sources, and map the resulting data into an interactive visualization. WorldLens aims to obviate the need for an a priori search term, and instead focuses on facilitating informative discovery across news content from articles, web video, and broadcast television.

Thesis Supervisor: Andrew B. Lippman
Title: Associate Director & Senior Research Scientist, MIT Media Lab

WorldLens: Exploring World Events Through Media

by

Jonathan Eliezer Speiser

The following people served as readers:

Signature redacted

Thesis Reader ....

Professor Deb Roy
Associate Professor, Media Arts and Sciences

Signature redacted

Thesis Reader ....

Professor César Hidalgo
Assistant Professor, Media Arts and Sciences

Acknowledgments

This work was supported by the Media Lab Digital Life Program and the Ultimate Media Initiative, which is in turn supported by Comcast, DirecTV, and Cisco.

The past two years have been an incredible experience, and I would like to take this opportunity to recognize the people who have helped me along the way.

First, I have to thank Andy Lippman for giving me the opportunity to come to Lab and for his unique perspectives which have undoubtedly influenced the way I think.

I would like to thank Deb Roy and César Hidalgo who served as my readers, for their support, creativity, and thoughtful feedback.

I would like to thank my close collaborators in E14-348C: Rob Hemsley and Dan Sawada.

I would like to thank my fellow students in the Viral Spaces group: Eric Dahlseng, Vivian Diep, Savannah Niles, Amir Lazarovich, Travis Rich, Jeremy Rubin, Shen Shen, and Eyal Toledano.

I would also like to thank my family and extended family: Eema, Abba, Bobbie, Eliyahu, Eva, Netanel, Yoav, Gavriel, Shoshana, Rebekah, Avi, Irit, David, Leora, and Ilan.

Last, I would like to thank my wonderful wife Tali. Thank you for encouraging me to apply to graduate school and for your ceaseless support these last two years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

What is happening in the world right now? This is a question that is easy to pose, but hard to answer. For the last century we have relied on professional curators to help us answer this question. Newspapers and television news programs helped twentieth century citizens identify issues to focus on.

Today's predominantly search-based world event interfaces make it very easy for us to find what we want, but not necessarily easy to find information beyond the borders of our familiar environments. Furthermore, the fact that search-based interfaces are heavily reliant on text to convey information makes it difficult to get an understanding of world events at a glance. Consequently, exploration is typically difficult and time consuming.

In addition to the need for a visual interface that affords exploration, there is a need to consider more diverse sources of information. In the past, established broadcasters such as CNN, Fox News and NBC News, were the main sources for news video production and dissemination. The rise of platforms such as YouTube, represent a challenge to the traditional model. A 2012 Pew study comparing trends of YouTube with broadcast news, remarked: "For the news industry, the growth of YouTube and other video sharing sites represent a significant opportunity and also a challenge" [5]. While original news production is available on YouTube, it remains largely outside the purview of current news aggregators, which typically focus on sourcing text articles from major news establishments.

To construct a portrait of world events at present, therefore we ought to look at the nexus of textual web content, web video, and broadcast television. In order to "look" at these everyday media sources in an automated fashion, two main components are required. The first component is an ingestion layer. This layer looks at sources of interest, such as news websites, internet video repositories or even cable television channels and extracts data. The second layer is an analysis layer. This layer is responsible for taking the raw data acquired from the first layer, and extracting or producing useful information from the source data, in order to impose structure between relevant pieces of media. These two layers are realized in this thesis through the development of Constellation, an overarching media aggregation and analysis framework.

1.1 Context of Thesis Work

The work described in this thesis grew out of the Ultimate Media program at the MIT Media Lab. Ultimate Media is a joint program between the Viral Spaces and Object Based Media groups, which was started in reaction to the recognition that our visual media experiences had remained both markedly underdeveloped and isolated. Traditional television and media systems were continuously built with the interests and limitations of content creators and manufacturers alike in mind. To address these shortcomings, the Ultimate Media program investigates frontiers of media experiences that include data-driven media presentation and generation, social integration, and new paradigms for remote control. The focus of the program is to concentrate on creating the best experiences for the people who use, interact and express themselves through media, without the confining shackles of traditional media business considerations.

Many of today's media interfaces from cable boxes to broadcast television, to Internet video repositories, are built by corporations with particular legal restrictions, content contracts, revenue generation strategies and agendas. These factors culminate in an experience that does not place end users at the focal point of media, but rather at the periphery, forcing end users to ultimately adapt their behaviors to the facilities and goals of a particular media conglomerate. The Ultimate Media program seeks to reverse the bit on traditional media interfaces and to enshrine users at the center of the media universe. WorldLens represents this ethos as it scours information and media across typically disparate sources of media and brings the resultant broadcast and web content together in a unified way, unlike many of the services of today.

1.2 Thesis Vision

The goal of this thesis is threefold. First, to develop a system capable of visually answering the question of what is happening in the world. Second, to illustrate the potential of interfaces that do not rely on a priori search terms to funnel users towards particular pieces of content, but instead enable open and informative exploration. The latter allows users to discover content they would not have searched for or otherwise encountered on their own. Third, to illustrate the value of adjoining related media from typically isolated content sources, such as broadcast television and web articles.

This thesis envisions the WorldLens as a fixture in the home. Instead of an empty black frame sitting in the living room in the morning, I envisage WorldLens serving as a lively window into world events. As the resolution of consumer televisions continue to increase, and touch-based interfaces continue to see increased adoption, the living room screens of the future are poised to play a much larger role in peoples daily lives, and it is in that context which I see WorldLens thriving.

1.3 Thesis Map

Chapter 2 of this thesis examines the background and history behind news aggregation, including both the underlying technical methods used to aggregate and cluster news. The latter portion of chapter 2 examines the development and use of 3D user interfaces in academia and mainstream applications, as well as existing news aggregation and visualization services.

Chapter 3 focuses on the architecture and implementation of Constellation, the media analytics framework. Chapter 4 focuses on design iterations for the WorldLens visualization. Chapter 5 presents the final version of the WorldLens interface and its associated interaction experience as well as an evaluation of the system. Chapter 6 examines future work and additional applications of WorldLens and its associated infrastructure. Rounding out the thesis, Chapter 7 offers conclusions and final thoughts.

Chapter 2

Background

2.1 Aggregators and Content Analysis

In 1945 Dr. Vannevar Bush imagined Memex, a machine that could store various types of content in microfilm form. Bush envisioned that the Memex would include "books of all sorts, pictures, current periodicals, newspapers" [3]. Consequently, Memex can be seen as an early concept for a content aggregator. Perhaps more importantly, however, Bush recognized that the importance was not merely creating a content repository, but more crucially, he recognized the value in understanding underlying structural connections inherent in the repository. As Bush wrote: "This is the essential feature of the memex. The process of tying two items together is the important thing."

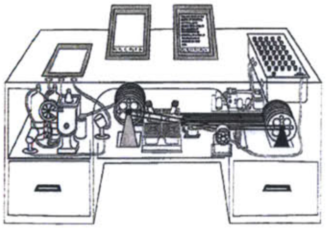


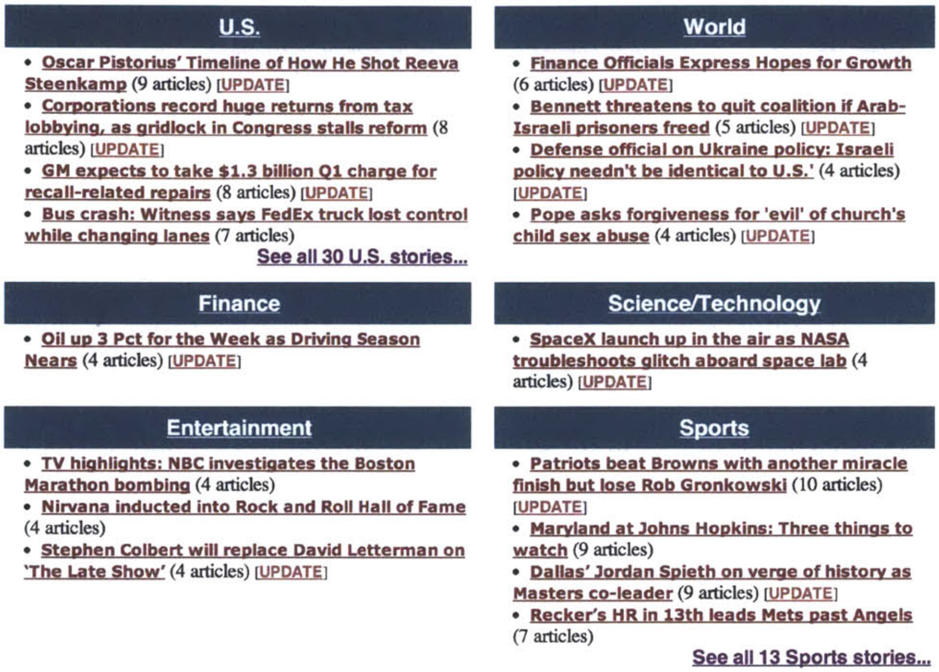
Figure 2-1: Memex illustration.

Nearly thirty years later, in 1972 Dr. Karen Sparck Jones had an idea that could facilitate Bush's "essential feature." Historically, in the area of document analysis, term frequency (tf) was regarded as a salient factor in understanding the importance of particular words to a document. In 1958, thirteen years after Bush published his vision for the Memex, Hans Luhn a computer scientist working for IBM had put forward what now seems like a trivial idea: "It is here proposed that the frequency of word occurence in an article furnishes a useful mesaurement of word significance [20]." Luhn had been working on automatically creating abstracts, and as part of that process he promoted the use of term frequency as a key "ingredient" in understanding the relevance of particular frequent words to a given article [22].

Going beyond Luhn's work, Jones key insight was to highlight the value of examining inverse document frequency. Whereas Luhn was primarily "thinking about word frequency within individual documents", Jones realized that not all words in a search query should be considered equal, and in fact terms could be weighted based upon a corpus of documents to provide better results [17].

As a very simplified illustration, consider the search query, "tallest building in Brooklyn." A naïve information retrieval system could have had any documents containing the terms "tallest", "building", and "in" on par with documents containing the term "Brooklyn". With this naïve approach it is possible, and indeed likely, that there could be many highly ranked search results that have nothing to do with "Brooklyn" at all, and thus are most likely useless to the querying user. Jones however proposed the use of inverse document frequency [15], which could weigh query terms in inverse proportion to their frequency throughout the document corpus. Therefore, frequent terms such as "in", "is", etcetera, would no longer be on par with less frequent terms. As a direct result of Jones work, Dr. Gerard Salton another pioneering figure in information retrieval, published the concept for using term frequency inverse document frequency weighting [31]. Combining both the aspects of the term frequency (number of times a word appears in a document) and the inverse document frequency (a measure of the words frequency throughout a corpus of documents) Salton developed a critical underpinning for the information retrieval systems. AltaVista, one of the first early web engines founded in 1995, used tf-idf [16]. The combined work of Jones and Salton would also provide an important basis for the first automated news aggregators.

In 2002 a team working at Columbia University launched NewsBlaster. NewsBlaster was an early Internet news aggregator that analyzed content from various sources, and could cluster related articles based on the textual content of the articles.



U.S.

- **Oscar Pistorius' Timeline of How He Shot Reeva Steenkamp** (9 articles) [UPDATE]
- **Corporations record huge returns from tax lobbying, as gridlock in Congress stalls reform** (8 articles) [UPDATE]
- **GM expects to take $1.3 billion Q1 charge for recall-related repairs** (8 articles) [UPDATE]
- **Bus crash: Witness says FedEx truck lost control while changing lanes** (7 articles)

See all 30 U.S. stories...

World

- **Finance Officials Express Hopes for Growth** (6 articles) [UPDATE]
- **Bennett threatens to quit coalition if Arab-Israeli prisoners freed** (5 articles) [UPDATE]
- **Defense official on Ukraine policy: Israeli policy needn't be identical to U.S.'** (4 articles) [UPDATE]
- **Pope asks forgiveness for 'evil' of church's child sex abuse** (4 articles) [UPDATE]

Finance

- **Oil up 3 Pct for the Week as Driving Season Nears** (4 articles) [UPDATE]

Science/Technology

- **SpaceX launch up in the air as NASA troubleshoots glitch aboard space lab** (4 articles) [UPDATE]

Entertainment

- **TV highlights: NBC investigates the Boston Marathon bombing** (4 articles)
- **Nirvana inducted into Rock and Roll Hall of Fame** (4 articles)
- **Stephen Colbert will replace David Letterman on 'The Late Show'** (4 articles) [UPDATE]

Sports

- **Patriots beat Browns with another miracle finish but lose Rob Gronkowski** (10 articles) [UPDATE]
- **Maryland at Johns Hopkins: Three things to watch** (9 articles)
- **Dallas' Jordan Spieth on verge of history as Masters co-leader** (9 articles) [UPDATE]
- **Recker's HR in 13th leads Mets past Angels** (7 articles)

See all 13 Sports stories...

Figure 2-2: Screenshot showing article clusters in Newsblaster.

To achieve the clustering, NewsBlaster built off the work of Jones and Salton and leveraged tf-idf [23]. Tf-idf is used in the context of Newsblaster to determine both the relevant broad category, and specific event cluster for a given article. The Newsblaster site still functions as of April 2014.

Also in 2002, Google launched Google News, the company's news content aggregator that gathers articles from sources across the world. The news displayed to users is selected based on attributes such as story location, time, and popularity. Google News is known for its use of collaborative filtering, whereby users actual consumption patterns and preferences are considered as part of its analysis [9]. While the precise mechanics of Google News are proprietary, it is known that traditional Google Search uses tf-idf as part of its search strategy. Mark Levene in his textbook An Introduction to Search Engines and Web Navigation, under a section titled Combining Link

Analysis with Content Relevance, writes: "The way in which Google combines the PageRank of a page with its content relevance measures such as TF-IDF is one of its well kept secrets" [19]. While the inner workings of Google News are not completely transparent, it is indeed likely that Google's News algorithms include tf-idf (or a similar variant thereof) as one of the metrics that it uses to evaluate and group news stories.

2.1.1 Named Entity Extraction

Beyond the macro picture of evaluating document level metrics, there are also more detailed pieces of information that can be useful and representative. A named entity extractor is a tool that given a string of text can extract and classify words of interest, including people, companies, organizations, places, etcetera. For instance, running the sentence "Washington D.C. is the capital city of the United States". through Calais [27] identifies two named entities in the sentence. It recognizes Washington D.C. as a city and the United States as a country. Calais uses machine learning, natural language processing and tagged data from Thomson Reuters data repository in order to discern named entities from the raw text. Named entity extraction tools used in conjunction with tf-idf techniques across a corpus, can be useful tools for both extracting keywords of interest from an item in the corpus (named entity extraction) as well as determining the relevance of particular entities to a given document (tf-idf).

2.2 News Visualization

In 1999, The City of News project from the MIT Media Lab, investigated novel 3D user interaction paradigms for news. City of News presented users with a virtual space in which they could see the relative importance of various stories, at a glance. For instance, the importance of a topic can be visually derived from size of a virtual building representing each topic. Topics are organized into "districts" of interest within the virtual space, so that users could navigate through financial, entertainment or shopping districts [34].

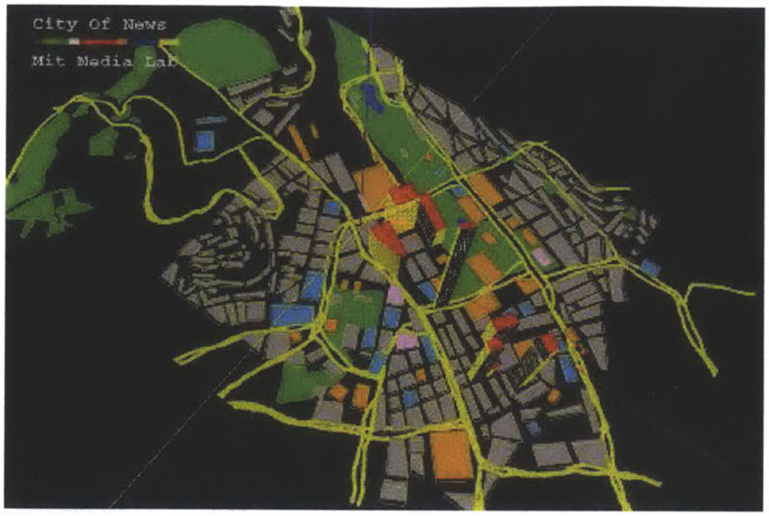


Figure 2-3: City of News.

Later work on the City of News also investigated the use of a gesture-based interfaces intended to allow users more direct manipulation of the 3D scene [35].

Newsmap is a treemap-based visualization of Google News. Using the treemap visualization, Newsmap indicates the importance of a story by the relative size of each story rectangle. Stories are also grouped by color to indicate various categories such as entertainment, health, technology or business.

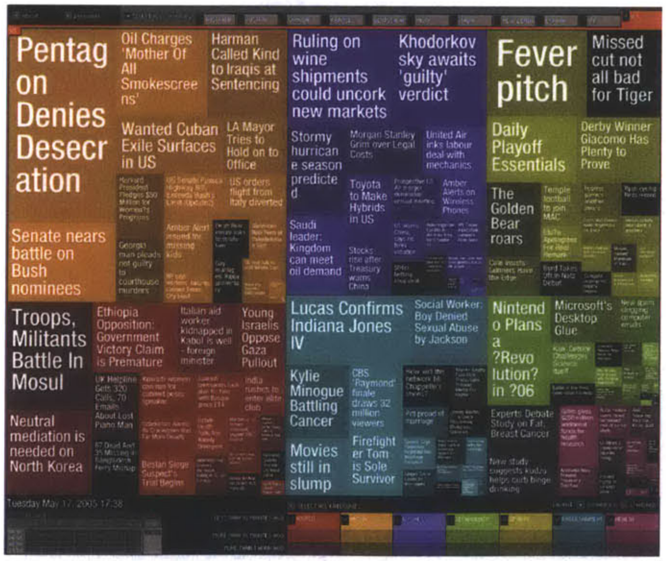
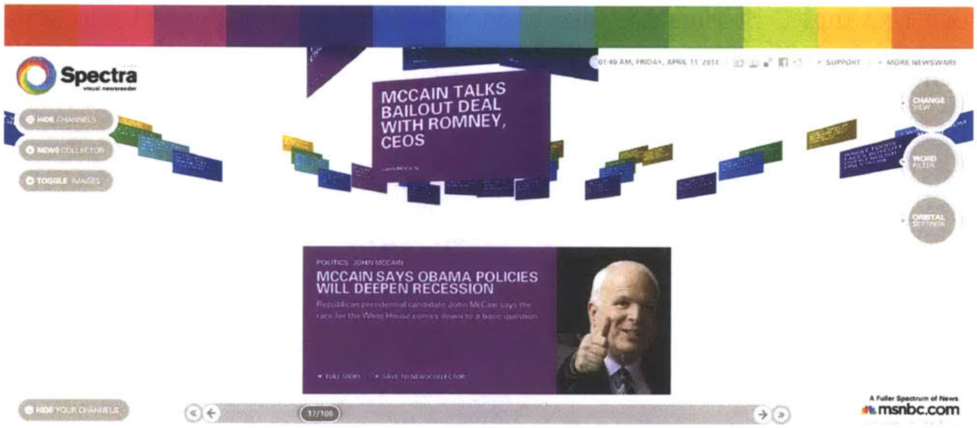

Pentag on Denies Desecr ation
Oil Charges 'Mother Of All Smokescree ns'
Harman Called Kind to Iraqis at Sentencing
Wanted Cuban Exile Surfaces in US
LA Mayor Tries to Hold on to Office
Senate nears battle on Bush nominees
Troops, Militants Battle In Mosul
Ethiopia Opposition: Government Victory Claim is Premature
Italian aid worker kidnapped in Kabul is well - foreign minister
Young Israelis Oppose Gaza Pullout
Neutral mediation is needed on North Korea
Ruling on wine shipments could uncork new markets
Khodorkov sky awaits 'guilty' verdict
Stormy hurrican e season predicte d
Toyota to Make Hybrids in US
Saudi leader: Kingdom can meet oil demand
Lucas Confirms Indiana Jones IV
Social Worker: Boy Denied Sexual Abuse by Jackson
Kylie Minogue Battling Cancer
CBS 'Raymond' finale draws 32 million viewers
Movies still in slump
Firefight er Tom is Sole Survivor
Fever pitch
Missed cut not all bad for Tiger
Daily Playoff Essentials
Derby Winner Giacomo Has Plenty to Prove
The Golden Bear roars
Nintend o Plans a ?Revo lution? in ?06
Microsoft's Desktop Glue
Experts Debate Study on Fat, Breast Cancer

Figure 2-4: Newsmap.

The interface supports filtering based on geographic data for a select group of countries, and also lets users limit the visualization, based on keyword search through articles titles, descriptions, or sources [36].

2.2.1 3D Visualization

Throughout the nineties, there was a flurry of academic activity to examine the use of 3D concepts for everyday interfaces [4, 21, 29, 28, 1, 34]. The work of human-computer interaction pioneers such as Stuart Card, George Robertson and Jock Mackinlay at Xerox PARC, examined the potential of overcoming the limitations of two-dimensional interfaces by intelligently using a three-dimensional representation that considered the biological underpinnings of the human perceptual system [21]. Yet, despite a history of academic work, todays popular media interfaces such as Google Search, Bing Search, YouTube, Apple TV, Netflix and news sites such as the New York Times, Washington Post and the Boston Globe, remain largely constrained to visualizations on a two-dimensional plane.



MCCAIN TALKS BAILOUT DEAL WITH ROMNEY, CEOS

POLITICS JOHN MCCAIN

MCCAIN SAYS OBAMA POLICIES WILL DEEPEN RECESSION

Republican presidential candidate John McCain says the race for the White House comes down to a basic question

Figure 2-5: Spectra.

One example of a 3D visualization applied to news is Spectra. Launched in 2008, Spectra by MSNBC provides a 3D news aggregator for NBC's news feeds. Users select the high level topics of interest such as health, technology, entertainment, politics etcetera. The visualization arrays the latest articles from the feeds on the topics selected by the user in a rotating 3D spiral. In addition to selecting broad topics, users can also perform keyword searches across the news feeds to include stories appearing in the feed about particular keywords of interest. Once a user selects an item of interest they wish to view, the stories are not embedded within the Spectra visualization itself, but rather the user is then redirected to a separate webpage where the content may be viewed.

As part of my thesis, I seek to reexamine the use of three-dimensional user interfaces for exploring world events in the current age of "Big Data" and when so much information is so readily capturable.

Chapter 3

Motivation

During early experimentations with user interface prototypes, it became clear that in order to develop dynamic, data-driven media experiences, we would need to invest a substantial amount of effort to building a suitable infrastructure. Towards this end, in collaboration with Dan Sawada and Robert Hemsley, we created a media gathering and analysis framework we call Constellation. The goal of Constellation is to create a smart repository for media that can power a host of user-facing applications that depend on visual media. Constellation provides the data that powers the WorldLens interface.

3.1 Constellation

In the World Wide Web, there are basic semantic markers that make it possible to extract information and order from hypertext documents. The all important hyperlink tag[1], for example, explicitly maps out connections between various webpages, and becomes a useful tool for analysis in applications such as Google's PageRank [24]. Other HTML tags such as, the header tags, emphasis tags (bold, italic, underlined) and more recently completely explicit OpenGraph metatags (e.g. og:description), provide markers for differentiating importance of content within a document. These

[1] The <a> tag in HTML specifies a link relationship from one webpage to another and typically contains a URL as well as text describing the link. For instance <a href="target.html">Description </a>

structures embedded in hypertext facilitate the creation of useful information retrieval applications.

Video content, however, does not contain quite the same structured and useful information at present. Data such as closed captions, audio and image frames are typically encapsulated within video container formats without semantic enrichment. [2] Therefore, unlike the web where the content creators are the ones explicitly making structures apparent, we must perform additional work in an effort to recover the original structure embedded within the video so that we may better analyze the content.

[2] Occasionally video captions may contain indications to identify speakers but this is neither consistent across videos nor standard in whether the name of the speaker is identified or if simply several marker characters are used.

3.2 Architecture

3.2.1 Overview Diagram

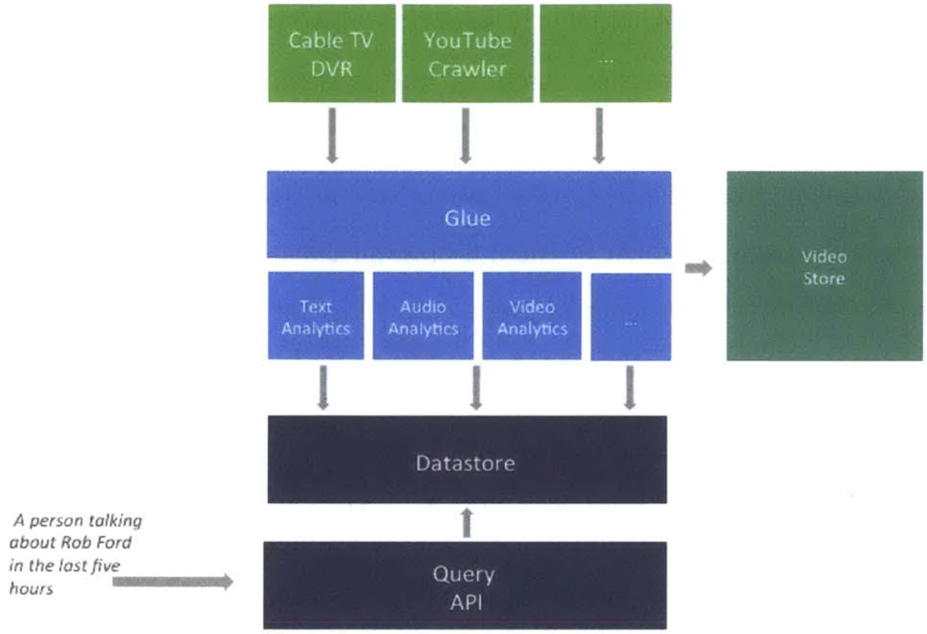


Figure 3-1: A high level overview of the architectural design of the Constellation system.

3.2.2 DVR and Web Crawler

The Constellation system begins with its inputs. At the Media Lab, we currently have thirty live feeds available to us from DirecTV. We created a programmable DVR that can record and transcode television programs from the Lab's UDP network streams. We also have access to data from Tribune Media Services (TMS) from which we are able to retrieve metadata regarding the current television lineup. With both the content from DirecTV and the metadata from TMS we are able to, for example, record every television program available to use under the category of news. We can also target individual programs of interest for either inclusion of exclusion from our recording. In order to both capture the streams from the network, and also put them in a format suitable for visualization purposes, we use FFmpeg[3] to transcode the streams.

In addition to content captured from live broadcast feeds, we also created a web video crawler that can scour the Internet for video and feed it through Constellation. The web video crawler currently supports video content from Youtube as well as the News Archive's Television Archive, and is readily extendable to other video content repository. The web video crawler is currently seeded with Youtube news channels. Starting from the seed news sources, the crawler downloads, transcodes, and downsamples videos which it encounters and then dispatches the downsampled content to Constellation, while retaining a copy of the original high resolution content for user applications. The crawler is built using the Python Scrapy [14] framework, which allows for the rapid development of web crawlers as well as RabbitMQ [32] which provides the ability to create robust and distributed message queues. By using RabbitMQ, the crawler instance is able to continuously crawl the web for video while adding metadata for extractable content such as Youtube video URLs to a work queue. Then, separate worker processes are run in parallel to take videos off the queue, download the content, perform transcoding and consequently send video through Constellation for processing.

3.2.3 Core

Once the DVR or crawler has captured, transcoded, and downsampled content, a notification message is sent to the Constellation Core via HTTP containing metadata about the incoming content in JSON format [6]. Constellation by design, will then download the relevant media content specified in the notification JSON message. To avoid creating a monolithic system, this layer of abstraction between the Constellation core and the content sources is included to enhance scalability and modularity. Due to this design, an arbitrary number of DVR or Crawler services can be spawned across various machines to feed content through Constellation. Consequently, the

[3] The program FFmpeg is a useful tool for capturing, encoding and manipulating a wide variety of video stream and file formats [2]

Table 3.1: Selected Constellation Modules

Module	Description
Broadcast	Sends out a message over the network to inform user-facing applications that a new video has been processed.
Face Detector	Uses OpenCV's Haar Classifier to detect faces within frames of video.
Frequency	Examines the closed captioning data and processes it through both the Open Calais and Alchemy API tools for named entity extraction.
Thumbnail	Creates thumbnail images to represent the video as well as an animated GIF of each frame at the scene boundary frames.
Scene	Detects scene cuts using the Openshot library

Constellation Core can simply download the files from the remote server, process the files, generate the appropriate metadata then delete the files after processing, as long term storage is provided by the original modules which retain higher resolution versions of the content for display purposes.

Once a video is downloaded successfully by the Constellation core, it is then ready for processing by individual modules.

3.2.4 Modules

In Constellation's modular architecture, we can write code that will plug in to the media processing pipeline. For instance, we can create a module that is notified once a new video is downloaded by Constellation, and then could process the video container to analyze anything from closed-captioning, to audio, to the actual video. Furthermore, we can also create modules that can provide additional information at arbitrary points in the pipeline.

The broadcast module can send a broadcast message to notify external listeners (such as user-facing applications) that a new video has finished processing. This message is sent over the Ultimate Media Events server (developed by Jon Ferguson) which is a websocket message bus, and which allows any application to broadcast

Table 3.2: Constellation Events

Constellation Event	Descipriton
StartFileDownload	Starting to download an incoming video
EndFileDownload	Finished downloading an incoming video
ProcessRequest	New data received
VideoReady	New video ready for processing
Cleanup	Finished all processing

messages to which desired client applications can choose to listen.

The face detector module runs through the frames of each video at a given interval (currently, it checks for a face every second in the video). When faces are detected, the module tracks how long the faces remain detectable, and provides a timeline of both when faces are viewable in scenes, as well as how many are detected. This module uses OpenCV [11] and its included Haar Feature-based Cascade Classifier [12]. In particular, we currently use the classifier to detect objects matching OpenCV's frontal face training dataset.

The frequency module processes the closed captioning data from a video, which is originally received in time coded SubRip (SRT) form. First the original time coded information is stored. Then the module converts the time coded data into plain text. Once in plain text form, the data is processed through both the OpenCalais and Alchemy APIs to extract named entities.

The scene module uses the open source Shotdetect tool created by John Mathe, to detect scene boundaries in video. In conjuction, with the thumbnail module this module allows us to generate a thumbnail for each distinct scene detected in the video, which we also use to generate a GIF containing the various scenes. Additionally, the scene module also helps us extract the relevant portion of a video when querying based on text. Instead of retrieving the entire video when text is matched, we can retrieve the relevant scene from in which the text is contained, by combining the data from the frequency and scene modules.

3.2.5 Database and Queries

The metadata generated by Constellation and the associated modules, are stored in a MongoDB database [7]. MongoDB was chosen due to its ease of use, compatibility with Python in which most of Constellation is written, and its NoSQL characteristics which lend flexibility. As opposed to traditional relational databases, such as MySQL or Postgres, MongoDB provides its users with the ability to dynamically update schemas. This makes adding or removing features and their associated persistent data quite simple in applications that run on top of MongoDB. In particular, the use of MongoDB allows us to rapidly add and remove modules to and from Constellation without having to update our database's schema each time a change is made.

While MongoDB does provide ease of use and flexibility for agile development, these benefits do come at a cost. In particular, complex queries involving nested data structures can be slow. To remedy this drawback, we utilize Apache Solr [33] to perform indexing on the MongoDB database to speed up queries necessary for retrieving information from particular modules (for example, looking up all videos mentioning the country of Russia).

3.3 Article News Gathering

The basis for the article news gathering is a crawler, which like the web video crawler, is written in Python using the Scrapy Framework. Scrapy provides a reasonable skeleton on top of which custom web crawlers can quickly be written, by handling the crawling queue, and the easy creation of crawling rules. For example, Scrapy provides the ability to blacklist particular domains or URL types, or trigger special functionality such as when a page with a particular URL structure is encountered. The news crawler is first seeded to crawl Google News, which serves as a basis for the world event model and provides a starting point for determining both story clusters, and trending topics at the time of the crawl. Links to related stories clustered together in Google News are extracted by parsing the structure of the news page and are associated as a base group of links on a related story. After the group of links are extracted, the crawler goes to the original news article pages and extracts the title of the article, publication date and time, as well as the page itself. Additionally, a screenshot of the page is captured at the time of the crawl by the Web Thumbnail Service (see more on the subsection for Web Thumbnail Service ahead). Once the data has been extracted from the page, the contents of each retrieved site's title and body, are fed through two named entity extractors: Alchemy and Calais. In an effort to reduce spurious named entity recognition, the output of each extractor is treated as a vote and only unanimously voted entities are then registered as entities for the given news event.

In order to relate videos to news content, we rely on the closed captioning and the time of publication. As with the news articles, we run the video closed captioning through both Calais and Alchemy named entity extractors and use results from both systems as votes and only select unanimously agreed upon entities. Since video content typically trails articles, when a new video comes through from the web video crawler and a broadcast message is sent, we query the news API for news events 24 hours before and after the event start time. Both the Alchemy and Calais APIs provide relevance estimates, which aim to predict the relevance of a given named entity extracted to the text submitted for analysis. If there are at least two entities in common between the video and a news story cluster with relevancy[4] thresholds above twenty percent, then a match is considered to have been found and an associated is created between the video and the news story cluster.

3.3.1 Web Thumbnail Service

The reason for gathering news articles is in order to display them, however it would be impractical to load thousands or even hundreds of nested web pages (i.e. iframes) into the browser's Document Object Model (DOM) at this time. To work around this limitation, as mentioned previously, each time a news web page is visited by the crawler, it dispatches a separate thread to capture an image of the page at the time

[4] The relevancy is a value returned by both the Alchemy and OpenCalais APIs and is a measure for the relevance of an entity to a given portion of text.

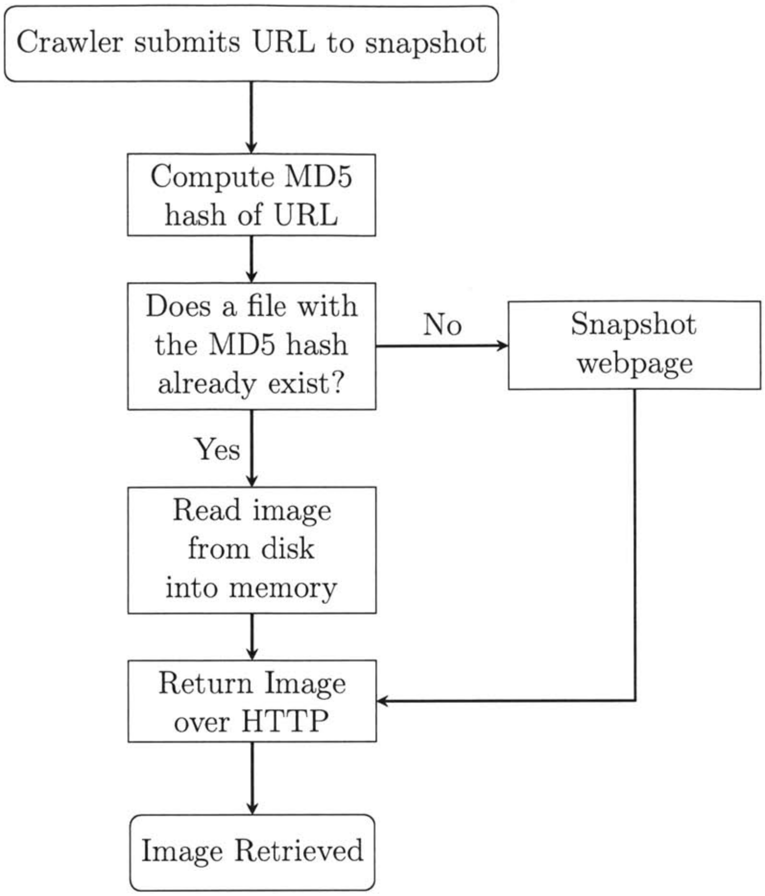

Crawler submits URL to snapshot
Compute MD5 hash of URL
Does a file with the MD5 hash already exist?
No
Snapshot webpage
Yes
Read image from disk into memory
Return Image over HTTP
Image Retrieved

Figure 3-2: Flow diagram for the Web Thumbnail Service

of the visit. Due to the large number of web pages whose thumbnails would need to be indexed, we needed to create a service that would not simply display an image on demand, but also that would cache the results.

In order to achieve this, we built a web and file system wrapper around the node-webshot library [18]. The node-webshot library itself is a wrapper around PhantomJS, which is a ''headless Webkit scriptable" browser client [13]. Images are captured, as necessary, by having a PhantomJS client navigate to a target page and then the rendered HTML page is converted in memory to an image.

The Web Thumbnail Service accepts a URL as input. First, the MD5 hash of the URL is computed. Then a file system lookup is performed to examine if a file exists in the web snapshot directory with the filename equivalent to the MD5 hash of the input URL. If a file corresponding to the computed hash exists, by default the service returns the existing file. If no such file exists, the service utilizes the aforementioned node-webshot library to capture an image of the page and returns the newly generated image.The service also accepts an optional parameter of a snapshot overwrite flag. If this flag is set to true, then the service will attempt to retrieve the latest version of the website and overwrite any previous image.

Chapter 4

Design Iterations

4.1 First Iteration

From the early stages of the Ultimate Media initiative, we were intrigued by the possibilities of creating a 3D user interface and experience. One of our first iterations towards a visual media content explorer was inspired by a visualization created by Bluefin Labs.

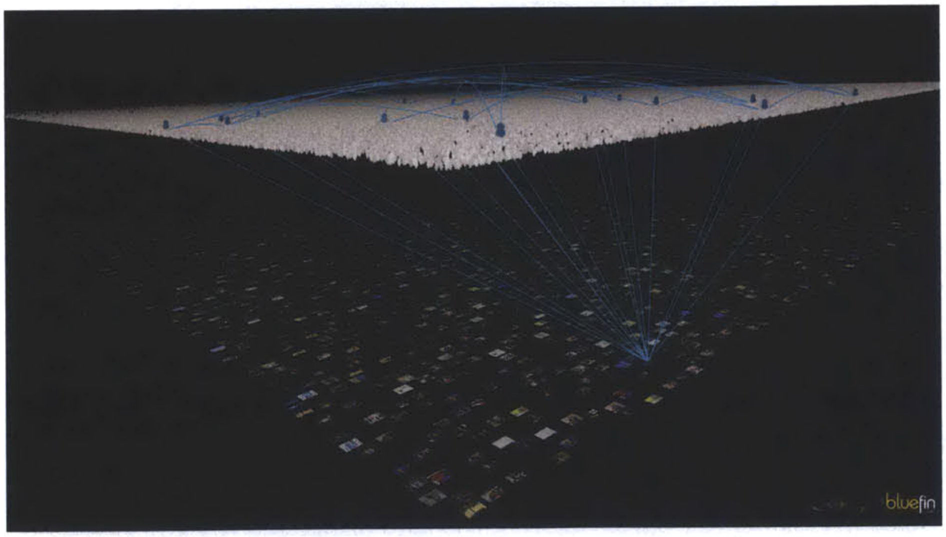


Figure 4-1: BlueFin Labs illustration

The visualization illustrates two planes: at the bottom of the figure is content plane and above it in parallel is the social plane. Lines highlight relationships between people and content, while arcs represent social connections between people. Building off this initial illustration, I created a first version of an interactive 3D content navigator.

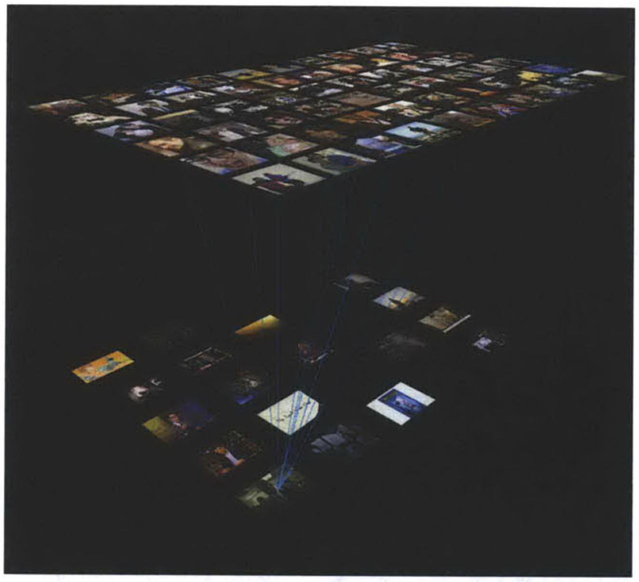


Figure 4-2: First design iteration for 3D media navigator

As in the Bluefin illustration, the first prototype had two planes: a social and content plane. Lines between the two planes highlighted relationships between content and people (for example, which people like which programs). Different categories of content could quickly be recalled from the plane through the use of spherical topic selectors. Once a topic selector was chosen, the related content for that topic would fly forward, and the user could simply click nearby videos to cycle through the set of content.

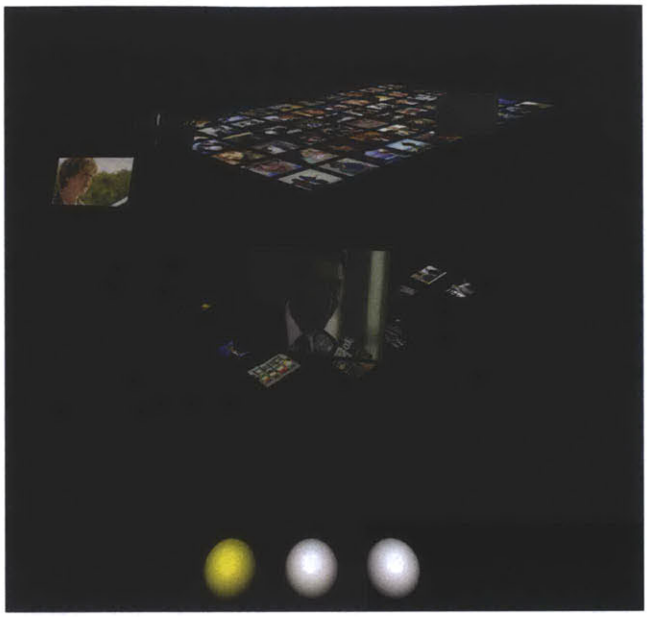


Figure 4-3: Selecting a topic sphere would pull relevant content to the forefront

The main impetus behind introducing the concept of a content plane was to challenge the traditional notion of channels. Channels are an essential component of a broadcast system and today's television interfaces are built around the structure of broadcasters and content providers. However, the notion of a channel does not necessarily reflect an optimal experience for users. With the content plane videos could be laid out in a grid, and users could utilize selectors to filter through to the content they liked and simply zoom in to start watching.

4.2 Second Iteration

While in the first iteration we began by looking at exploring narrative entertainment, my focus quickly shifted towards examining news content. While the focus shifted to news, I still wanted to carry over some of the principles we had started focusing on in the narrative space. A second design prototype looked at mapping news content to a sphere. The idea was to bring to life the convenience of scrolling across a globe to quickly and seamlessly discover content and news around the world.

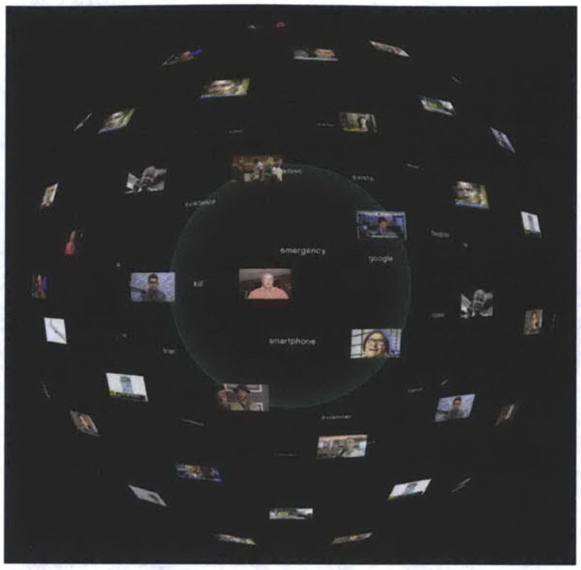


Figure 4-4: Spherical arrangement of content.

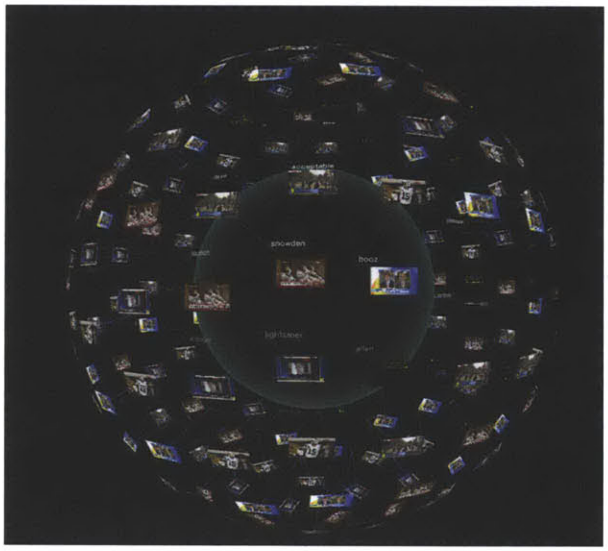


Figure 4-5: Adding in lines to show connections.

The design also included an inner word cloud as an atmospheric layer that would provide additional information relating to keywords in nearby content. This addition was meant to begin a journey towards creating data-driven interfaces. Other steps in the direction of utilizing data behind the content of the globe arrangement motif included an iteration where links were drawn between the inner word cloud layer and video. One drawback, however to this particular approach was that the 3D lines tended to clutter the visualization and it was difficult to get an understanding of the related content at a glance when attempting to highlight multiple connections at once.

4.3 Third Iteration

A third design prototype looked at a more, two-dimensional grid like structure while retaining use of the z-axis for descriptive categorization.

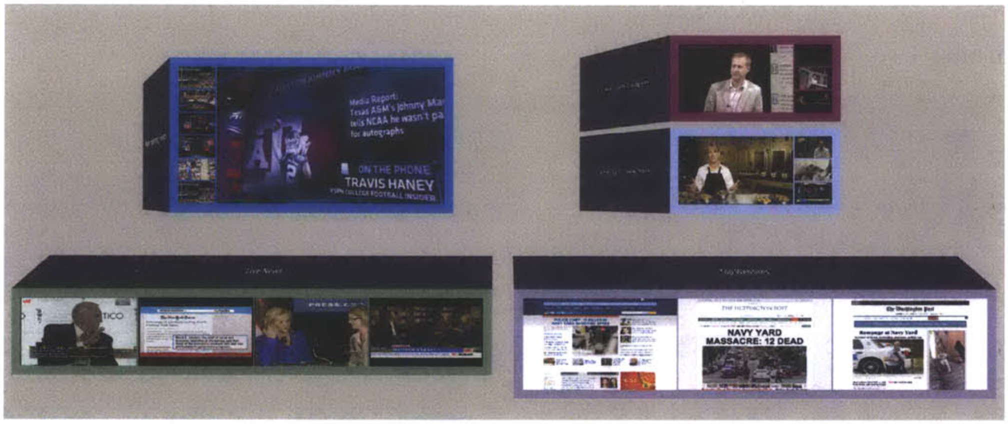


Figure 4-6: 2D layout with z-axis used for description.

This design iteration examined using distance in the x and y planes to indicate topical similarity. While previous iterations focused solely on visualizing video content, this iteration began considering the integration of other media such as still images and webpages.

One of the concepts that this iteration considered exploring, but was ultimately abandoned, was a *tabula rasa* approach, where users would start with a minimally populated interface and could then populate the space with keywords, which would be represented by the colored group planes. The idea was to allow users the option to generate, prioritize and layout content planes of interest. Instead of remaining static these content planes could be "pinned" to keywords or topics such that the latest videos for the given keyword or topic would appear in the particular plane. The interface supported spatially rearranging these planes via dragging and dropping.

Additionally, this iteration experimented with the inclusion of live content from broadcast television. Live television was incorporated in HTML 5 video by capturing and transcoding the UDP webstream and converting it into an HTTP WebM container format stream using the VP8 and Ogg video and audio codecs respectively.

4.4 Fourth Iteration

A fourth design prototype returned to the basics of a fundamentally two-dimensional grid, but retained the zoom capabilities of a camera that could pan over a large sheet of content.

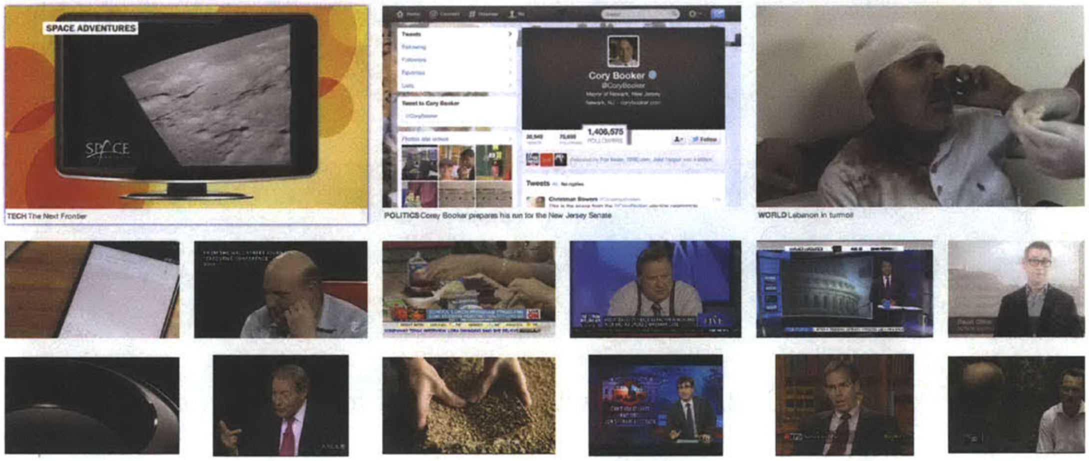
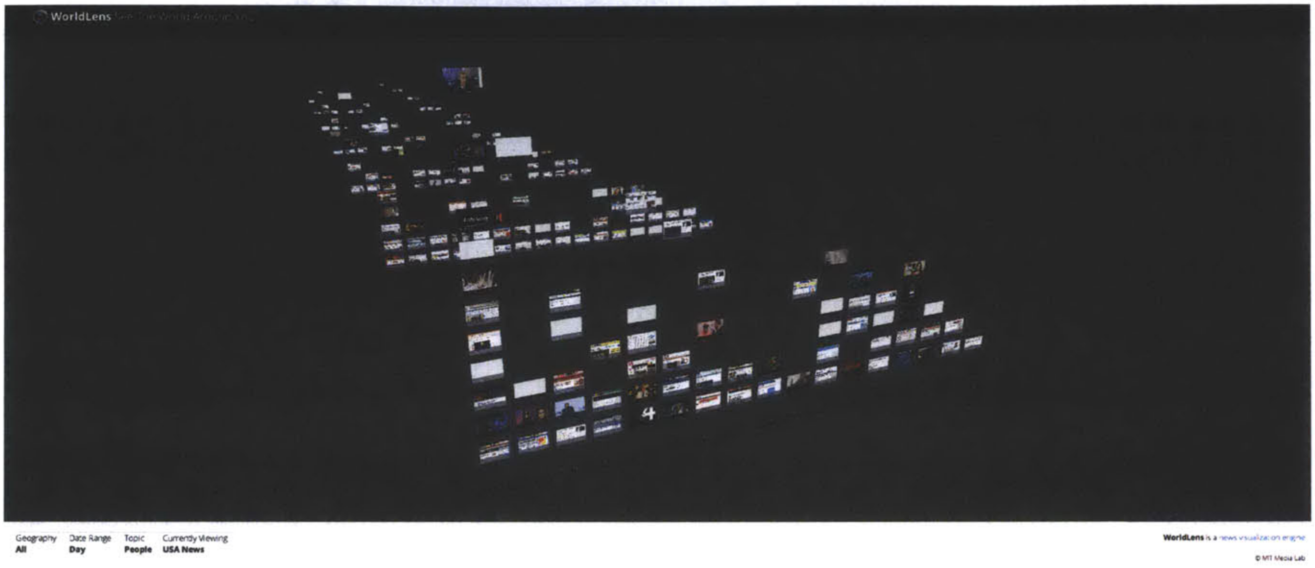


Figure 4-7: 2D media grid

In this iteration, elements could be rearranged as users could drag, drop and delete items from the grid at will, and resize items. The concept at this stage was to explore the notion of allowing users the ability to semantically "pin" content. For instance, the interface explored allowing users the ability to pin videos relating to "politics" to the top of the media grid.

4.5 Fifth Iteration

The fifth iteration examined the use of a three-dimensional grid. In this iteration, news items were laid out in space such that the x-axis represented the time of a particular news event, the y-axis represented media coverage, and the z-axis provided space for perusing the related media to the story. With the final use case of a touch screen in mind, this iteration integrated the use of multitouch gestures such as pinch to zoom, pan, and swipe.



Geography All
Date Range Day
Topic People
Currently Viewing USA News

WorldLens is a news visualization engine

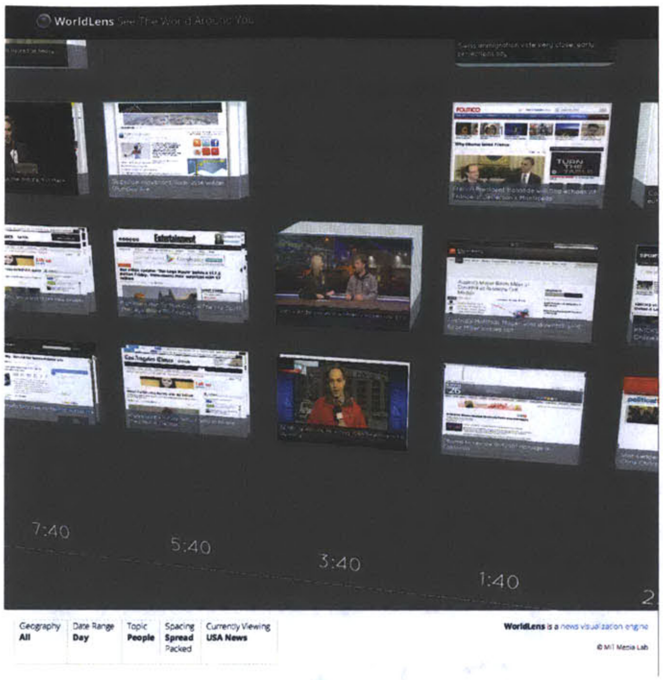
© MIT Media Lab

Figure 4-8: The WorldLens 3D media grid

4.6 Sixth Iteration

The sixth iteration refined the original approach of the original three-dimensional grid approach. The spacing between items on the z-axis was dramatically reduced, as this appeared to provide a clearer representation. With items far away, as the camera panned or rotated, items in the distant background would become distracting would often confuse the visual aesthetic of discrete grid rows and columns. By packing items closer together, the visual appearance of a three-dimensional grid remained more intact.

This sixth iteration also included additional components beyond the three-dimensional grid. This iteration included the ability to rotate the camera such that the grid could be viewed as a histogram highlighting the most frequently mentioned stories, as well as basic feature for filtering stories by country.



Geography All | Date Range Day | Topic People | Spacing Spread Packed | Currently Viewing USA News

WorldLens is a news visualization engine

© MIT Media Lab

Figure 4-9: Early 3D media grid

The impetus for adding these additional views, was primarily driven to facilitate the exploration process for users. By including these two additional avenues, the World Lens provides users three primary dimensions among which users can freely explore depending on how they want to understand what is happening, be it based on time, category, or location.

Throughout these designs and preliminary user tests, certain lessons emerged. Arbitrary use of 3D or even 2D space can be confusing and underutilizes the affordances of space. Lines must be used judiciously so that they serve to clarify and not clutter visualization. Text remains an important way to learn more, so even in visualizations that rely heavily on images and video, there should a way for a user to get more textual information as needed.

Chapter 5

Interface, Interaction and Evaluation

5.1 Interface and Interaction

The final version of the interface includes elements drawn from the progressive evolution of previous design iterations. While my personal interests in the visualization lay deeply tied to the temporal aspect of news, I found through discussion with users that a topical view and a geographic view were elements of exploration that merited consideration for inclusion in the interface.

5.1.1 Grid View

The interface begins with a presentation of the news arrayed in a three-dimensional grid. The x-axis of the grid represents time, the y-axis is inversely proportional to story frequency, and the z-axis contains the depth, or related content for each story.

Users can pan over the news grid, pinch to zoom in or out, and double tap a story to activate the original content. When a user double taps a story item that is a news article, the item expands and the original webpage is presented, with which the user can interact. This allows users the ability to see stories at a glance and in such a way that they can easily read headlines and see prominent images in article pages, but

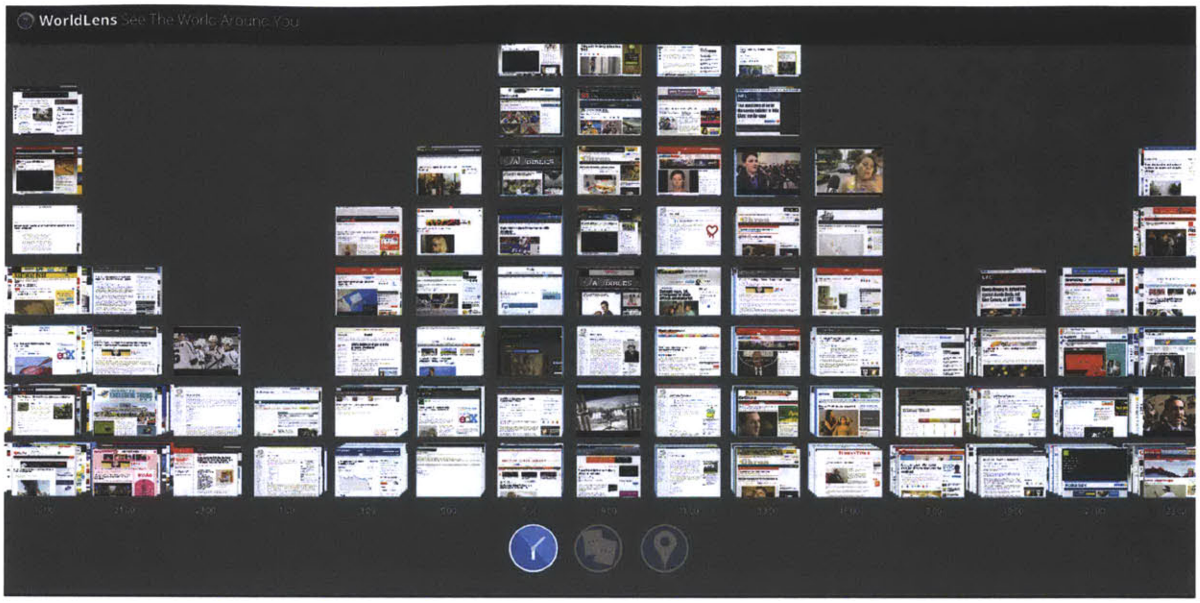
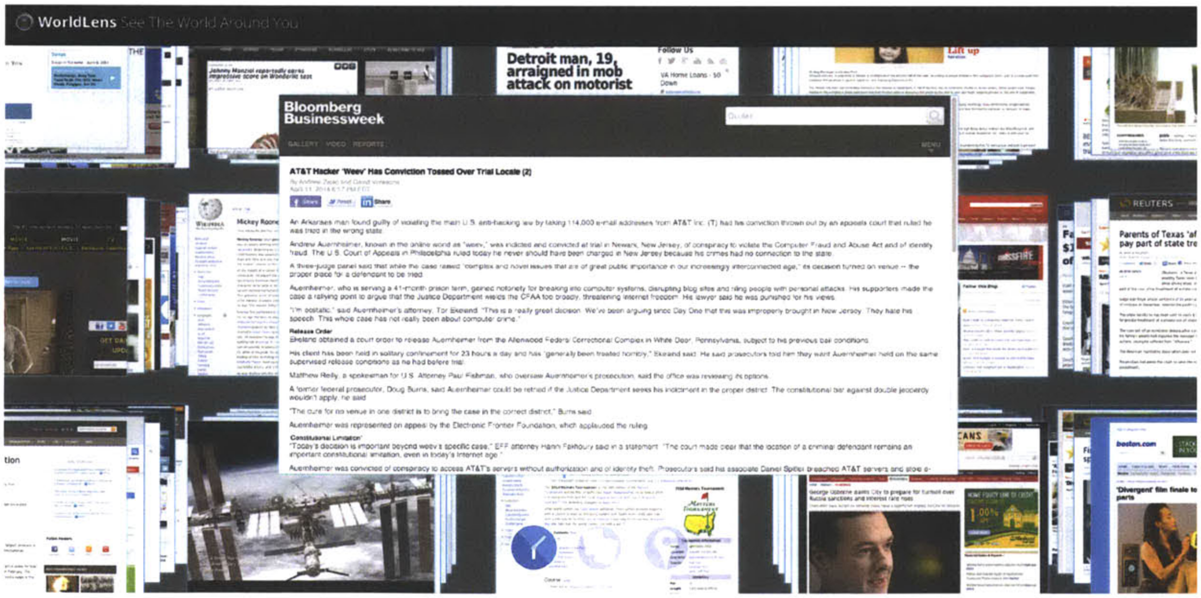


Figure 5-1: The grid view.

at the same time easily access the original article for more in depth information as needed.


WorldLens See The World Around You
Detroit man, 19, arraigned in mob attack on motorist
Follow Us
VA Home Loans - $0 Down
Lift up
Bloomberg Businessweek
GALLERY VIDEO REPORTS
MENU
Share
Tweet
Share
REUTERS
Parents of Texas 'af pay part of state tre
George Osborne alerts City to prepare for turmoil over Russia sanctions and interest rate rises
'Divergent' film finale to
Masters Tournament

AT&T Hacker 'Weev' Has Conviction Tossed Over Trial Locale (2)

By Andrew Zajac and David Voreacos
April 11, 2014 6:17 PM EDT

An Arkansas man found guilty of violating the main U.S. anti-hacking law by taking 114,000 e-mail addresses from AT&T Inc. (T) had his conviction thrown out by an appeals court that ruled he was tried in the wrong state.

Andrew Auernheimer, known in the online world as "weev," was indicted and convicted at trial in Newark, New Jersey, of conspiracy to violate the Computer Fraud and Abuse Act and of identity fraud. The U.S. Court of Appeals in Philadelphia ruled today he never should have been charged in New Jersey because his crimes had no connection to the state.

A three-judge panel said that while the case raised "complex and novel issues that are of great public importance in our increasingly interconnected age," its decision turned on venue -- the proper place for a defendant to be tried.

Auernheimer, who is serving a 41-month prison term, gained notoriety for breaking into computer systems, disrupting blog sites and riling people with personal attacks. His supporters made the case a rallying point to argue that the Justice Department wields the CFAA too broadly, threatening Internet freedom. His lawyer said he was punished for his views.

"I'm ecstatic," said Auernheimer's attorney, Tor Ekeland. "This is a really great decision. We've been arguing since Day One that this was improperly brought in New Jersey. They hate his speech. This whole case has not really been about computer crime."

Release Order

Ekeland obtained a court order to release Auernheimer from the Allenwood Federal Correctional Complex in White Deer, Pennsylvania, subject to his previous bail conditions.

His client has been held in solitary confinement for 23 hours a day and has "generally been treated horribly," Ekeland said. He said prosecutors told him they want Auernheimer held on the same supervised release conditions as he had before trial.

Matthew Reilly, a spokesman for U.S. Attorney Paul Fishman, who oversaw Auernheimer's prosecution, said the office was reviewing its options.

A former federal prosecutor, Doug Burns, said Auernheimer could be retried if the Justice Department seeks his indictment in the proper district. The constitutional bar against double jeopardy wouldn't apply, he said.

"The cure for no venue in one district is to bring the case in the correct district," Burns said.

Auernheimer was represented on appeal by the Electronic Frontier Foundation, which applauded the ruling.

Constitutional Limitation'

"Today's decision is important beyond weev's specific case," EFF attorney Hanni Fakhoury said in a statement. "The court made clear that the location of a criminal defendant remains an important constitutional limitation, even in today's Internet age."

Auernheimer was convicted of conspiracy to access AT&T's servers without authorization and of identity theft. Prosecutors said his associate Daniel Spitler breached AT&T servers and stole e-

Figure 5-2: The grid view with a news article selected.

For video content, the content selected by the user with the double tap action is played. Users can swipe an event stack to see more content about the same story, thereby allowing users to explore the z-axis of the grid.

5.1.2 Topic View

Users can also switch to a topical view, which presents a 3D bar chart style visualization of the most trending story groups in the news. In this view of the data, users can also swipe through the grouped stacks to get a glance of media about a particular story.

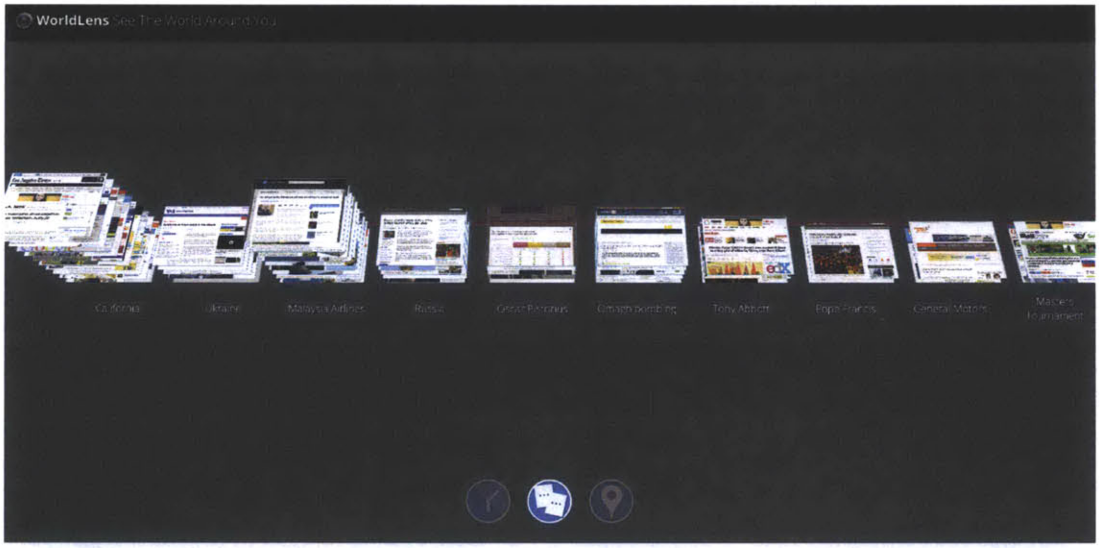


Figure 5-3: The topic view.

Further, users can also double tap a grouped topic stack to launch into a timeline grid for the story. This view affords users a sense of how a given topic has ebbed and flowed over time.

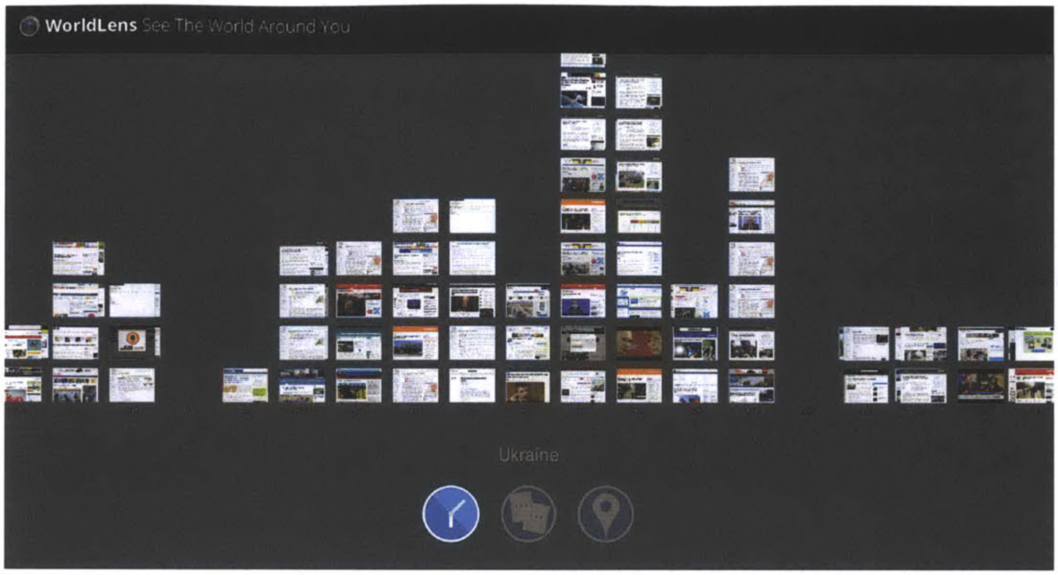


Figure 5-4: The topic history view.

5.1.3 Geographic View

The geographic view of the world is rendered using the D3 Datamaps library [10]. The Datamaps library provides a convenient SVG based representation of the world, making it easy to individually address countries of interest.

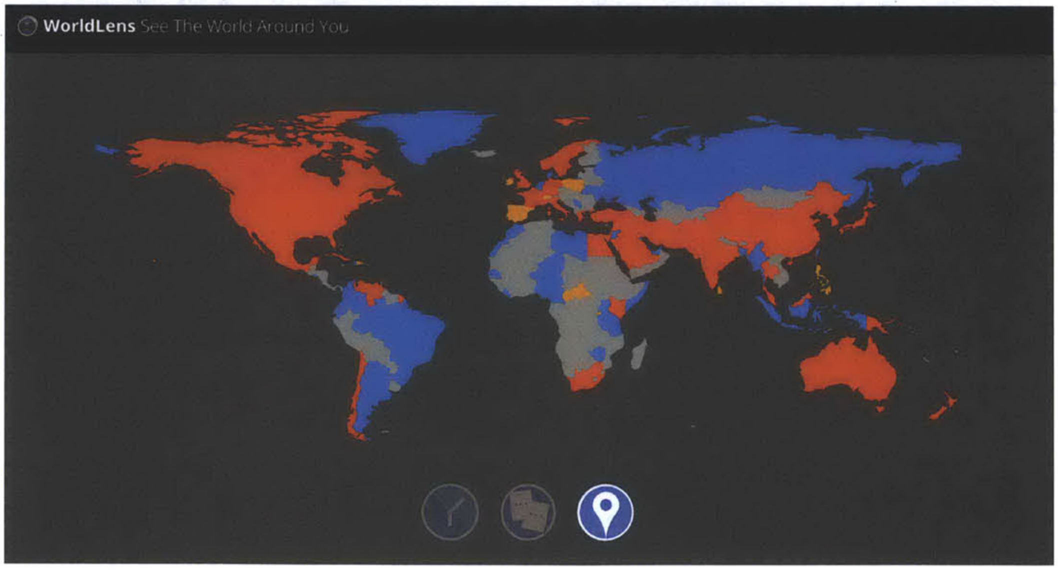


Figure 5-5: The geographic view.

The map view allows users to quickly browse the world for news. The map is augmented with a heatmap visualization, illustrating on a color scale the countries with recent activity. The color of each country is determined based on a score calculated by computing the time delta for each story cluster discussing the country times the number of related media items in each story cluster. The data are divided into three groups, where the lowest third of scores are colored blue, the next third are orange, and the highest scoring third are colored red.

Users can simply pan and zoom over the map. When users see a country of interest, they may select a country by tapping it. The visualization will then shift to nearly center the selected country and will display a stack of related stories. The stack of stories are kept slightly off center with the three-dimensional camera view so as to highlight the depth of the stack of stories related to the particular country.

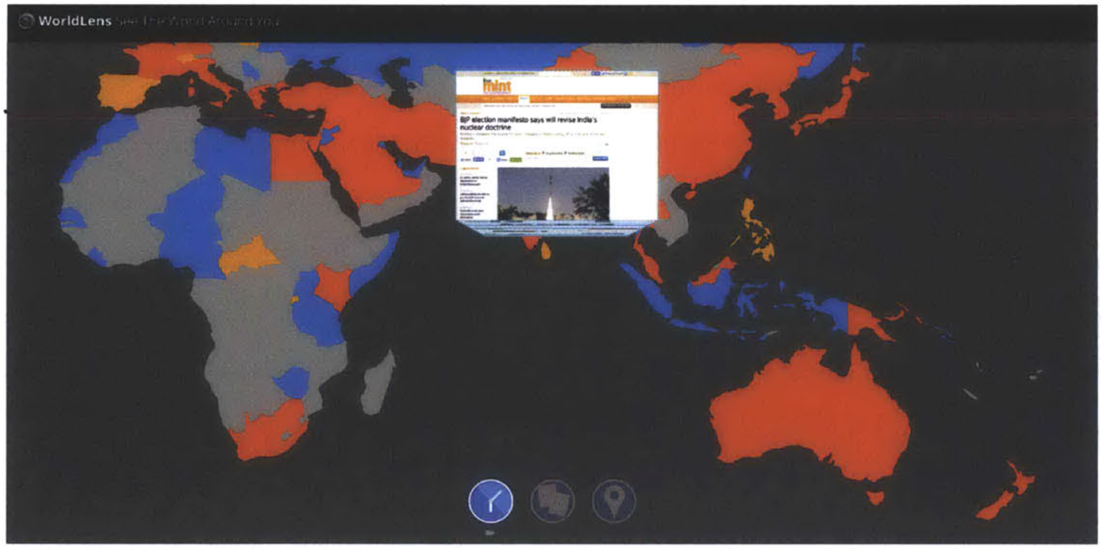

WorldLens See The World Around You
mint
BJP election manifesto says will revise India's nuclear doctrine

Figure 5-6: The stack of related stories in the map view.

Once a country stack is active, users may swipe backwards or forwards to move through the stories which are initially displayed in reverse chronological order. When a user finds a story of interest they may double tap the story to expand and activate the related content.

5.2 User Evaluation

A user study was conducted to examine the WorldLens system (see Appendix B). Users were asked to perform a series of tasks that involved navigating the WorldLens interface, as well as tasks that were aimed at evaluating users' understanding of the system. After the tasks were performed, users were asked a series of questions to evaluate their experiences.

5.2.1 Equipment

Evaluations were performed running the WorldLens system on a Planar 84-inch multitouch display. The display was connected to a custom built computer with the following core components: an Intel i7-4771 3.5 gigahertz quadcore processor, 16 gigabytes of Corsair DDR3 RAM, an NVIDIA K5000 containing 4 gigabytes of GDDR5 video memory, a 1 terrabyte hard drive with 7200 RPM from Seagate, and a gigabit ethernet network interface from the ECS Z87H3-A2X motherboard. The computer was running a 64-bit version of the Ubuntu 13.04 operating system.

5.2.2 Collected Data

By way of background, users were asked regarding any prior experience with 3D interfaces. 75% of the users surveyed responded that they used 3D interfaces (excluding video games) less than one time per month.

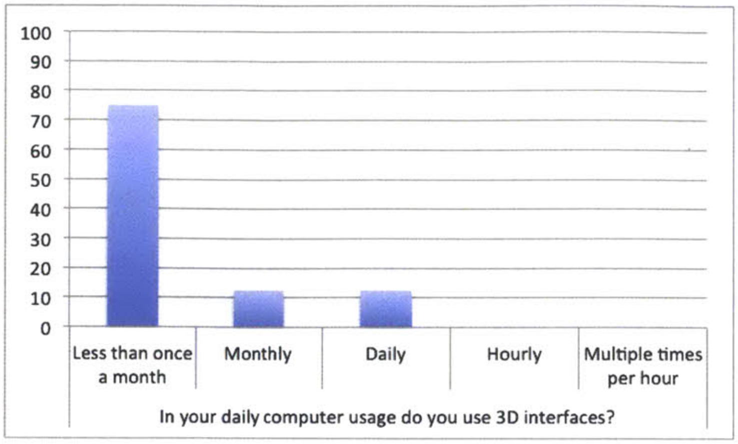


Figure 5-7: Usage frequency of 3D interfaces (excluding video games)

These responses align with my initial observation in 2.2.1 that the mainstream adoption of 3D interfaces remains limited.

Additionally as a background question users were asked what they considered to be the most important facet of a news story.

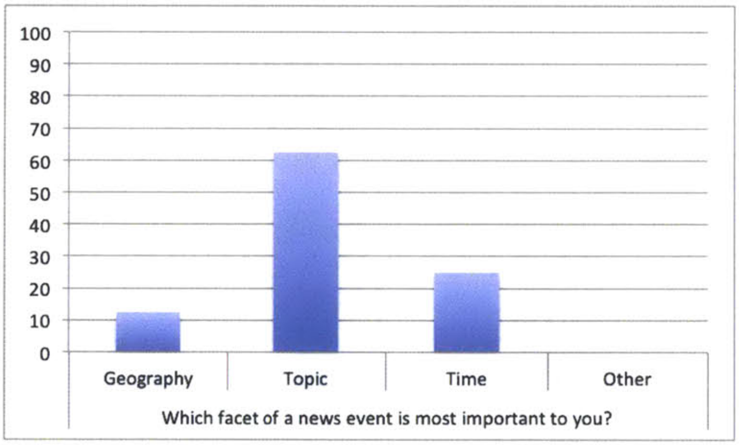


Figure 5-8: Most important facet of a news story

Over 60% of responses indicated that the topic of a news story was the most important facet of a news story. This was an interesting result for me, as I had previously considered the temporal facet (i.e. the recency) as the most important facet.

Fully 100% of users responded that using WorldLens they would be at least more likely to explore news from other countries.

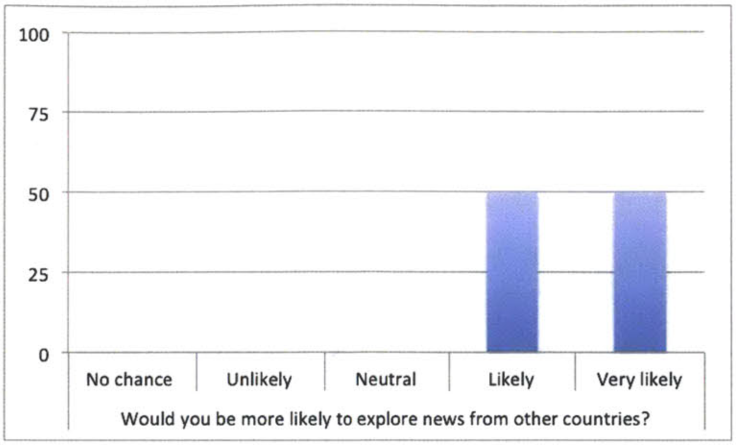


Figure 5-9: Map exploration inclination after using WorldLens

50% of users indicated that they prefer a 2D interface for news exploration.

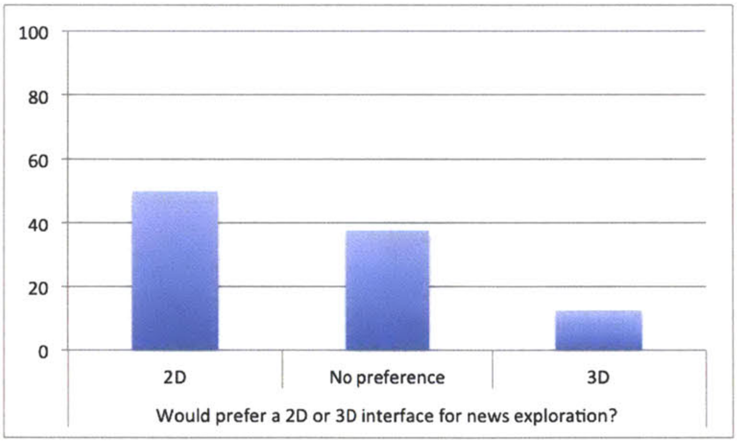


Figure 5-10: Dimensional preferences for news exploration

The aforementioned preference notwithstanding, 75% responded that they felt WorldLens strongly provided them with an understanding of what is happening in the world.

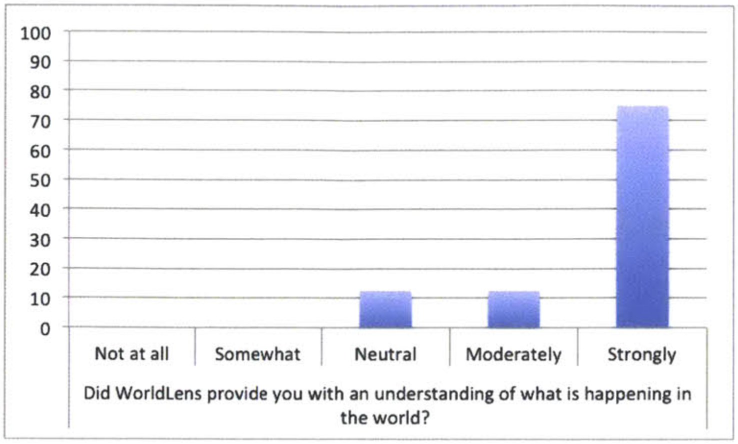


Figure 5-11: World happenings

87.5% of users reported the overall usefulness of WorldLens as good or excellent.

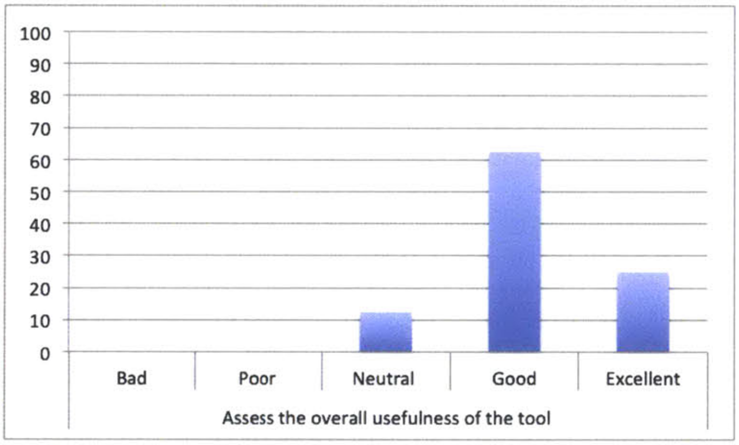


Figure 5-12: Overall usefulness

5.2.3 User Feedback

Users provided valuable feedback based on their experience utilizing WorldLens. Some of the feedback included enhancements around interactions, for instance: providing multiple way to minimize content in WorldLens after it is expanded, and adding a labeled axis to highlight the spatial mappings of the 3D time grid. Additionally, several asked about the possibility for versions of WorldLens that could run on phones, tablets or laptops with multitouch.

Chapter 6

Future Work and Additional Applications

6.1 Mobility

The ubiquity of smartphones and tablets and consequently the large number of users who rely on their mobile devices for the latest information, make the development of a mobile interface a promising next step. It turns out that since the WorldLens user interface was written using WebGL, it actually is technically able to run on modern mobile devices supporting web implementations of OpenGL. However, as the experience was developed with a large screen in mind the visualization is very taxing on utilzation across: CPU for animating many elements simultaneously, network for simulatenously loading many images, and memory (both main and GPU) usage for storing the data structures and loading images off the network and into textures that can be manipulated in 3D.

One quick and potential solution would be to simply scale down the number of elements, so that the current visualization runs smoothly on mobile devices. However, I think that the mobile experience deserves a solution that isn't a mere retrofitting, as other design considerations come into mind with the mobile environment such as intermittent network connectivity and battery life.

Beyond ostensible constraints the mobile environment also offers potential advantages such as improved location resolution. While modern web geolocation tools can give a rough approximation of a user's location, in the WorldLens living room environment this location is largely static. With mobile, this information is not only available at a higher resolution with GPS and WiFi positioning data, but it is also dynamic and can provide new opportunities to explore the news in the world around. Perhaps a scenario where a user can simply hold up their phone and see an augmented overlay of the surrounding news and it need not just show local news. Location history may also provide intelligent clues for daily routines, which can also be rolled into features for mobile news delivery.

6.2 Personalization

There are several additional features that could be useful in the WorldLens system in the realm of personalization. Currently, the system does not employ a personalization mechanism. One benefit of not having a personalization system, is that it can help avoid "filter bubbles." Eli Pariser writes in his book "The Filter Bubble: What the Internet is Hiding From You":

> Ultimately, the filter bubble can affect your ability to choose how you want to live ... You may think you're the captain of your own destiny, but personalisation can lead you down a road to a kind of informational determinism in which what you've clicked on in the past determines what you see next - a web history you're doomed to repeat. You can get stuck in a static, ever narrowing version of yourself - an endless you-loop. [25]

That said, there are certain aspects of WorldLens where a selective application of personalization could be useful. For instance, features could be added so that users could customize the visualization based on the textual or temporal extent of content, and the system could learn user habits automatically by monitoring the system usage. Perhaps at eight o'clock in the morning before commuting to work, a user only views news items that are brief and can be finished in a minute or two, versus perhaps the evening when he or she may have more time and prefers to view longer content.

6.3 Meaningfulness Per Unit Time

In the context of timeboxed viewing, another potential area for exploration that captured my interest would be to try and rank content by the useful data conveyed per unit time. For instance, given two videos that are discussing the Ukraine, is it possible to determine which video a user would be better off watching for purposes of efficiently learning about the Ukraine.

Part of the problem will be determining the "entropy" of the video's meaningfulness, which may be relative to the user. In absolute terms, and for any user, if all the video is doing is repeating the same one fact over and over again, then it clearly has low entropy in terms of meaningfulness. A more complicated, but perhaps more interesting approach, would also consider the video's entropy in relation to the data a user already knows. In relative user terms, if all the video is doing is repeating the same facts a user is already aware of, then it clearly has low entropy meaningfulness to the user. It would also be interesting to examine questions of: could I get the same information diet faster from reading newspaper article A versus video B. It would be interesting to examine questions such as: for which types of stories does video consistently trump text (or vice versa), which sources across the various media types provide the most efficient news content, and which provide the least efficient.

From a purely factual point of view, the implementation could begin with a simple approach such as measuring the facts mentioned in a video divided by its length, or the facts contained in an article relative to length of the article.

6.4 Knowledge Mapping

Towards this end, and in the spirit of Vannevar Bush's forward thinking, one concept that has piqued my curiosity is the potential for understanding the knowledge individuals hold. What if we could capture every piece of information individuals encountered? If we could actually understand all the news content a person was aware of, we could specifically present stories targeted to filling gaps in peoples' knowledge of current events or at the very least hide redundant floods of "breaking news". Perhaps today this could be achieved through a combination of audio and video monitoring. Devices such as Google's Glass are potentially well positioned to capture all the information we collect in the course of a day. Perhaps within the century, we will have brain-computer interfaces which will sit at the nexus of the sensory inputs of our eyes and ears. Either way, the ability to understand what we do and do not know (what Donald Rumsfeld termed "known knowns" and "known unknowns" [30]) could fundamentally alter how we understand the world events around us, and relieve some of the mental congestion that results from the deluge of redundant information.

Chapter 7

Conclusion

This thesis illustrates a system for aggregating media across multiple source types, as well as an interface that enables people to explore the world around them. The WorldLens system affords users a rich, visual, interactive portrait. From news articles to video, users can readily discover world happenings across content platforms that are typically isolated. The goal of this work is primarily to enable people to better answer the question "what is happening in the world?" By aggregating content, generating metadata and creating multiple views, the WorldLens system enables users to answer that question in terms of time, geography and topic.

Along the way to achieving the vision for WorldLens, we created the Constellation system which provides an extensible framework for media analysis, and a foundation on top of which several applications. From news aggregation and remixing to exploration, to social video tracking, multiple additional media rich application have been constructed. We anticipate the Constellation system will provide a fruitful blueprint for future work as well as a useful platform for future media applications and experiments.

Appendix A

Constellation Wall Visualization

A.1 Overview

The Constellation Wall was created to provide a sense of the inner workings of the Constellation system. The six foot by eight foot wall was constructed using wood and drywall. The wall provides room to manage power cabling to eight iPads which are mounted on the wall and display content passing through the Constellation framework. Additionally, a short-throw projector mounted on the ceiling augments the experience with graphical overlays on the wall and around the iPads.

A.2 Interaction

By default, the wall initally shows the last 8 videos processed through the system. Additionally, using events from Constellation, the wall dynamically updates as new content passes through the system and new videos are automatically added into the display.

There are two basic forms of interactions with the wall. Users may tap the screen of any of the playing tablets. When this happens, the touched iPad begins to "glow" (via an projected overlay), and the video momentarily pauses as the metadata for the corresponding video shown is gathered. Once all the metadata is gathered and ready, a control signal is sent from the server, and "comets" are overlayed via projection shooting out from the glowing iPad that was touched. These comets represent the metadata gleaned by Constellation, and a list of the metadata templates which are visualized are listed in Section A.4. As soon as a comet contacts an iPad, the display of the iPad is updated to reflect the metadata represented by the comet. Users can tap again to restore the visualization to its original state.

Additionally, using a tablet installated in front of the wall, users may travel through time to explore the processed videos over time. By twisting a virtual dial on the control tablet, users may navigate the time dimension of the videos Constellation has processed. As users move through time, a timeline is projected onto the top of the wall which in addition to serving as a way to jumping to particular points in time, includes information on the number of videos processed.

A.3 System Architecture

The wall visualization system's software was built using HTML, CSS, and JavaScript. Node JS [8] and Socket.io [26] are used to manage communication between the various iPads. A coordinator server was built to handle the querying from the Constellation API. The coordination server manages the various time sensitive aspects of the visualization, such as animation synchronization, as well as the data gathering. Instead of having each client fetch data itself, the coordinator server simply supplies each tablet client application with the data it needs.

A.4 Templates

The visualization system includes a series of visualization templates, which illustrate particular pieces of data that Constellation records about videos. Currently these include templates that visualize the following metadata:

- Subtitles

- Named entities

- Location

- Scene cuts

- Related news articles

A.5 Images of the Exhibit

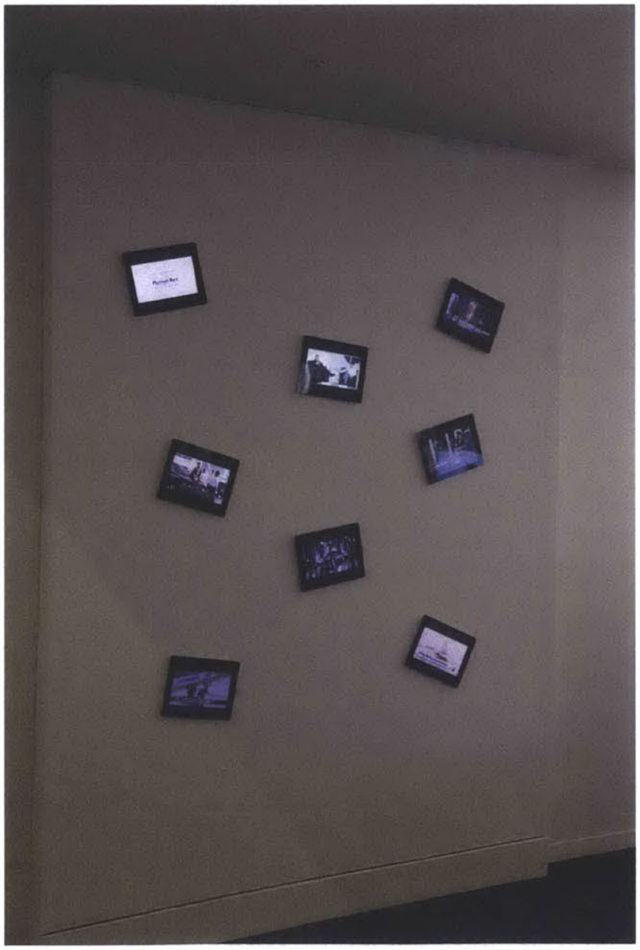


Figure A-1: The Constellation Wall.

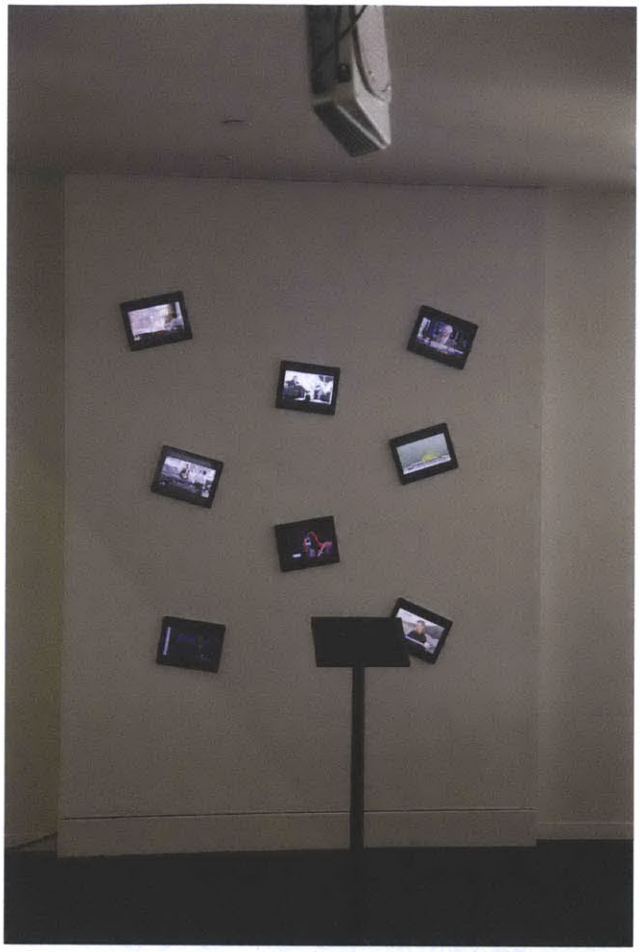


Figure A-2: The Constellation Wall, short-throw projector and kiosk.

Appendix B

User Evaluation

B.1 Survey

B.1.1 Background

How often do you check the news?

Less than once a month	Monthly	Daily	Hourly	Multiple times per hour
□	□	□	□	□

Which facet of a news event is most important to you?

Geography	Topic	Time (Recency)	Other
□	□	□	□

In your daily computer usage do you use 3D interfaces? (excluding video games)

Less than once a month	Monthly	Daily	Hourly	Multiple times per hour
□	□	□	□	□

B.1.2 Tasks

Locate a story and view related content items from at least 3 different sources. Did the related content impact your understanding of the story?

No Impact	Minor Impact	Neutral	Moderate Impact	Strong Impact
□	□	□	□	□

Locate the latest major news story. Were you able to easily accomplish this?

Not at all	Somewhat	Neutral	Moderately	Strongly
□	□	□	□	□

Did the ability to view multiple news sources within the same interface impact your news experience?

No Impact	Minor Impact	Neutral	Moderate Impact	Strong Impact
□	□	□	□	□

Navigate to the topic view. Does this view reflect current trending topics?

Not at all	Somewhat	Neutral	Moderately	Strongly
□	□	□	□	□

View the history of two topics. Did the history view aid in your ability to understand the topic?

Not at all	Somewhat	Neutral	Moderately	Strongly
□	□	□	□	□

View the latest news from any African country. Were you able to easily accomplish this?

Not at all	Somewhat	Neutral	Moderately	Strongly
□	□	□	□	□

View atleast 3 news story clusters with associated video. How do you rate the relevance of video to the news cluster?

Bad	Poor	Neutral	Good	Excellent
□	□	□	□	□

B.1.3 Post Usage Questions

Would you use a 3D interface for news?

No chance	Unlikely	Neutral	Likely	Very Likely
□	□	□	□	□

Would prefer a 2D or 3D interface for your news exploration?

2D	No Preference	3D
□	□	□

Would you be more likely to explore news from other countries?

No chance	Unlikely	Neutral	Likely	Very Likely
□	□	□	□	□

Which view(s), if any, best presented any news stories you were unaware of but found interesting?

Geography	Topic	Time (Recency)	Other
□	□	□	□

Did WorldLens provide you with an understanding of what is happening in the world?

Not at all	Somewhat	Neutral	Moderately	Strongly
□	□	□	□	□

Would you use WorldLens?

Not at all	Somewhat	Neutral	Moderately	Strongly
□	□	□	□	□

Assess the overall usefulness of the tool.

Bad	Poor	Neutral	Good	Excellent
□	□	□	□	□

B.1.4 General Comments

Please include any additional thoughts, comments, suggestions or questions you may have in the space below.

Comments:

Bibliography

[1] Benjamin B Bederson and James D Hollan. Pad++: a zooming graphical interface for exploring alternate interface physics. In *Proceedings of the 7th annual ACM symposium on User interface software and technology*, pages 17–26. ACM, 1994.

[2] Fabrice Bellard. FFmpeg multimedia system, 2005.

[3] Vannevar Bush. As we may think. 1945.

[4] Stuart K Card, George G Robertson, and Jock D Mackinlay. The information visualizer, an information workspace. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 181–186. ACM, 1991.

[5] Pew Research Center. Youtube and news. 2012.

[6] Douglas Crockford. Json, 2001.

[7] Ryan Dahl. Mongodb, 2012.

[8] Ryan Dahl. Node.js: Evented i/o for V8 Javascript, 2012.

[9] Abhinandan S Das, Mayur Datar, Ashutosh Garg, and Shyam Rajaram. Google news personalization: scalable online collaborative filtering. In *Proceedings of the 16th international conference on World Wide Web*, pages 271–280. ACM, 2007.

[10] Mark DiMarco. Datamaps: Customizable svg map visualizations for the web in a single Javascript file using d3.js, 2013.

[11] Willow Garage. Opencv, 1999.

[12] Willow Garage. Cascade classification, 2013.

[13] A Hidayat. Phantomjs: Headless Webkit with Javascript api.

[14] Pablo Hoffman. Scrapy, 2008.

[15] Karen Sparck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21, 1972.

[16] Karen Spärck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 60(5):493–502, 2004.

[17] Karen Sparck Jones. Statistics and retrieval: Past and future. In *ICCTA*, pages 396–405, 2007.

[18] Brenden Kokoszka. Webshot, 2012.

[19] Mark Levene. *An Introduction to Search Engines and Web Navigation.* John Wiley & Sons, 2011.

[20] Hans P Luhn. The automatic creation of literature abstracts. *IBM Journal of research and development*, 2(2):159–165, 1958.

[21] Jock D Mackinlay, George G Robertson, and Stuart K Card. The perspective wall: Detail and context smoothly integrated. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 173–176. ACM, 1991.

[22] Inderjeet Mani and Mark T Maybury. *Advances in automatic text summarization*, volume 293. MIT Press, 1999.

[23] Kathleen R McKeown, Regina Barzilay, David Evans, Vasileios Hatzivassiloglou, Judith L Klavans, Ani Nenkova, Carl Sable, Barry Schiffman, and Sergey Sigelman. Tracking and summarizing news on a daily basis with Columbia's Newsblaster. In *Proceedings of the second international conference on Human Language Technology Research*, pages 280–285. Morgan Kaufmann Publishers Inc., 2002.

[24] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: Bringing order to the web. 1999.

[25] Eli Pariser. *The filter bubble: What the Internet is hiding from you.* Penguin UK, 2011.

[26] G Rauch. Socket.io, 2012.

[27] T Reuters. Opencalais, 2009.

[28] George G Robertson, Stuart K Card, and Jack D Mackinlay. Information visualization using 3d interactive animation. *Communications of the ACM*, 36(4):57–71, 1993.

[29] George G Robertson, Jock D Mackinlay, and Stuart K Card. Cone trees: animated 3d visualizations of hierarchical information. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 189–194. ACM, 1991.

[30] Donald Rumsfeld. *Known and unknown: a memoir.* Penguin, 2011.

[31] Gerard Salton, Anita Wong, and Chung-Shu Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620, 1975.

[32] Pivotal Software. RabbitMQ - Messaging that just works, 2007. [Online; accessed 11-April-2014].

[33] Apache Solr. Project page, 2012.

[34] Flavia Sparacino, Alex Pentland, Glorianna Davenport, Michal Hlavac, and Mary Obelnicki. City of news. *Ars Electronica Festival, Linz, Austria*, pages 8–13, 1997.

[35] Flavia Sparacino, Christopher Richard Wren, Ali Azarbayejani, and Alex Pentland. Browsing 3-d spaces with 3-d vision: body-driven navigation through the Internet city. In *3DPVT*, pages 224–233, 2002.

[36] Marcos Weskamp. Newsmap. *Webdesigning Magazine, June*, 2004.