KALEV H. LEETARU

CAREER AUTOBIOGRAPHY

MIDDLE SCHOOL AND HIGH SCHOOL

As a child, Kalev Leetaru would wake up early on weekends to spend hours building elaborate worlds out of legos and would take apart old electronics to learn how they worked. In eighth grade his parents bought him a robot for Christmas that could be programmed from the computer. He didn't like the interface it came with and so wrote his own programming software. His father, who was just beginning to write web pages, suggested he make a version of it for web authoring, and a few months later in 1995 he founded his first web company, while still in eighth grade. Anticipating the interactive web, one of their early products permitted creating fully interactive web pages, a decade before the Flash-powered online experiences of today. Two years later he sold that first company and moved into digital image protection on the web, where he grew the company to include a vertical reseller network spanning multiple continents. Three years later a venture-backed company readied to purchase the company, but when its funding fell apart from the Dot-Com crash, and with Kalev's entrance to the University of Illinois as a National Merit Finalist, he eventually wound down the company.

UNDERGRADUATE YEARS

While still in highschool, Kalev had begun an unpaid internship at the National Center for Supercomputing Applications (NCSA), the home of Mosaic and the modern web, one of the few highschoolers to have worked there at the time. In fact, his internship brought his career full circle, as his eighth grade science student teacher, who also worked at NCSA, had invited him that summer to work on a project with him. Instead, Kalev's company had taken off, but, four years later, with the company winding down, Kalev was given a second opportunity to work at the Center. As a University of Illinois freshman, Kalev developed the NCSA VIAS project, a global web crawling, analysis, and monitoring service designed to monitor the entire web. Before Google was a household name and "social media" and "industry mining" weren't yet on the horizon, VIAS monitored mailing lists, USENET groups, and the entire web and collected every public piece of information on a given topic, extracting key information like company names and people, and generating detailed analysis and trending reports summarizing entire fields. The VIAS project was used in several corporate intelligence initiatives and received considerable recognition, including the NCSA Private Sector Program Technology Development Award in 2002 and the 2003 NCSA Industrial Grand Challenge Award, which recognizes pioneering solutions to fundamental problems of substantial economic and scientific impact that "enable major competitive breakthroughs" for industry.

As a sophomore he developed an immersive virtual reality application called ShadowLight that allowed users to don a pair of special goggles and controller and use their hands to sketch out their own virtual worlds in space. Showcased at a Chicago gallery opening, it was one of the first virtual reality applications that permitted both freeform and structural design entirely in an immersive environment, with an interface intuitive enough to be used in projects with the US Army, middle school outreach, and the University of Illinois School of Architecture for two and a half years in senior and graduate-level design courses. Still an undergraduate, he was accepted to present at ACM SIGGRAPH and SPIE, two prominent computer graphics conferences.

Kalev also helped lead many of the weekly public tours of NCSA's visualization and virtual reality facilities, and designed and wrote all of the visitor brochures that translated the complex technology of the displays for the general public. In addition to building the scheduling and demo management software that oversaw the facilities, he also took over much of the day-to-day maintenance responsibility of the million dollar CAVE facility for two years, including troubleshooting and maintenance of production audio and projection equipment. He also co-led and then took over NCSA's participation in the annual University of Illinois College of Engineering Open House, the largest university engineering exhibition in the country. Under his leadership, NCSA expanded its participation from a single exhibit to a showcase of all its visualization facilities, becoming one of the most anticipated and popular exhibits at each year's Open House.

Strongly interdisciplinary, he has worked on countless large scale projects with the humanities, arts, social sciences, engineering, business, and hard sciences. Harkening back to his web roots, Kalev recognized that a common need among many of the communities he worked with was the ability for non-technical users to easily create functional and updateable web presences. This led to the creation of the Editable Web Browser™ and Invisibase™ during his junior year, part of the Total Productivity Suite™ that also included online document sharing and collaboration, digital multimedia management, and intelligent website content delivery platforms he developed. The Editable Web Browser built on NCSA's history of developing the first modern web browser by extending the notion of the browser to include an "edit" button. Rather than using separate software to create and edit web pages, a user browsing to a page he or she had editing access to would simply see an edit button light up inside the browser and could instantly change the live page, with an advanced authentication mechanism that didn't require remembering passwords. Invisibase offered the same capabilities for advanced web-based databases, allowing non-technical users to create fully datatyped databases inside enterprise platforms like Oracle with a simple and intuitive web-based platform. The underlying technology was so novel that it resulted in a US patent which has since been cited by four subsequent Microsoft database patents.

Another feature of the Total Productivity Suite was a content management system for websites that allowed web pages and sections of web pages to be assigned metadata that would selectively display them based on the current user's security privileges, browsing history, self-identified interests, and any range of other available metadata. Instead of static documents, web pages became dynamic content containers where any portion of a page could change in realtime based on the user accessing it. The tools were widely used across the University of Illinois campus by both student groups and university units as part of a pilot program and a training workshop was held at the National Science Teacher's Association meeting in Chicago in 2005.

A computer science undergraduate in the fifth-ranked program in the country, Kalev elected to write a senior thesis to fulfill his undergraduate writing requirement and generate new publishable research, rather than take a traditional undergraduate literature course. Under Pulitzer-nominated historian Vernon Burton, he applied his computer, writing, and research skills to create the UIHistories Project, [1] a history of the University of Illinois' entire physical plant, spending his lunch hours over a summer in the University of Illinois Archives writing biographies of more than 300 buildings. To make the histories come alive, he personally scanned more than 30,000 pages of material from the University Library, which he combined with another 50,000 pages of scanned material donated by the Archives. This was used to create an interactive website devoted to the University's history, with a searchable archive of more than 700 works, [2] including all extant campus maps, dedication proceedings, course catalogs, and countless brochures, books, and other materials dating through the University's entire 140-year history.

---

[1] http://uihistoriesproject.chass.illinois.edu/

[2] http://uihistoriesproject.chass.illinois.edu/cgi-bin/rview?REPOSID=8

Complementing the UIHistories' historical view of campus, he personally took more than 80,000 photographs documenting every building and space on the University's campus through the four seasons across more than six years, despite having never owned a camera before starting the project. Once again, the resulting image database was cataloged and made freely available via a website, called UIPhotos/Phantasm, [3] with more than 22,000 images licensed for publications, presentations, and displays over the last five years by every unit within the University and its alumni, at a total combined gift-in-kind value to the University of $8.09 million: nearly $1.6 million dollars per year.

In March 2005, roughly half of the nineteenth century plaster models from the University's Math Models Collection (170 in all) were removed from their museum cases during a building renovation and Kalev took the opportunity to professionally stage and photograph each model, capturing more than 2,300 images. The resulting gallery is available online, [4] together with scans of several early catalogs and descriptions of the models. The UIHistories and UIPhotos/Phantasm projects are one of the largest university-focused digital history and photographic archives in the world, and among the largest digital history projects created by a single individual. During the process of cataloging and publishing the more than 80,000 images in the collection, Kalev discovered that contemporary image management platforms couldn't scale to collections of that size and complexity. So, he developed a brand-new software platform that could handle the complex workflow of the site, including a completely automated electronic licensing system that handles all aspects of image licensing and rights management.

His first year on campus, Kalev was named University of Illinois Student Employee of the Year, the first freshman ever even considered for the honor. His sophomore year, he was featured in Fortune Magazine in a profile of national university student technology leaders. Kalev's first biographical profile came as a senior with a 5-page portrait in the Kauffman Thoughtbook as their featured Student-Entrepreneur. [5] Over the course of his undergraduate career, his student research led to three United States patents, placing him in the top 102 most prolific patentees in the University's history, the top 16 in the UI College of Liberal Arts and Sciences, and number one in Computer Science and the Social Sciences. Furthermore, his more than 50 University Invention Disclosures place him among the University's most prolific inventors ever in terms of disclosed innovations.

GRADUATE SCHOOL

Given his strong roots and ongoing research at the University of Illinois campus and his focus on the *application* of information technology to real-world problems as opposed to the *theory* of computing, Kalev applied to the doctoral program at the University of Illinois Graduate School of Library and Information Science (GSLIS). He was one of the few candidates accepted into the #1 ranked program in the country without an intermediate Masters degree. Continuing at NCSA, now as a Graduate Research Assistant, Kalev worked on a number of projects relating to large-scale data analysis and web mining.

One such project, part of his early doctoral research, focused on the ways in which institutions could compile all of their disparate data warehouses and external datasets and divine new meaning and insight into the operation of their institutions. Called the Profile of a Campus Project, [6] the initiative presented its findings in 2006, at the start of the University's Strategic Planning process, with six core

[3] http://uihistoriesproject.chass.illinois.edu/photoarchive/
[4] http://www.mathmodels.illinois.edu/
[5] http://www.kalevleetaru.com/Publish/KauffmanThoughtbook2005_LeetaruBio.pdf
[6] http://www.kalevleetaru.com/profileofacampus/

focus areas covering the breadth of available data at the University of Illinois. Insights on Units drew from the University's human resources database and compiled all departmental appointments listed for faculty/staff on campus. Personnel with multiple appointments were used to connect their buildings and departments based on shared research areas and the resulting "campus collaborative network" showcased the ability of a dataset created for one purpose (payroll) to shed light on very different attributes of the University's structure (collaboration). Focus on Engineering drew from an annual reporting database compiled by the College of Engineering, searching through major citation indexes to list all publications and presentations given by engineering faculty/staff each year. This was used to explore collaborations and shared research themes among engineering departments: once again applying a database created for one specific purpose (annual reporting) to a very different cause (measuring interdisciplinary research collaborations).

Emphasis on Education used Kalev's national 40-year graduation dataset to rank the University through time among its various peer groups in each field it awarded degrees. Quick Graphs showed how visualization techniques could be used to display key indicators in a dashboard-like format for administrators, while Inside Campus used a complete dump of the University's key performance metrics data warehouse to chart all 244 departments across 440 indicators with 9 years of reporting data. It focused on computing underlying trends and using them to rank departments, showing whether they were improving on each indicator. Finally, the Campus Concept Network was a unique take on how an institution's web presence could be used to better understand itself. This involved downloading the 100 most popular pages from all 272 of the University's departmental websites. Computer text mining techniques were used to identify all person and concept mentions on each page, resulting in a semantic network of 109,253 unique entries with 637,971 connections between them. This was used to understand which departments promoted the same research themes on their websites (indicating potential collaborative opportunities) as well as to help units understand the primary images they conveyed through their web sites.

When the National Archives and Records Administration (NARA) approached NCSA about working on a series of pilot projects relating to its future Electronic Records Archives (ERA) platform, Kalev was technical lead on several components, including hardware platform benchmarking and large-scale data visualization. One question NARA posed was whether the document archives of the future still needed to run on multimillion dollar supercomputing systems, or whether clusters of cheap PCs could do the job just as well. A wide range of technical benchmarks were conducted across NCSA's entire range of hardware platforms, representing most storage systems available at the time. NARA was also interested in how it could better facilitate interaction with the large datasets it collects. The public Enron email collection, totaling more than a half-million emails sent within Enron before its collapse, was used as one of the selected databases, with a variety of display and interaction tools developed to support querying it. One such tool used NCSA's tiled display wall to interactively render the entire network of emails at full resolution on a wall-sized display comprised of 40 individual projection screens seamlessly joined together.

When Hurricane Katrina struck in 2005, Kalev worked with Dr. Robert Gillio (one of the senior physicians who developed the environmental monitoring regimes for first responders after the 9/11 attacks) to address the need for a comprehensive, scalable, HIPAA-compliant medical records and tracking system for the Medical Reserves Corps deployed under the Office of the United States Surgeon General. Though the pilot system he developed for the project was never deployed in production, it served as a critical prototyping mechanism to understand the architecture and design needs of future EHR/EMR medical records systems and made recommendations to the CCR medical record standard.

Working with the Office of Naval Research, Kalev was the chief architect and developer of the ONR's "federal funding search, discovery, and analysis system." This web portal compiled a daily list

of all grant and contracting opportunities from across the entire federal government, performed a variety of data and trend mining analyses, and offered advanced trend mining, sophisticated visualizations and pattern and spatial analysis in a single integrated "funding opportunities portal" to support national small business and entrepreneurship addressing federal needs.

On behalf of the Ewing Marion Kauffman Foundation in 2006, Kalev led the technology and data management side of an ambitious project to catalog every faculty member at a US university who had recently taught an entrepreneurship-related course. Such courses need not be taught in a business school and may not have "entrepreneurship" in their title (such as engineering courses on forming startups or art courses that focus on writing grants for new works). This involved constructing a massive data integration platform that could accept data streams from automated crawlers, human searchers, data analysts, and precompiled reports. A custom-designed interface adapted traditional CRM methodology to a large-team parallel workflow, with a specific focus on conflicting information sources. The system included a suite of advanced "fuzzy" matching tools to track faculty that had moved between institutions over time, changed names, or had typographical errors in their information, linking duplicate and updated information. This system was paired with a similarly significant effort to compile summary reports of all accredited university websites in the United States and prepare search documents for trained human searchers to track down peripheral and nuanced references.

Another project with the Ewing Marion Kauffman Foundation involved the construction of a master database that permitted the tracking of every single postsecondary degree awarded by any institution in the United States from 1966-present. Initially created to allow the evolution of the US business degree to be tracked over four decades, the final database includes all degree types. Unlike NSF WebCaspar or similar resources, which track only at the line-group level (which groups administrative support positions with MBA degrees, for example), the resulting database includes a record of every college and university degree issued over nearly 45 years. This involved a massive data integration effort, including obtaining file conversions of 30-year-old magnetic tape and scanned copies of printed documentation and manually reconstructing legacy data formats, which had changed almost yearly. All available crosswalk and translation tables had to be evaluated and integrated to permit tracing of single degree lines across six different taxonomies. The only database of its kind in the world, the resulting archive was most recently used in a study examining the last 40 years of undergraduate women engineering students, tracing the unique trajectories of each degree, and documenting new institutional impacts on degree conferral. The forthcoming study, to be published in the Journal of Women and Minorities in Science and Engineering, relied on a custom large-computing statistical environment Kalev developed, which extracted every crosswise relationship from the full dataset, utilizing a series of automated data mining processes to uncover a set of novel findings regarding the role institutional type plays in the engineering environment for women.

Named a University of Illinois Graduate Scholar in Entrepreneurship in 2009 for his research on examining national US news coverage as a proxy for industry intelligence gathering, he also served as Project Manager for the University's RiverWeb Project. Under his leadership, RiverWeb was modernized into to a rich interactive online experience, with a searchable digital library integrated alongside thematic narratives. He also launched and oversaw RiverWeb's GIS mapping initiative, which assembled all available commercial and governmental GIS resources for the Illinois city of East St. Louis and produced more than 500 maps and analytical works exploring aspects of the city's evolution over the last 100 years.

Kalev has worked extensively on projects and initiatives to help promote the University of Illinois. Most recently he led the creation of the forthcoming "Student Innovation at Illinois" portal, a collaboration with the University's Technology Entrepreneur Center to catalog and promote student

research at the University. Currently in pilot mode, the site already contains the complete exhibition lists of the last several Engineering Open Houses, which are normally removed from the web each year to make room for the following year's exhibit list. Once released in 2011, senior design courses, competitions, and student submissions will all be integrated into the database. Ultimately, the site's goal is to offer profiles of nearly every student-led research project at the University over time, providing a unique window into the incredible ingenuity of University of Illinois students, which is not preserved or promoted through any other venue. In conjunction with this project, Kalev is also leading an initiative with the Engineering Open House committee to digitize all 100 years of its exhibition guides and internal documents and create an online interactive exhibition tracing a century of student engineering projects on campus. In 2005, 2006, and 2010 he also visited all of the Open House exhibits across the entire university and captured at least one photograph of each exhibit, resulting in a visual time capsule numbering several thousand images.

Throughout his graduate coursework, Kalev has tried to leverage his course assignments into publications or larger research initiatives. His very first course paper as a graduate student, on the rise and evolution of instant messaging, was turned into a book chapter one year later. [7] All four courses his second year of graduate school resulted in publications or works currently under preparation for publication. One paper was published in the Journal of International Communication, [8] while two others are in preparation for journal submission. His forth course, an independent study, resulted in a 189-page book manuscript surveying the field of digitization, which is currently under review. The final chapter of that manuscript was the first quantitative technical comparison of Google Books and the Open Content Alliance and was published in First Monday, [9] leading to an interview in Que Leer, the Spanish cultural magazine, titled "What will the world of the library look like in 2020: Six experts predict the future." [10]

Kalev's third year of graduate school resulted in a second 147-page book manuscript: an introduction to the field of automated content analysis, currently under final review by Routledge. A second course paper applied automated tone analysis to the CIA's unclassified global news monitoring archive to trace out worldwide reaction before, during, and after the 2003 Iraq invasion. The opening chapter of that paper was the first unclassified longitudinal analysis of the CIA FBIS and British Intelligence SWB open source intelligence news monitoring archives, exploring questions such as topical and source biases and source/sink geographic diffusion patterns. The CIA ultimately published the analysis in the unclassified March 2010 issue of its scholarly journal, Studies in Intelligence. [11] A study derived from the paper, examining the personalization of the Iraq conflict in the world's media, will be presented at the forthcoming International Studies Association Annual Convention [12]

His fourth year of graduate school resulted in one journal publication and two more papers under preparation for publication submission. One of the papers under preparation is a 187-page manuscript applying computerized content analysis tools to the complete population of 18 million

---

[7] *Instant Messaging as a Hypermedium in the Making*, in Handbook of Research on Computer Mediated Communication. (2008). Editors: Kelsey, Sigrid & St.Amant, Kirk. IGI, Inc.

[8] *An Open Source Study of International Media Coverage of the WorldCom Scandal*. (2008). Journal of International Communication Vol 14, Issue 2.

[9] *Mass book digitization: The deeper story of Google Books and the Open Content Alliance*. First Monday. Vol. 13, Issue 10. (October 6, 2008).

[10] http://www.kalevleetaru.com/queleerinterview.html

[11] *The Scope of FBIS and BBC Open Source Media Coverage, 1979-2008*. (Unclassified). CIA Studies in Intelligence, Vol. 54, No. 1. (March 2010). Pp. 51-71.

[12] *Do We Have a Stake in This War? A WorldwideTest of the In-Group Out-Group Hypothesis Using Open-Source Intelligence* with Scott Althaus. International Studies Association Annual Convention 2011: Global Governance: Political Authority in Transition. Montreal, Canada. March 16-19, 2011.

New York Times articles 1946-2005 to characterize the paper's coverage of war through the last half-century. The second paper used a similar approach to contrast the New York Times' and Wall Street Journal's coverage of presidential campaigns over this period, exploring the issue of bias in political coverage. Finally, a six-year archive of the Drudge Report totaling nearly 1.6 million snapshots was used to examine the role of social media in the mainstream news cycle and develop a case model for characterizing news websites. The headlining issue of the July 2009 issue of First Monday, [13] the study has twice been cited in Columbia Journalism Review and an investigative report by journalist Greg Beato using the study's data led to an editorial piece by the CRJ Editors on the ethical considerations of link aggregators. [14]

Just as his very first graduate course paper was published, so too did his final graduate paper go on to have a public impact. Kalev's last course was an independent study with Dr. Paul Magelli examining the last half-century of New York Times coverage of higher education. Titled the "Soundbite University," the study examined more than 18 million documents, comprising the entire run of the New York Times from 1945 to 2005, for all references to United States research universities and using spatial, temporal, and institutional indicators to examine how coverage has changed over this period and the characteristics most commonly associated with elevated national press visibility. One of the most surprising findings was the transition of the research university from a newsmaker to a news commentator, suggesting a need for universities to profoundly alter the ways in which they interact with the press. The study was profiled in the Fall 2010 issue of the American Council on Education (ACE)'s The Presidency. [15]

Collaborating again with his father, a researcher in carbon capture and sequestration (CCS) at the Illinois State Geological Survey, they presented their first coauthored study on the state of global media coverage of climate change and CCS in 2006. [16] One component of this study analyzed the ownership of the websites of the top 100 Google search results for CCS, examining potential biases in the information presented to the public. A second collaboration traced the last half-century of national news coverage of climate change, characterizing the evolution of energy in the public consciousness and the interconnection of coverage with external factors such as federal funding priorities. Building on this work, Kalev ultimately launched the Carbon Capture Report [17] in March 2009 to monitor global news and social media coverage of climate change and the energy sector on an ongoing basis. A total of 15 topic areas are now monitored by the site, including Climate Change, Carbon Capture, Carbon Credits, Alternative Energy, Renewable Energy, Green Energy, Biofuels, Geothermal, Hydroelectric, Natural Gas, Nuclear, Solar, Wind, Coal, and Oil. Algorithms determine the geographic location of coverage, cluster duplicate coverage, measure the overlap between blog and news content, compute emerging topics, and crowdsource and rank mainstream news by its popularity in the social media sphere. Automated Biographical Databases extract the names of people and organizations in the news and compile timelines over time of the relationships between the actors in the CCS space. Filters offer categorized access to coverage of activism, documentaries, legislation, and projects.

Over the past seventeen months the site has grown to become the premiere source of global insight on climate change and the energy industry, totaling more than 14.3M hits in April 2010. With

---

[13] New media vs old media: A portrait of the Drudge Report 2002-2008. First Monday. Vol. 14, Issue 7. (July 6, 2009). Cover article of issue.
[14] *The Link Economy*. (2009, November 10). The CJR Editors.
http://www.cjr.org/news_meeting/the_link_economy.php
[15] *The Presidency*. (September 2010). American Council on Education.
[16] *The use of data mining methods to evaluate public interest in carbon sequestration*. Fifth Annual Conference on Carbon Capture and Sequestration, Alexandria Mark Center, Alexandria, Virginia, May 8-11, 2006.
[17] http://www.carboncapturereport.org/

subscribers in more than 80 countries, the site has been ranked in the top 60,000 web sites globally, was listed in the top 10 results for nearly a quarter-million search terms on Google in April and is in the top 20 most-tweeted sites in the world on CCS. Among its subscribers today are most large energy, environmental, policy, governmental, and environmental financial and legal services firms in the world.

A completely new generation of the site was released in October 2010 which continues to build the site towards an intelligence and research platform, expanding the automated biographical database to include news outlets, blogs, twitter accounts, and locations. It is now possible to select a specific city or geographic region and see all coverage in a specific industry referencing that location through time, the pattern of tone of that coverage, and the top outlets covering that location, among other attributes. Individual news outlet profiles offer a complete overview of that outlet's coverage of each industry (for example, the New York Time's coverage of CCS), along with its tone and geographic focus. A specialized tonal analysis module offers unique features like intensity of "language personalization," a key indicator of popular mobilization, and the person database now recognizes names transliterated from more than 7,500 languages. A completely redesigned interface incorporating interactive visualizations makes it much easier to interact with the tremendous amount of analytical findings presented by the system each day. Finally, a new ranking system monitors the entire network of information flows across all mainstream and social media outlets covering each industry and calculates the "resonance" (importance) of each outlet to that community.

CLINE CENTER FOR DEMOCRACY

In 2006, while continuing his graduate work, Kalev joined the University's new Cline Center for Democracy as full-time academic staff with a mandate to help position it at the forefront of computational social science. The Center's two signature projects are its Comparative Constitutions Project (CCP) and its Social Political Economic Event Database (SPEED). The CCP project aims to collect all written constitutions worldwide 1789-present and codify them under a 1,200-question protocol. Under Kalev's technological leadership, the project transitioned from a largely manual workflow to a centralized data management infrastructure. Digitization was revamped using a high-speed dedicated document scanner and high-throughput enterprise digitization workflow. A specialized web portal system was developed to centralize all documents into a single interface supporting provenance tracking, advanced metadata capabilities, and fulltext searching across the entire collection. A new web-based protocol system capable of scaling to large numbers of users was developed with realtime progress reporting and built-in support for multiple coder reconciliation. Finally, a new message board system was created with a range of "awareness" indicators allowing faculty to analyze and visualize the communicative patterns across their entire team, proactively identifying problems before they impact coding progress. The resulting system's progress indicators have proved sufficiently comprehensive that the team has been able to eliminate their weekly all-staff meetings.

The SPEED project is an effort to compile a global event database codifying every major social, political, and economic event in the world from 1946 to present. Textual news reports are transformed into codified database entries with date, latitude/longitude location, and more than 1,600 variables, including connections to related events. This required building a global news monitoring infrastructure from scratch capable of capturing news from every country of the world and constructing a workflow that leverages the efficiency and scalability of automated text mining in concert with the interpretive sophistication of trained human analysts. From an initial pilot project clipping articles out of the paper New York Times with scissors, Kalev transitioned the project into a turnkey digital infrastructure, developing tools such as a document categorization system capable of filtering out irrelevant articles from a collection of more than 70M documents at greater than 98%

accuracy. Automated reliability checks, queuing systems, historical geocoders, and flexible lexicon and XML protocol management tools are all available from a single web-based portal capable of handling more than 30 students working simultaneously across an archive of more than 70M documents. Today, the SPEED project is one of the largest open source intelligence initiatives in academia, and is home to the only complete archives of the declassified CIA Foreign Broadcast Information Service and British Intelligence's Summary of World Broadcast collections.

Launched in the leadup to World War II, FBIS and SWB have operated continuously for nearly 70 years, monitoring the radio stations, television stations, newspapers, magazines, trade journals, and all other "white" and "grey" literature in nearly every country of the world. A representative sample of the daily news in each country is translated into English and recorded in a permanent archive in Washington, DC. More than 50 years of this content exists only in microfilm and microfiche format, requiring the digitization of several million pages of material to integrate it into an automated workflow. When commercial scanning firms proved too expensive and unable to meet the Cline Center's archival needs, Kalev designed and launched one of the highest-volume university-based microform digitization facilities in the country, turning out in excess of one million scanned pages of material per month. Commercial microform scanners are not designed to operate 24 hours a day, requiring down time each evening for recalibration. Operating the scanners at maximum archival resolution 24/7 for more than three months required collaboration with the manufacturer to adjust their software and workflows to accommodate this continual operation. Additionally, when budgetary constraints prevented the purchase of a $250,000 enterprise digitization platform, Kalev developed his own workflow system from scratch leveraging the Center's existing infrastructure investment.

In 2008, when one of the Center's faculty members commented that a citation to a White House press release now led to a non-existent URL, Kalev leveraged his experience using the Internet Archive for historical web research to track down the history of the page's changes. Collaborating together, they produced a study called "Airbrushing History" [18] that documented how the official public record of the Coalition of the Willing that backed the US Invasion of Iraq in 2003 had been deleted, altered, and postdated over a period of more than two years. In several cases, countries were added or removed from dated press releases while keeping the date unchanged, giving the impression that those countries had always (or never) been a member of the coalition on that date. The resulting study was featured in the New York Times [19] and received international coverage, including a commentary by the New York Times Editorial Board. [20]

As a personal initiative, Kalev worked with the Society of Women Engineers for six years, first becoming involved as an undergraduate, assisting numerous initiatives, and serving as their Staff Advisor for his final year. He co-led two signature initiatives with their section president, one addressing the gender diversity climate in the University of Illinois College of Engineering and one tracing the society's history, both leading to publications in the national society's professional magazine. [21] [22] Under the leadership of UIUC Section President and Diversity Lead Jessica Wood, SWE launched an ambitious diversity initiative beginning in 2007 that involved two email surveys of the entire female engineering student population at the University, outreach to student organizations,

---

[18] http://www.clinecenter.illinois.edu/research/airbrushing_history/

[19] http://www.nytimes.com/2008/11/25/washington/25documents.html

[20] http://theboard.blogs.nytimes.com/2008/11/25/president-bushs-coalition-of-the-willing-or-orwell-comes-to-iraq/

[21] Diversity from the Ground Up. SWE: Magazine of the Society of Women Engineers. Society of Women Engineers. Conference 2008 issue, pp. 40-46.

[22] Your Collegiate Section History is Waiting to be Discovered. SWE: Magazine of the Society of Women Engineers. Society of Women Engineers. Fall 2008 issue, pp. 62-64. The history initiative this article was based on was a winner of the SWE National 2010 History Award.

including meeting with all women-serving engineering student groups, and discussions with faculty and sponsors across the campus and College. Using a Six Sigma approach, the resulting 37-page analytical report [23] [24] drew from interviews, surveys, and campus and US Department of Education data to paint a full portrait of the current state of the gender diversity climate with the College of Engineering. This report and a subsequent recommendations proposal [25] was widely distributed throughout the College administration, helping to drive new initiatives aimed at improving the environment for female engineering students.

The history initiative researched the founding and evolution of the University of Illinois section of SWE, uncovering its place as the fourth collegiate chapter of the national society. It was discovered that 2010, the 60th anniversary of the national society, was also Illinois SWE's half-century anniversary and this became the theme of the chapter's winning bid to host the 2010 SWE regional conference at Illinois. A history section was ultimately added to the UIUC SWE website, with narratives and scanned historical documents tracing its 50 years of evolution on campus, and was honored as the only student chapter to be recognized by the 2009 National SWE History Award. Kalev's more than half-decade of service to the society was recognized with the University of Illinois Society of Women Engineers Award in 2008.

More recently, his work has focused on international media and how the news can be used to understand the broader world. He lectures on corporate intelligence and social media, including both its business and intelligence uses and has presented to Fortune 100 companies and professional organizations such as NACRO (Network of Academic Corporate Relations Officers). He was an invited participant at the 2007 and 2008 Director of National Intelligence Open Source Conferences, among the select few academics invited to attend. He has also been an invited speaker, panelist, and discussant at numerous scholarly venues including Harvard, Columbia, Stanford, UC Berkeley, and the Library of Congress, and an invited coauthored paper was presented at the 250th Anniversary Conference of The Royal Norwegian Society of Sciences and Letters (DKNVS). His work has been profiled in venues as diverse as the New York Times, Columbia Journalism Review, MSNBC, Que Leer, US News & World Report, Politico, Library Quarterly, and the American Council on Education's The Presidency.

In November 2010, the Library of Congress flew Kalev to Washington for a two-day summit on the future of citizen journalism and digital news, where he was one of just 25 external advisors interacting with 20 key Library of Congress staff. Kalev was one of the keynote presenters on the current state of digital news and research, where he presented a global perspective to the issue of citizen journalism and news analysis.

Today, Kalev has more than twenty global projects underway on news flows and public perception. One involves modeling all global news interactions across all countries in the world over the last several decades, resulting in tens of trillions of connections. Using some of the most advanced computation techniques and resources available, such work aims to significantly enhance the understanding of both global news flows and societal interactions.

[23] http://www.kalevleetaru.com/Publish/SWE_Summary_Of_Diversity_Status.pdf
[24] http://www.kalevleetaru.com/Publish/SWE_Dean_Letter.pdf
[25] http://www.kalevleetaru.com/Publish/SWE_Final_Proposal.pdf