Inventory Building and Utilization with Information Technologies:
The Use of Database and Pattern Matching Systems in Teaching and Research

Dr. Joshua Franzel and Dr. James Lee
American University
Washington DC, USA

Inventory Building and Utilization with Information Technologies:
The Use of Database and Pattern Matching Systems in Teaching and Research

Dr. Joshua Franzel and Dr. James Lee
American University, Washington DC, USA

Abstract

This article discusses the life cycle of two innovative information technology products that have both a teaching and research application. The dimensions of the story include the history, technologies, development, user base, and future plans for these information technology tools that have elements including online databases to pattern matching system software. These two related projects focused upon here, organize and offer hundreds of case studies on trade, environmental, and conflict issues for use by academicians and practitioners alike. The construct and application of the pattern matching systems and their underlying databases provide a conceptual foundation that may be useful in a wide range of disciplines, ranging from Public Administration to International Trade.

As web development and web use technologies have experienced greater adoption by universities and colleges in almost all aspects of operations, so too have educators incorporated them into their teaching and research. The Mandala Projects are a linked system of online research projects at American University. Two of its applications, those using database and pattern matching systems, are the focus of this article. These two projects exist as online journals, but are constructed in a manner that includes not only textual presentation, but also coded categorical case studies. The two projects are the Trade and Environment Database (TED) and the Inventory of Conflict and Environment Database (ICE). Since 1991, the TED and ICE projects have produced over 900 case study projects (around 800 from TED and 150 from ICE) that are available on the web. The cases studies are researched and programmed by graduate students in the School of International Service and are now used by academicians and practitioners around the world as reference and teaching tools. Typing the names of the projects into Google.com search engine reveals 400 sites that link to TED and 250 to ICE.

The article reviews the background of the TED and ICE databases and search engines, problems encountered during their development and successful solutions, current uses and user base, future plans for the two projects, and an in-depth outline of applied web technologies. It is also a more general exemplar of research that identifies sets of related case studies of interest, highlights trends in cases, and organizes large amounts of case data. Hopefully from this discussion a conceptual construct will emerge that others can adopt, modify, and improve in their own teaching and research.

Key words: innovative research tool, pattern matching, trade, environment, conflict.

Background

The Trade and Environment Database (TED) project began in 1991 under the direction of Dr. James R. Lee of the School of International Service at American University (AU) in Washington D.C. It was an outgrowth of a new class he continues to teach, "Computer Applications in International Relations Research". Within a short time of the class's inception, he and the students were developing some of the first generation of web pages at the educational institution by using HTML for programming. During the semester, students carry out case studies and publish them in the form of web sites, on issues of global trade and its relation to culture, human rights, labor rights, and other non-economic issues. Early case studies focused on such topics as Thai cigarette import bans, Ecuadorian oil export disputes, and hazardous Italian trade exports to Lebanon. The diversity of topics grew as the number of cases and trade itself grew over time. As the number of cases increased, so did the need for their organization and relational coherency. This lead to the creation of the TED database and a search package, using categorical case studies, that allows users to query the database on the basis of coded attributes, which grew from forty-two in 1992 to over 800 in 2005.

In 1997, as the number of TED cases grew, it became evident that a substantial portion of case studies dealt specifically with environment and local/regional disagreements that included physical violence or the threat of it. This led to the creation of an offshoot collection of studies entitled the Inventory of Conflict and Environment (ICE). Topics covered in this group include the environmental and economic repercussions of the Persian Gulf War on Kuwait, herbicide use in the Vietnam War, and Canadian and Spanish fishing disputes.

In 2001 Dr. Lee and Rick Hink, at the time a graduate student in International Studies, built a search engine for TED cases that searched on free text in case abstracts and on the coded categories (using Macromedia Cold Fusion with an Microsoft Access database as a backbone). In 2002, the two built a pattern matching system that permitted the input of coded attributes for a new or prospective case and matched the case to prior cases so that a ranking of similarity was constructed. It was also possible to weigh the ranking system. These similarities are thought to be at minimum a set of "lessons learned" for the decision maker or researcher, both in policies to pursue and to avoid. In 2003, Dr. Lee and Joshua Franzel, at the time a doctoral student in Public Administration and Policy, using TED as a model, built a similar but improved version of the sort and pattern matching system for ICE to allow users to sift through over 150 related cases.

Web Technology Used

The Trade and Environment Database and Inventory of Conflict and Environment use the same technologies. When students complete the web sites for their projects, which are typically served from their personal network drives during project development, they are asked to provide all site files (.html, images, sound/video clips, and associated documents) on a compact disk or zip disk thus allowing their sites to be transferred to the TED or ICE servers by the instructor or teaching assistants.

Along with a textual report, students provide categorical coded information related to 28 categories in TED and 16 in ICE. For example, cases might be coded as to the relevant continent, the regional focus of the study (South Asia, Eastern South America, etc.), the case's habitat (temperate, tropical, etc.), the case's industrial focus (mining, oil, tourism, etc.), applicable environmental issues (health, sea pollution, human rights, etc.), scope of issue (sub-state/NGO, unilateral, bilateral, regional, etc.), applicable policy issue (export/import ban, quotas, intellectual property, etc.), time period (ancient, middle, modern), duration of case (10, 100, 1000 years), the trigger of the issue (access to water, political disagreement, terraforms, etc.), outcome of issue (yield, victory, in progress, etc.), and level of issue/conflict (threat, high, medium, low, harm).

The coded information is then entered into one of two Microsoft Access databases, one for TED and one for ICE, along with the URLs pointing to each case on the TED or ICE server. From the databases, the search engines and the pattern matching systems are built. Each project has a search engine that can either use inductive or deductive approaches. Both reference the project's respective Access database, but process the information differently.

The "deductive" search engine format allows the user to carry out two different types of searches. The first type is a basic keyword search in which the user can type one to three terms into the provided entry fields and the search engine queries abstracts in the respective database. The Cold Fusion output provides the user with all applicable cases (with title, abstract, author's name and URL), that match the requested keywords in any of their categories, abstract, or title discussed previously. Also the user is provided the number count of cases in the database that match their requested keywords and links to all applicable online reports. This is an "open" search system.

The search engine also has a section with drop-down menus for searching each category with associated choice options, as discussed previously. The user can choose to specify search preferences for each category or just a few, leaving the other categories unspecified (it is an "intensive" rather than an "extensive" algorithm). The engine, as it did for the keyword function, will then offer the cases (with title, abstract, author's name and URL) that have Access category attributes that match the specified user selections. This is a "closed" search system

The "inferential" search engines for both TED and ICE offer a pattern matching system and an additional weighting component for the user. Similar to the drop-down menu selection approach above, the user is offered categories with associated choice options. In addition to choosing an option for each category, the user is also given the option of weighting (1-10, default is 1) the importance of each category in the query for cases. Thus, a decision maker might want to give a preference to African cases, for example, because the current issue at hand is in Africa. Cases are then ranked in order of a ratio score, highest to lowest. The ratio score is derived from the sum of the weights of all retrieved matches with the user's preferences divided by all of the possible weights from all of the categories added together. This ratio score, it is believed, yields a measure of how important the case ranking should be considered with respect to decisions in a new or prospective case, given the user's preferences.

Developmental Problems and Solutions

As is the case with any project, the TED and ICE development team had to address several issues during and after the database and search engine creation steps. Three specific problems surfaced which may be very important when one is attempting to create and implement such research and to develop the requisite coding, database, and pattern matching tools. Some major problems are (1) underestimating production time, (2) assuring student coding reliability, and (3) protecting of intellectual property rights in an Internet arena and preventing unauthorized use of materials.

Even at the graduate level, students in non-Information Technology fields often do not have sufficient training on web technologies. While they can be taught to "hand" code HTML with applications like Microsoft Notepad/Wordpad or to develop sites with programs such as Macromedia Dreamweaver, their projects still may not be ready for web posting due to format and compatibility issues. Even if taught proper linking techniques (i.e., relative pathnames), students often link to pages, images, and files with site-specific, absolute pathnames (Lemay, 2000, p. 99). This has and does cause sites to fail when the sites are transferred to another folder or server; a process that occurs when sites are moved to the TED or ICE servers from the students' individual development network drives. The development team assigned to preparing the students' sites for usage in the databases created two solutions approaches. First, they keep a running list of common mistakes students make from semester to semester and inform the instructor to reiterate solutions to these mistakes at the onset of the course when the initial web development and site definitions occur. Also, the team found that, no matter how much training is given to students, mistakes will still happen and time will be needed for correction. The team has opted to allot a month after the semester is over to review the content and format of the sites before the semester's batch of new cases is entered into the databases. Thus, there was recognition of the need for postproduction of the cases.

A second important problem addressed during development and implementation of the models deals with student coding reliability. As noted, students provide categorical information in regards to the continent, region, habitat, industrial, environmental issues, scope, etc. of their cases for the databases. While they are prompted to provide information on a set number of categories with a set number of potential values, there is still a large amount of potential variation between how one student categorizes a case in relation to another student with a similar case. Without such inter-reliability the overall value of the full set of cases is diminished and thus its usefulness reduced.

To combat such coder reliability issues, the ICE and TED teams have incorporated three methods. The first is having the instructor use a few students' projects as examples of how one should categorize them. This gives the students a reference of how different cases fit into various categories. Second, along with the examples offered in class, students are given documentation for reference that outlines categorization standards for each of the database / search engine models and also provides some examples. The third solution minimizes categorization variation among student case studies by reviewing the cases as they are transferred from the students' network drives to the database servers.

A third major problem of note deals with the improper use of ICE and TED case material. In fact, the volume of these case studies has grown over time and the site is visited by millions each year. Equivalent privately run web resources would ordinarily ask for some sort of fee to pay for such services. In this respect, it can be said that the project has a value and thereby a potential for abuse. The development team typically will allow anyone to use the information, images, and documents offered in both databases' case studies as long as those using the material use proper reference citation (citing ICE and/or TED). No commercial use is allowed. However, groups and individuals will often take ideas and the actual files from the case studies and use them as their own. One case study from TED wound up on the web site of "cheaters.com", a site name that needs little embellishment. To protect the rights of the students, the instructor, and of the university, the development team uses online search engines like Google.com, Yahoo.com, and Altavista.com to search and monitor TED and ICE information being used by others. Also, an informal network of the development team, previous students, persons who have used the information properly, and others who are interested in the databases and engines in general, has informed and often identifies, through normal web surfing, TED and ICE information that is being used improperly.

While instances of abuse are rare, a basic procedure is in place as to how to approach such issues, with a general escalation of responsibility. Once a person or group is identified as using case study material improperly, the development team will typically send an email or call the party asking them to either correctly cite the information and/or take the information off of their web site. If this first step does not work, an email will be sent to the party's network administrator. These two steps usually are effective in correcting the problem of misrepresented material, but should they not work, referring the issue to American University's general counsel is an infrequent, but available, third step.

Current Uses and User Base

In 2005, about 4 million people visited the Mandala Projects' web site that includes the Trade and Environment Database and Inventory of Conflict and Environment. The development team attempts to keep track of who is using the databases and engines and how they are using the information. For those visiting the TED and ICE sites, the team uses web statistics applications, run by American University, to monitor users' entry points (referring sites), points of origination, and other data characteristics. To account for those using the information on their own site, public search engines are used by the team for information monitoring. Four groups appear to be the primary users of the models' information. They are universities, private groups, governmental entities, and non-governmental organizations (NGO).

ICE and TED case study information is often used and cited on student, department, and research sites at Columbia University, Georgetown University, The University of North Carolina, Tulane University, Florida State University, The University of Lyon, Monterrey Technical Institute of Mexico, and many other institutions. It is also used by private news and research groups like Yahoo.com's social science directory on peace and conflict studies, Cable News Network's (CNN) In-Depth Specials on the environment and first Gulf War, and Omniseek.com's coverage on Bolivian history.

Government organizations that have used the case studies include, the National Oceanic and Atmospheric Administration in their Damage Assessment and Restoration Program's documentation, National Institutes of Health in their Environmental Health Perspectives Journal, and U.S. Agency for International Development's (USAID) coverage on Czech and Romanian environmental issues. From the non-profit / non-governmental sector, examples of organizations who have used the case studies are United Nations Educational, Scientific, and Cultural Organization (UNESCO) with their work on the management of scarce resources, United Nations Environment Programme (UNEP) Post Conflict Assessment Unit's index on the history of conflict in Iraq, and The Environmental Literacy Council's reporting on the Iraqi Marsh Arabs.

Future Plans for TED and ICE

The development of the TED and ICE databases and pattern matching systems that include over 900 case studies are a decade in the making. Adding more cases will continue. It must be acknowledged that these projects have required an intense amount of focus on the development of the pattern matching tools. But this grander expansion of the uses of categorical case studies and web technologies only suggests some of the horizons for exploration.

The case studies have been used in a variety of distance learning efforts starting with actual courses and connected to educational partners in the Middle East and East Asia. The computer aspect of the class makes it amenable to online teaching and learning. Other classes and students have contributed to TED and ICE, ranging from Sweden to Costa Rica. In 2003, Truman State University in Missouri established an entire class devoted to producing case studies and George Washington University graduate students regularly contribute case studies. TED and ICE cases also are offered online as part of totally online distance course covering the associated topics of trade, environment, and conflict, which began to be taught in the summer 2004 at American University.

Conclusion

The usefulness of the TED and ICE is apparent in the fact that their longevity has allowed the development of a basic set of materials that is now large enough to have useful application to both theory and policy. It has proven useful as an innovative research tool where none existed in the two respective fields of inquiry. These tools can be the catalyst for other efforts, as Mendelev did in developing a basic set of case studies of chemical elements in a tabular form. A small investment carried on religiously over a long time can reap substantial teaching, research and policy benefits, especially when combined with emerging, effective, and ever-developing information technologies. A group of students and faculty using these web based information technology tools in tandem can have a significant impact, slowly, but incrementally, over a long time horizon. These can result in large-scale teaching and research activities.

About the Authors

Dr. Joshua Franzel is an Affiliated Graduate Fellow with the Center for Teaching Excellence at American University and a current Presidential Management Fellow with the United States Government. He can be reached at: josh.franzel@alum.american.edu

Dr. James Lee is the Associate Director for Technical Support and Training for the Center for Teaching Excellence and a faculty member in the School of International Service at American University. He can be reached at: jlee@american.edu

Sources

Cable News Network. A decade since Desert Storm: Environmental impact - Oil fires and spills leave hazardous legacy. Retrieved January 20, 2004 from http://www.cnn.com/SPECIALS/2001/gulf.war/legacy/environment

Environmental Literacy Council. Marsh Arabs. Retrieved January 11, 2004 from http://www.enviroliteracy.org/article.php/598.html

Lemay, Laura (2nd ed.). (2000). Web Publishing with HTML 4. Indianapolis, IN: SAMS.

Mandala Projects. Retrieved January 25, 2004 http://www.american.edu/TED

Mandala Projects. Inventory of Conflict and Environment. Retrieved January 25, 2004 http://www.american.edu/TED/ice/ice.htm

Mandala Projects. Trade & Environment Database. Retrieved January 25, 2004 http://www.american.edu/TED/ted.htm

National Institutes of Health. Environmental Health Perspectives Journal. Retrieved January 15, 2004 from http://ehis.niehs.nih.gov

National Oceanic and Atmospheric Administration. Damage Assessment and Restoration Program. Retrieved January 15, 2004 http://www.darp.noaa.gov

Omniseek.com. Lifestyle /History /South America /Bolivia. Retrieved January 20, 2004 http://lifestyle.omniseek.com/srch/%7B53422%7D

United States Agency for International Development. Regional Map and Country Links. Retrieved January 20, 2004 from http://www.usaid.gov/locations/europe_eurasia/countries

United Nations Educational, Scientific and Cultural Organization. Managing scarce resources: a challenge for democracy. Retrieved January 15, 2004 from http://portal.unesco.org/en/ev.php@URL_ID=14727&URL_DO=DO_TOPIC&URL_SECTION =201.html

United Nations Environment Programme (UNEP) Post Conflict Assessment Unit. Iraq Background Reading Materials. Retrieved January 11, 2004 from http://postconflict.unep.ch/iraqreading.htm

Yahoo.com. Social Science > Peace and Conflict Studies. Retrieved January 20, 2004 from http://dir.yahoo.com/Social_Science/Peace_and_Conflict_Studies