Searching the World Wide Web – Finding the Right Information the First Time

Ms. Kristine Donate, Florida Atlantic University

Twenty years in the RF Analog/Mixed Signal Design and EDA software industries doing design, customer support, application engineering, technical writing, training development and delivery, project management, AE and business management. Current focus is on telephony and IVR technologies.

Realization of software development as a true passion of mine led to the pursuit of a Master’s degree in Computer Science at Florida Atlantic University, graduated in December, 2013.

Dr. Ravi T. Shankar, Florida Atlantic University

Ravi Shankar has a PhD in Electrical and Computer Engineering from the University of Wisconsin, Madison, WI, and an MBA from Florida Atlantic University, Boca Raton, FL. He is currently a senior professor with the Computer and Electrical Engineering and Computer Science department at Florida Atlantic University. His current research interests are on K-12 education, engineering learning theories, and education data mining. He has been well funded by the high tech industry over the years. He has 7 US patents, of which 3 have been commercialized by the university. He has published earlier work on the use of the semantic web for medical applications at another conference. This work is part of an ongoing teaching and research project that leverages our collaborative teaching in smart phone app development. We plan to leverage this in generalizing the course offering so other interdisciplinary groups’ efforts are facilitated.

Dr. Diana Mitsova, Florida Atlantic University

Diana Mitsova has a background in research design, statistical and spatial analysis, as well as environmental planning and modeling using geographic information systems, and interactive computer simulation. Her primary area of research involves the impact of urban development on ecosystems and other environmentally sensitive areas. Her recent publications focus on the impact of climate-related stressors on coastal communities and the implementation of planning approaches related to enhancing coastal resilience to natural hazards. Her research has been funded by the National Park Service through FAU Environmental Sciences Everglades Fellowship Initiative, USGS, and The Nature Conservancy.

Mr. Francis Xavier McAfee, Florida Atlantic University

Francis X. McAfee, Associate Professor in the School of Communication & Multimedia Studies at Florida Atlantic University (FAU) merges his background as a ceramic sculptor and printmaker with new digital technologies. After graduating with a BFA in Art in 1989 he joined the Florida Center for Electronic Communication (CEC) as a lead artist creating animation for applied research projects. These computer animated films were nationally and internationally screened in New York, Chicago, Hollywood, San Francisco, and Tokyo in industry recognized competitions as the International Video Art Competition, the New York Festivals, and the American Film Institute.

McAfee is also active in web-based virtual reality projects. His research includes digital archaeology of a deteriorating ancient tomb in Sicily to help preserve and visualize its’ characteristics for future study. His collaboration with Florida International University’s International Hurricane Research Center showed how certain roof construction materials may become projectiles during high wind events. For the FAU Center for Environmental Studies’ Sea Level Rise Summit McAfee lead a student team to produce a short animated video showing what might happen to the neighborhood around the Miami Freedom Tower if sea level rises to its full potential impacts. The video was picked up by National Public Radio and other media outlets. In 2007 he helped visualize the research of FAU’s Ocean Engineering using animation for a competition for a State of Florida Center of Excellence. FAU won the completion and has since been named as a national research center, Southeast Regional Marine Renewable Energy Center. Since 2009 he has collaborated with colleagues in Computer Sciences and other colleges to form cross-disciplinary student teams that create software applications for Android mobile devices.

©American Society for Engineering Education, 2015
McAfee compliments his professional activities with volunteer service for ACM SIGGRAPH. He served on their executive committee as Director for International Chapters and has organized local Fort Lauderdale chapter events for over 20 years.
Searching the World Wide Web – Finding the Right Information the First Time

I. Abstract
This paper addresses a real-world problem on how to help a multi-disciplinary team become more productive. A group of faculty members from various disciplines as related to the development of a smart phone app brought together undergraduate students from these various disciplines and helped them develop a state-of-the-art App. We found that communication across the various discipline-specific cultures was the most challenging, both at the levels of professors and students. Misunderstandings were common because of differences in their intra-disciplinary languages, tools, physical models, and skill sets, and intra-disciplinary cultures in terms of mental models of other disciplines and stereotyping. However, we also found that much of the uncertainty and friction can be overcome by providing a solid foundation of expectations, milestones and communication mechanisms from which to work. This paper addresses one important aspect of that effort, viz., of providing access to comprehensive app-specific content knowledge on the World Wide Web so the groups can progress without much communication gaps. Though this paper addresses issues specific to content of the apps, the same approach can be applied to providing a common platform for communication with regard to engineering, aesthetics, marketing, and other aspects of app development as well. The semantic web approach is generic enough to be applied for any multi-disciplinary setting. The effort is worth the results in terms of reduced misunderstandings and enhanced productivity.

II. Introduction
As the World Wide Web (web) continues to expand, the information available via the Web is an ever-growing forest of useful and non-useful details. The casual user may be surviving their attempts to query the web by typing a few keywords into their favorite browser while performing an initial search. Typically the initial search is followed by one or more searches each with a refinement of the search keywords based on the search results. The resulting task of separating the relevant search result from those that are distracting and non-helpful often presents the user with a daunting and time consuming task. An example of application of our approach is the recent decision by Google to rank their results based on the ‘truth’ value of the hits\(^1\). Information overload is tamed; if one disputes the relevance of Google’s algorithms, it is easy to apply their own algorithms and build another parallel search engine. The community will ultimately decide the value of either or both.

Today, teams of students at our university are working with each other across multiple departments such as Computer Science, Urban Planning, Communication and Multimedia, etc. to develop mobile applications providing a variety of end-users with a wealth of environmental and community information as described in our 2013 ASEE Conference Paper\(^2\). Our paper will describe the effort underway to assist the students development teams find the information they need on the web by leveraging the Semantic Web. Starting with an introduction of the Semantic Web, we will explain how information and knowledge can be more efficiently extracted from the deepest and widest recesses of the web. We will detail the various methods being employed to collect information from the student teams such as surveys and presentations along with question
and answer sessions. Students are providing strategic key words used and websites found while researching aspects affecting the applications they are developing such as educational information, funding/spending, community impact and subject matter fundamentals. Next the concepts and descriptions of an ontology and taxonomy will lay the foundation of our premise. The student provided keywords and the most applicable website they lead to are being categorized and organized to provide a structural framework upon which the ontology for our search tool is based. A presentation follows of the taxonomy defining the relationships between keywords and the location of most relevant information associated with the specific keywords on the web. Finally, we will leverage existing work utilizing the semantic web and storage of keyword collections, or profiles, in a Semantic Search Engine tool. Through the modifications described to the existing tool, we will demonstrate how to optimize and drive maximum efficiency while searching the web and various on-line databases ultimately aiding the student teams to be more efficient in their research efforts.

III. Introducing the Semantic Web
According to Merriam-Webster Dictionary, semantics is the study of words and phrases within a language to grasp meaning and understanding. This is the fundamental concept behind the semantic web. Instead of just searching the content or keywords of a web page hoping for a match, by providing structure and relationships within the page data, a semantic web application can find the most relevant information during an initial search. Furthermore, the located information can in turn be used to uncover additional related or contingent information used by our smart applications which understand the meaning and relationships of the data extracted from the web.

In 2001, Berners-Lee presented in Scientific American his vision of the semantic web not as a replacement of the web, but instead, a refinement and enhancement of the web. As he did with the web, Berners-Lee defined the components of the semantic web; the ability to leverage the universality of the web, a language and method that provides meaning to the data, and a structure with which different databases may be logically related. The key behind use of semantics and the semantic web is to use the natural language, which has a contextual meaning making the search more effective.

The web, as defined by its very name, is a universal resource comprised of data and servers around the globe. The limitless reach of the Internet permits a student, anywhere in the world, to learn about certain unique situations elsewhere; for example, the native South Pacific and Indian Ocean fish species that now invade the waters off South Florida. Information is also available regarding the hazard this non-native fish is causing in the South Florida waters and what various entities are doing to combat the problem. However, all this information is disparate and it is left to the student to relate the data from the various sources into a cohesive collection of information.

The key to provided meaning starts with the use Extensible Markup Language (XML) and Resource Description Format (RDF). Not constrained by a fixed nomenclature, XML tags can be completely customized to assert whatever definition is necessary such that data within a source, say a web page or a database, has been identified. For example, let’s say you have a data source with two objects. You want to provide identification in some way stating these two
objects are ‘people’, that is, each one is a ‘person’ object. Using XML and a ‘person’ tag you have defined, you can represent your object by surrounding it with the <person> tags. You could do the same with an address, phone number, title, and so on. Unfortunately, although we have identified an object within our data, it still has no meaning or relative definition for use by an application. We need to introduce a hierarchy of sorts and provide a manner in which multiple objects may be identified as the same type, or class of object, yet each object is distinct from the other. Furthermore, we must define a meaning or relative association to our object.

Once the individual aspects of the data have been identified, RDF is used to provide logic and meaning through the use of RDF triples. Similar to the way we use English language with nouns, adverbs and verbs to construct a sentence to define who does what and how that action is done, RDF triples state the relationship between bits of data. For example, let’s say you have a data source in which each of two objects are identified as a person and you want to define a relationship between these two objects that they “know” each other. Using RDF triples, or statements, containing a subject, predicate and object, you can do just that as seen in Figure 1 from Hebeler.

![Figure 1 - Sample RDF/XML](image)

Applying our example, the subject is the first person, the predicate is the property defined as “knows” from FOAF, and the object is the second person. FOAF stands for ‘Friend of a Friend’ and provides a dictionary of people-related terms that can be used in structured data, as with RDF.

With our data meaningful, we can now define the structure through which data within and among knowledge bases can be related. An ontology provides the relationship and logic, or rules, among definitions and knowledgebases. The classes we have discussed thus far and the properties associated with the classes are part of a taxonomy within the ontology. It’s the addition of logical rules with the taxonomy within our ontology that enables our search application to “understand” the data provided and complete its goal. Our complete ontology will be detailed in section V.

IV. Our Specific Environment

Several professors in varied disciplines at our university are collaborating to provide a truly unique approach to educational experience. Over the course of three semesters, undergraduate, graduate and high school students work together from the Engineering, Business, Arts and Anthropology/Sociology departments to create apps for Android smart phones. In each of the Fall, Spring and Summer semesters, teams of cross-department students both in on campus classes and on-line, work to design the components and platforms for the applications; build prototypes; creating marketing materials and deliver the completed apps. During the Fall 2013, students in computer engineering and multimedia studies teamed with graduate and undergraduate students in urban planning to explore domain-specific applications related to environmental issues and sustainability (see Table 1).
<table>
<thead>
<tr>
<th>Application Topic</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marine Debris</td>
<td>Education on impact of marine debris on marine animal</td>
</tr>
<tr>
<td>Sea Turtles</td>
<td>Identify, conserve, document and reporting of sea turtles</td>
</tr>
<tr>
<td>Lion Fish</td>
<td>Identify, handling, documentation and reporting of LionFish</td>
</tr>
<tr>
<td>Shark Migration</td>
<td>Identify, handling, documentation and reporting of LionFish</td>
</tr>
<tr>
<td>Beach/River Health</td>
<td>Provide map based ratings and advisories</td>
</tr>
<tr>
<td>Green Neighborhood</td>
<td>Rank and compare neighborhood for environmental awareness</td>
</tr>
<tr>
<td>Rainwater Harvesting</td>
<td>Educate and promote use of capture/use of rainwater</td>
</tr>
<tr>
<td>Environmental Education</td>
<td>Educate children regarding their impact on the environment</td>
</tr>
<tr>
<td>Solar Energy</td>
<td>Promote and assist with adoption of solar energy</td>
</tr>
<tr>
<td>Community Gardens</td>
<td>Educate and map area of urban farming</td>
</tr>
<tr>
<td>Brownfields</td>
<td>Promote awareness of environmental impact areas.</td>
</tr>
</tbody>
</table>

| Table 1 - Android Application Topics |

The student projects were distinct enabling the teams to use real databases while creating useful prototypes to serve local communities ensuring applications subsequently developed will go beyond simply providing information. The prototypes merged the skills of all three groups of students into holistic end products incorporating user interfaces, graphics and animation to create educational and interactive apps. The initial conceptual models of each prototype were further enhanced by adding content map layers, improved user interfaces, Bluetooth, multi-threading, web services and performance optimization. Students developed skills in using mapping applications, interacting with layers, features and attributes, and running analysis tasks. Many apps included games and maps as tools for engaging the user. Several groups interacted with local stakeholders to improve the currency and usefulness of their respective app. Students also created promotional presentations using video production for the purpose of marketing their apps. At the end of the fall semester, panel of judges representing local industry professionals evaluated the prototypes. These interactions help the students to connect with industry stalwarts, be motivated to achieve higher, impress these professionals and find local jobs. The teams who developed the top-ranked apps (the top 4 in Table 1) were invited to showcase their work at conferences and other venues designed to connect mobile technology professionals and businesses. Figures 2 and 3 below show splash screens from top two ranked Apps. For complete documentation of these apps, please visit our Bitbucket site.

Research is a natural part of academic pursuits and students in the Android app development program needed to understand the nature of the topics their apps would address. The explosion of the internet and access to information on the World Wide Web (WWW), or the web as it is also known, has changed the paradigm of conducting research. Gone are the days of searching the “stacks” of the old library as are the hours pouring through microfiche and microfilm. Today, simply sit down in front of a computer with your favorite web browser and you quickly have access to published media, videos, recorded classes, presentations, etc. In fact, according to Google, 1 billion searches are conducted every day through Google’s search engine. Utilization of the web has simplified gaining access to research materials. No doubt everyone has utilized the web to find information, be it the price of the latest technical gadget or information on migration patterns of the Black Tip Shark. And yet the web is a fairly young technology emerging in the early 1990’s out of efforts at the European Center for Nuclear Research (CERN).
It is at CERN that Tim Berners-Lee first assembled the components necessary for the realization of the web: a browser (client), a server, a protocol for client/server communication (HTTP) and a language for presenting information (HTML)\(^\text{10}\). Since then the growth of the web has been nothing short of explosive. The rapid growth itself caused one of the inherent problems of today’s web: Getting the correct information and leveraging found information effectively continues to beset web users. Improvements are being made at record pace to increase the speed and efficiency of our searches. Research has shown that 98% of web searches users only examine the links presented on the first page of results\(^\text{11}\). Therefore, getting the right information to not only the first page, but the top of that first page is critical for getting the right information to the user. Search Engine Optimization (SEO) is a practice which seeks to make sure web page content contains the most relevant information, such as keywords, users are likely to utilize when searching. Customers are lining up for SEO technicians to optimize their web sites as the SEO service industry is forecasted to grow from $2.1 billion in 2011 to over $3.5 billion in 2016\(^\text{12}\). And yet, the typical browser still remains ignorant about the link between its

---

\(^\text{10}\) 

\(^\text{11}\) 

\(^\text{12}\)
keywords and content. This is where the concepts behind the Semantic web come into play. This is the problem the Semantic Web is poised to solve.

V. Methods
Even before the Fall semester started, students in the Urban Planning department began assembling a list of topics the applications would address. Topics spanned environmental issues as well as energy related topics. During those initial weeks the Android apps teams refined their
conceptual understanding of the technology and its' feasible mapping to develop apps for urban planning. During these interactions it was easy to see patterns developing to relate the topics to one another. For example, topics such as shark migration and the invasive lionfish could be part of a larger marine life category. As a natural part of understanding the issues various applications would address, students started their research using the web. The initial list of topics assembled contained not only the application topics, but also a basic description of the associated issue and the URL to a web site containing key information.

Over the course of the semester, the Android apps teams gathered to share experiences and present their progress. During one of these first gatherings, a presentation was made by each of the Android apps teams that detailed how they were conducting the research, the information they uncovered and primary location where the information was found. And although work done for this semantic search project would not directly aid these Fall 2013 Android apps participants, students were extremely receptive to the notion that their contribution would benefit future students. At the two-thirds mark into the semester, the Android apps teams were significantly progressing on their applications. During another round of presentations, groups demonstrated beta versions of their works. Each group’s presentation led to question and answer sessions on how the application would engage the user, but also how the application would lead the user to additional information. A very interesting aspect of their work came to light, not only had the students conducted research to understand the significance of their topics. But they had also uncovered existing databases, government agencies with on-going projects, and social media sites centered on the issue at hand, among other related information. It became clear a simple list of topics, keywords, and web sites, was too limiting. Instead, our list also needed to capture the nature of the keywords and websites along with their relationship to the application topic. Furthermore, with the students now entrenched in the detailed implementation of their applications, a less intrusive method was required to facilitate collecting keyword and web site links, and information about their relationship to the issues. A questionnaire was used to gather information (via the web, of course) from the Android apps students. The questionnaire consisted of seven two-question sets. The first question in each set asked the student to provide a list of keywords used to find the most useful web site for finding information about a particular aspect of their application issue. The second question of each set asked the student to respond with the web site URL the keywords had led them to. Each of the seven sets of questions addressed a different aspect of the students’ research; subject matter information, community impact information, scientific and/or educational information, governmental entity information, funding available or monies being spent, websites for application user to sharing information and data/databases. Although it would seem collecting a single questionnaire from each team would provide the required data, further consideration revealed each student contributed to their teams efforts in a unique manner. The end result was the product of each team, but within a team each student offered a different perspective and had done research of their own. It was determined that responses to the questionnaire (see Figure 4) from each student instead of just one per team would provide a richer and more complete collection of keywords and web sites.

Finally, the structure of the on-line questionnaire was set so that each question must have an appropriately formatted response in order to be submitted. As this process was in an academic environment, participation was encouraged by making completion of the questionnaire part of
the students’ coursework. Not too surprisingly, the response rate was over fifty percent. Results were maintained in a spreadsheet and the data feed right into our semantic web ontology.

In addition, a learning outcomes assessment was conducted using pre-course and post-course questionnaires which included thirty-two questions grouped in six learning outcome categories:

**Urban Planning Application Research**

Please complete this questionnaire with respect to the research performed for your Android Application development project.

* Required

1a. Subject matter information: Please list between 3 to 10 keywords you would use to find the most helpful websites while gathering this type of information.

Think about the research you did for background information on your application’s subject. Use a semi-colon to separate individual keywords from keyword groups. For example: green sea turtles; rainwater harvesting; solar; invasive species.

1b. Subject matter information: Please enter the URL (link) to the most helpful website you found while gathering this type of information.

For example: [http://epa.gov/](http://epa.gov/)

---

**Figure 4 - Sample Urban Planning Application Research Questions**
programming skills, critical thinking; problem solving, logistics of app elements functionality, target audience, and presentation of the app prototypes. Fifty-five students enrolled in the concurrently taught courses from the three disciplines (i.e., computer engineering, communication and multimedia studies, and urban planning) completed the questionnaires. Knowledge gains were assessed on a Likert scale from 1 to 5 where 5 indicated the highest level of self-reported improvement. Using the Wilcoxon Mann Whitney test-statistics, we tested the null hypothesis that the median difference between the pre- and post-test observations is zero, that is, $H_0: \mu = \mu_0$. Results of this survey instrument are included the Results section.

VI. Results

Just as Berners-Lee originally described, in order for the Semantic Web to succeed, there must be a structure to the information enabling computing applications to “understand” the web content they access\(^9\). Our goal was to create an ontology providing relationships and structure for the topics themselves, but also to the keywords and information sources about those topics.

Reviewing the app topics it was easy to discern two primary categories, energy conservation and environmental issues as demonstrated in Figure 5. By starting at the top with fairly broad categories, we ensure that the ontology can be further refined to include topics and issues addressed in future Android app development cycles.

Figure 5 - Urban Planning Ontology Topic

Moving a step down the ontology, under energy topics we have defined a class specific around solar energy. Moving back to the environmental side, several classes were defined to capture the marine life, conditions/hazards, sustainable communities and climate associated with environmental issues tackled by the app development teams. At this point we noticed that certain issues should be more appropriately identified under multiple 2nd level classes. However, for this revision of the ontology, we leaned towards a straight-forward approach and determined
the parent/child relationship was best inferred if not directly implemented. Hence, the marine life class was subdivided into marine life – conversation and marine life – invasive species as seen in Figure 6. The three remaining environmental topics (climate, sustainable communities
and conditions/hazards) are maintained at the same level of hierarchy as our two marine life classes. In order to provide our search application with a good starting point for its web search, each topic class has associated primary web sites where the students found the most helpful information for the given aspect of their research. Attribute properties were defined to hold the web site information corresponding to the area of research: CommunityImpactSearchSeed, DatabaseSearchSeed, FundingSearchSeed, GovernmentEntitySearchSeed, SciEduInfoSearchSeed, SubjectMatterSearchSeed and WebsiteSearchSeed. Each topic class was assigned attribute properties with the corresponding web search seed URLs.

Our final classes are those directly associated with the corresponding keywords. These are the classes our search application will utilize. However, we must be able to stipulate which keywords are to be used when searching for the particular aspect of information associated with the given issue. That is, which keywords should the search application use when search is done of funding information around community gardens instead of social networking information? To solve this confusion, additional attribute properties were defined within the ontology to designate when a keyword should be used during a search within a given information aspect: isCommunityImpactKeyword, isDatabaseKeyword, isFundingKeyword, isGovernmentEntityKeyword, isSciEduInfoKeyword, isSubjectMatterKeyword and isWebsiteKeyword. Each keyword is assigned the appropriate attribute property such that the keyword class is designated as a part of one search aspect over the other. See Figures 7 and 8.

![Image](image_url)

**Figure 8 - Example keyword attribute property**
As mentioned under ‘Methods’, learning outcomes assessment was also conducted. The results from the pre- and post-testing of self-reported learning outcomes indicate cross-disciplinary knowledge gains in all six categories. The most significant knowledge gains occurred in critical thinking and problem solving skills as well in presentation skills. Increased awareness of the target audiences and gains in programming and app elements logistics were also reported.

VII. Discussion
Although we could have taken many different directions when defining our ontology, our path was guided by the existence of search application tool as described by Islam, Freytag and Shankar in 2013. The Diabetes Management tool utilizes the semantic web to find information pertaining to drugs used by diabetes patients\(^\text{13}\). To use the application, a patient will select the category (currently associated with types of diabetes management drugs available). The application will then present the user with a list of keywords associated with the selected category. When the user clicks the search button, the application will search a pre-determined set of databases (currently a collection of U.S. Food and Drug Administration web-sites) to find articles relative to the keywords. The search results are sorted based on the user’s preferences to rank the most relevant articles higher than others thus ensuring the patient finds the information they need after just one search.

The goal is to modify the application for more generic searches such as our Android app research with a few key updates. First, the primary web site used to start the searches will be set by the attribute property within the ontology. Second, an aspect preference will be added to the application interface so the user can select the aspect (subject matter information, funding sources, government entity, etc.) of information they seek. This preference will drive the use of the corresponding web site seed along with the presentation of the associated keywords to the user as defined by the attribute property also in the ontology. And third, the refined list of keywords, instead of all the keywords under a given topic, is used by the application for the searches. Figure 9 contains a sample mockup of the modified search application including the

![Figure 9 - Sample Search Application Mock-up](image-url)
user’s ability to set the search aspect. Note the categories reflect the various Android app topics and the keywords associated with the search aspect are displayed.

VIII. Conclusion
Access to all the information available on the World Wide Web is truly vast. Now more than ever, searching the web takes time and expertise to locate and extract the information most relevant. Applying the principles of the Semantic Web, meaning and reasoning, can make searching the web more effective and efficient. This was demonstrated by defining an ontology around everyday topics used by multi-departmental teams of students developing Android smart phone apps. An existing application will be updated to realize the web searching utilizing our Urban Planning ontology. The ontology was defined in such a manner for easy expansion and optimization. For example, we plan to also enhance the search application such that users can add topics and keywords through the application, which will in turn be added to the ontology. The topic and keywords will be added as classes at the appropriate level in the ontology and assigned the appropriate attribute properties. There is no limit to the amount of information to be made available on the web. Therefore, with continued improvements, both the ontology and search application will prove valuable tools for researchers in the very near future.

IX. Bibliographic Information
7. FOAF Vocabulary Specification 0.98. URL: http://xmlns.com/foaf/spec/.
8. Bitbucket open source site: https://bitbucket.org/shankarfau/profile/teams - it has the entire App related documentation, code, and assets of our collaboration.