Teach Astronomy – A Comprehensive Online Astronomy Education and Outreach Resource

Chris D. Impey
Dept. of Astronomy, the University of Arizona
933 N Cherry Avenue, Tucson, Arizona 85721, USA

Kevin K. Hardegree-Ullman
Dept. of Physics and Astronomy, the University of Toledo
2801 West Bancroft Street, Toledo, Ohio 43606, USA

Anand Patikkal
Amazon Headquarters
410 Terry Avenue North, Seattle, Washington, 98109, USA

Carmen L. Austin
Dept. of Astronomy, the University of Arizona
933 N Cherry Avenue, Tucson, Arizona 85721, USA

Abstract

A web site called Teach Astronomy (http://www.teachastronomy.com) has been created to serve astronomy instructors and their students, amateur astronomers, and members of the general public interested in astronomy. The content includes astronomy articles from an introductory level textbook and from the online resource Wikipedia, short video clips, astronomical images, podcasts, and recent news stories. This article describes the technology behind the delivery of those learning resources, which is relevant to the capabilities and limitations of the web site. One key innovation is the Wikimap, a Flash-based tool that presents the visual results of a real-time clustering analysis of hundreds or thousands of text items, displaying the item that best matches the search term and most closely related items. The clustering is carried out in a Lucene index, and it can operate on any database containing items of text. The astronomy content is routinely updated, in some cases daily. Due to the prevalence of smartphones, tablets, and other handheld devices, a simplified non-graphical version of the interface was developed using custom style sheets. Teach Astronomy has a large following of students taking introductory astronomy classes and members of the public with a recreational interest in astronomy. In the past year, there have been 250,000 unique visitors. Currently we are developing a new interface that uses HTML5 instead of Flash to display the Wikimap, an app version of the website for use on smartphones and tablets, and tool to support an instructor and learner community.

Keywords: Astronomy education; online resources; electronic textbook; multimedia content; human-computer interface; lifelong learning

1. Introduction

Since British physicist Tim Berners-Lee published the first publicly available web site on the Internet on August 6, 1991, the Internet has grown by leaps and bounds to become a pervasive aspect of modern life (Xiaojing, 2011), and Internet traffic is expected to grow by a factor of a thousand in the next twenty years (Saleh, 2011). Information floods the Internet, and navigating and harnessing that powerful resource has been a task taken on by giants such as scientists and technology companies. Search engines have become so invaluable that "Google" is used as a verb to describe the action of any general web search. The highest ranked web sites in the world in terms of traffic are Google, Facebook, YouTube, the Chinese search engine Baidu, Yahoo, Amazon, and Wikipedia (Alexa, 2015). This list illustrates the demand for quick general, social, and/or encyclopedic information. Wikipedia in particular has expanded enormously since its inception in 2001. Nearly five million articles are now available on the English language version of the site and most of those articles have been written, edited, and maintained by members of the general public (Lih, 2009). Given this, the generally high accuracy of the Wikipedia science articles is remarkable (Giles, 2005). Science is well served by the Internet, although the roughly ten million unique visitors of two of the top sites, hosted by NASA and the National Optical Astronomy Observatory (NOAO), are dwarfed by the half billion unique monthly visitors of Wikipedia (The Wikimedia Foundation, 2015). The Internet has not only given the public much better access to scientific information, it has changed the way science is carried out (Shankar, 2008). Astronomy has made particularly good use of the Internet for outreach in addition to education (Author, 2012).

The digital age has ushered forth a new set of tools for those in academia to use to disseminate and assimilate knowledge. Novel web technologies, high-speed Internet connectivity, and the increasing presence of mobile Internet-capable devices make information immediately accessible to increasing numbers of people. As astronomy educators, we have created an astronomy portal to make the task of finding general astronomy content easier. The goal is to provide thousands of undergraduate students, amateur astronomers, and informal learners with an easy-to-use web site containing a comprehensive array of astronomy information. We have utilized readily available technologies such as Adobe® Flash®, Apache Lucene™, and various programming languages to create a search and display interface called a Wikimap to help users easily navigate and discover relationships between content topics. This paper gives an overview of the technology, processes, data hosting, and development of the tools employed on Teach Astronomy. For information on the instructional motivation for Teach Astronomy and its intended audiences, see Author et al. (2013). Figure 1 illustrates the growing use of Teach Astronomy since its initial release in 2012. We note that the articles on the web site have been used as the online textbook for two massive open online classes (MOOCs) in the past two years. These courses on Udemy and Coursera have reached over 70,000 informal learners. Teach Astronomy continues to reach a mixed audience of students of introductory astronomy and lifelong learners.

2. Content Clustering

The backbone technology behind the Teach Astronomy web site is a database which allows any text-based content to be clustered according to keyword similarity. There are several traditional ways to demarcate related items on a web site. One is architectural, where the items are assigned to a particular web page or are part of a directory structure. The other is based on metadata, with items tagged by suitable keywords. The problem with the architectural solution is that it has to be coded by hand, so it is inflexible and requires additional coding to change. The limitation of the metadata solution is that metadata has to be added by hand and it requires a schema for defining appropriate keywords and applying them consistently. In this project, we wanted a scheme that was flexible and adaptable, and which could be applied to any text-based content.

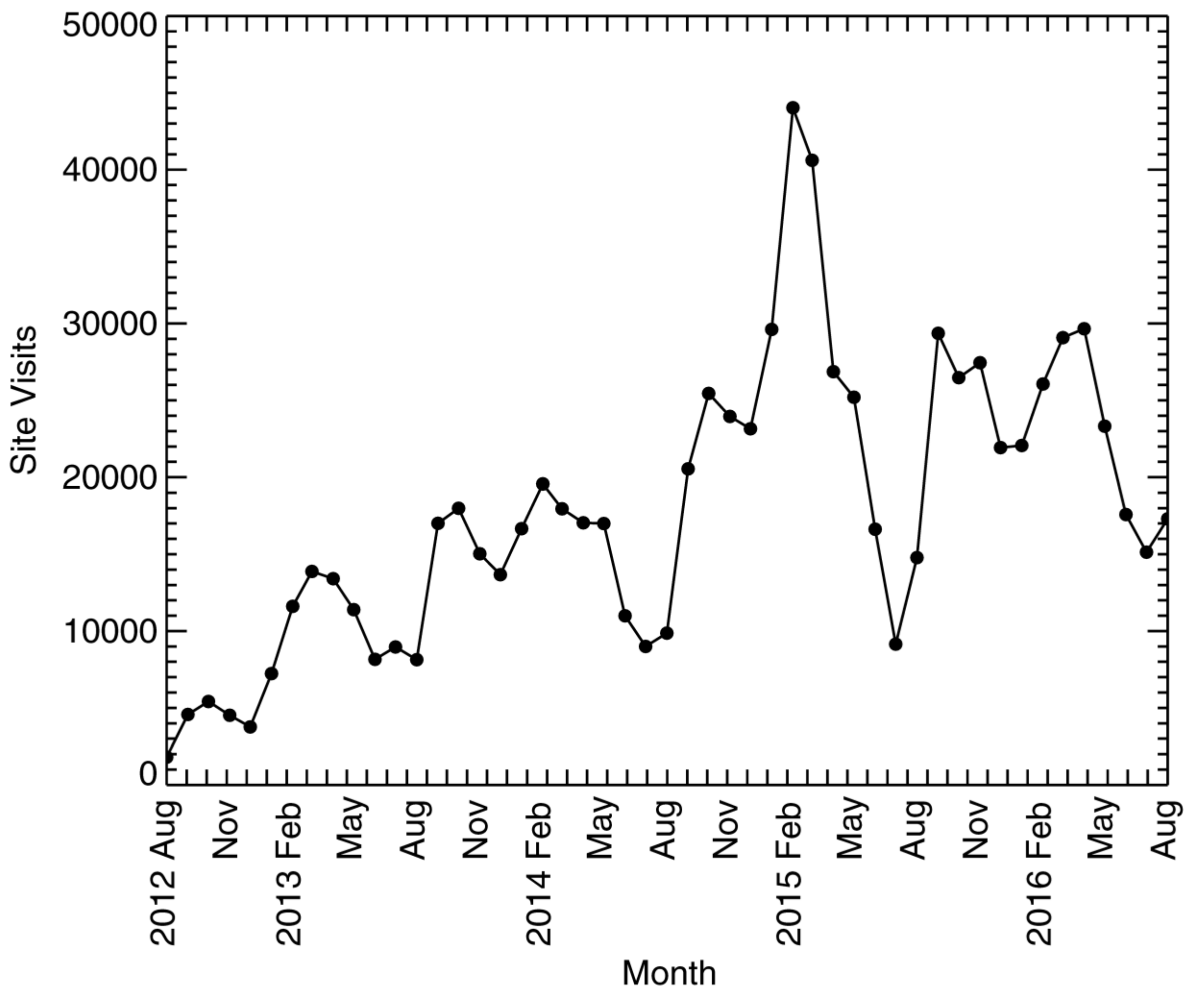

Figure 1. Monthly visits to Teach Astronomy. The number of monthly site visits to Teach Astronomy has been steadily increasing between August 2012 and January 2016 with noticeable drops during standard school break periods.

The first step of the process is to create a document index of text elements, using Apache Lucene™. For articles, the text of each article occupies a single document within the Lucene index. For images, the document consists of image captions. For video clips or podcasts, the index is populated with text transcripts of the video or audio files. Any text-based information can be clustered by creating a Lucene index. We have written a custom algorithm based on the Lucene index that takes the text from an indexed document and calculates a keyword frequency, where stop words and common words such as articles are discarded (from the index, not from the content). It then finds relationships between keyword frequency among articles and returns a ranked result based upon a search query. For any particular primary article, the result is a ranked list of closest items in the index based on keyword overlap. We modified the default Lucene search results using our own algorithms to improve the search technology. In particular, we added title filtering, which requires that the title of the primary item is contained within the text of any other items returned by the search. The title acts the same way as any other keyword, but it is given an added weight which yields better search results since the title strongly determines content in most cases. We inspected the results of over one hundred searches with and without title filtering, and using our general astronomy knowledge, we were able to judge that title filtering clearly improved the search results.

The clustering derived from a Lucene index is excellent but not infallible. Keyword overlap as a measure of the strength of relationship (proximity) between any two items of content works best when the database items all have homogeneous length and keyword density. Some keywords are specific to astronomy (like cosmology or parallax) while others have non-astronomical meaning (like stars or sky), and a much larger number of words are unrelated to astronomy and so their matches act like noise. Yet another group of words only have astronomical meaning when used in conjunction (like black hole and white dwarf). The algorithm behind the Lucene index cannot distinguish between these categories. However, clustering works because a significant number (or multiplicity) of overlapping astronomical keywords between two items is very unlikely to occur unless they have similar subject matter. Clustering becomes less reliable when text items are short, because the number of keywords is low and any overlap may depend on a handful of words. The key output from Lucene is an algorithmic "distance" between any two content items. This metric incorporates both keyword overlap and keyword density so it normalizes for article length and can handle situations where content items vary in length.

The results from a search query—the primary text item and the nearest *N* matches returned by Lucene—are put into an eXtensible Markup Language (XML) file. The XML file is parsed by our Flash tool to return search results in a visual display called a Wikimap. The entire process takes less than one second, and all computations are performed on a small five node computing cluster using quad core processor PCs. Content items are stored in the RAM of the computing cluster to return search results faster, unless prohibited due to licensing agreements from partner organizations. None of the procedures and methods are specific to astronomy, or indeed to any academic subject. The process described in this paper can be used to cluster any text content.

3. User Interface

The display interface for our search technology is called the "Wikimap." The Wikimap is an Adobe® Flash®, henceforth Flash, tool that displays results from a search query with the initial result showing as dark blue in a central node, and ten outlying nodes connected radially to the center and moving outwards in a spiral pattern from one to ten. The peripheral nodes contain links to the items most related to the central node. Item relevance is indicated by the distance from the central node, with the most relevant item appearing closest to the center. Navigation or “surfing” the Wikimap can be done by clicking on any of the external nodes. When an external node is clicked, a new Wikimap is displayed, clustered around the node which was just clicked, and a new set of ten closest results is displayed (Figure 2). This process can be repeated as much as desired, so sequentially clicking outlying nodes is a way to navigate the landscape of content based on similarity between the content items. Hovering over any node with the mouse will give the user options to view the item or perform a search for that item on Google.

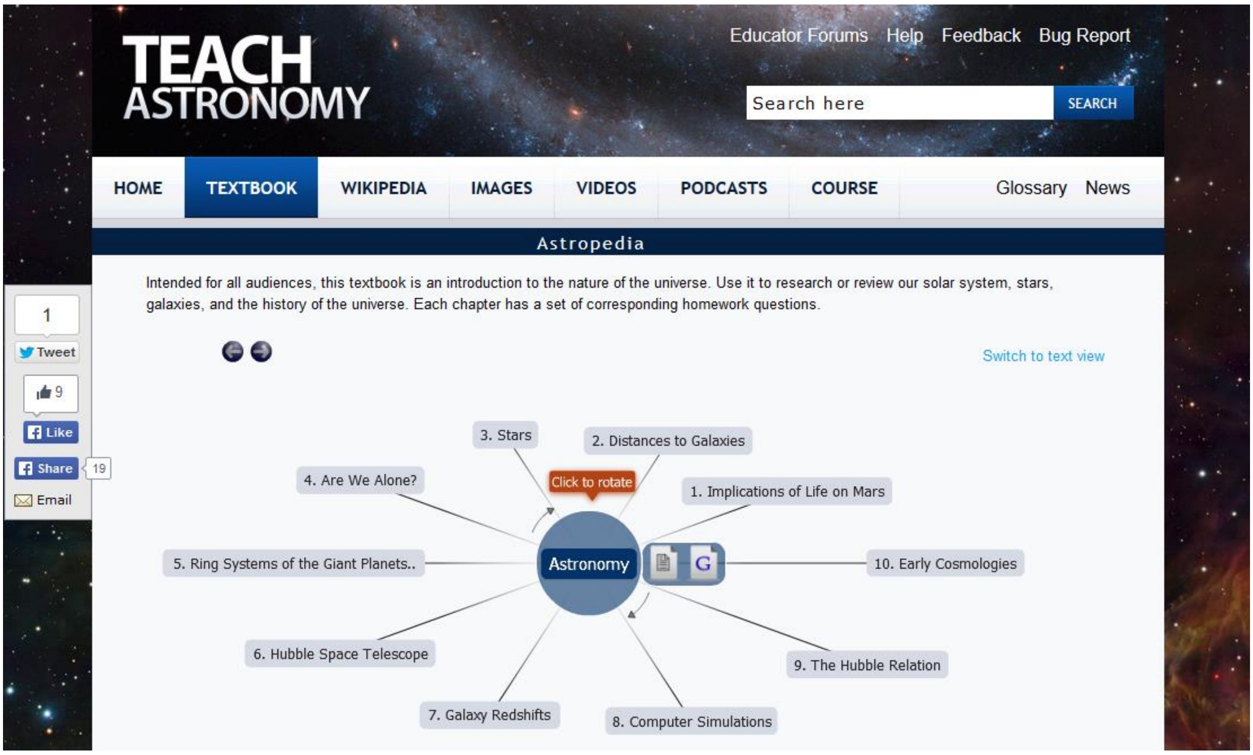

Figure 2. The Wikimap. The Wikimap view of textbook content on Teach Astronomy with “astronomy” as the search term showing the ten closest articles to astronomy spiraling outward from the search term.

The code behind the Wikimap is a flexible general graphing mechanism. In addition to the radial mapping interface described here, the Wikimap can be modified to display as a horizontal or a vertical tree layout which allows a hierarchy with multiple levels of nodes. Also, more than one node can generate a set of closest items so multiple parent-child relationships are possible. The style and colors of the Wikimap are adjustable, and arrows near the central node allow nodes to be rotated, so that overlapping titles can be read. The nodes in this implementation display only the title of an item of content, but the nodes are wrappers of great flexibility. In this astronomy teaching use of the Wikimap, the nodes contain either text or image thumbnails, but in principle many formats can be contained within a node, including web links or URLs, videos, animations, and sound files. Although only ten outlying results are displayed at a time, this choice is purely for aesthetic reasons, given the limited screen real estate. The number of nodes can be arbitrarily large and we have tested a paging mechanism so that the nearest 50 or 100 items to the central node can be displayed ten at a time. Visual navigation of web content can done well when there is metadata or an organizational scheme, but in their absence a solution of clustering by Lucene plus display by Wikimap is very effective.

4. Astronomy Content

Teach Astronomy is host to a large amount of information: 20 GB and 55,000 items. Updating and hosting this information depends on the nature and source of the content. The core resource is a set of 500 textbook articles, based on a textbook originally written by one of us (Author) and William Hartmann. After copyright reverted to the authors and then to Author the textbook was divided into separate articles for each subsection, with any explicit page and figure references and forward and backward references removed, so that each article could stand alone. Images were layered into the articles using public domain open copyright sources. The textbook article content is hosted, maintained, and updated on our own servers. Two comma separated text files separately contain article text and links to the images. A parser reads the information from the text files and displays it on the web site, allows the images to flow into the text, and provides a tool to click on any image and resize it larger. The textbook articles are updated about once a year by astronomy Ph.D. students at the University to follow the rapidly evolving state of research. With most astronomy textbooks costing over $100, there is a huge benefit to students in having access to a free, open source, learning resource.

The second major text resource is based on Wikipedia. Elsewhere, we have commented on the generally high quality of Wikipedia for science content (Author et al., 2013), and it continues to be an extraordinary resource, with major astronomy articles updated regularly, and many new articles written to keep up with the pace of discovery. Wikipedia provides its entire text database as a free download, and we use the articles in accordance with the attached provisos, with a link back to the article on the Wikipedia web site. We regularly download this database and host the 44 GB of information on our servers. We have filtered down the entire database to about 42,000 astronomy-related articles. This was done by selecting 25 high-level astronomy topics and then running them through our search and clustering algorithms using the entire Wikipedia database. Articles returned from those searches were placed in a “walled garden” of astronomy content. Given the “fuzzy” nature of the clustering, this process involved an algorithmic choice on the smallest amount of keyword overlap any article could have while still being likely to be about astronomy. The goal was 90% completeness and 90% reliability, verified by manual inspection of 13,000 articles. These 13,000 articles were assigned a number based on their relevance to the field of astronomy: 3 indicated a core topic in astronomy such as star or planet or galaxy, 2 was assigned to articles that were pertinent to astronomy, but not necessarily a core topic, such as a famous astronomer, space mission, telescope, or historical article, and 1 was assigned to articles that had a connection to astronomy but were either on obscure or specific (but not noteworthy) astronomical objects. Articles with no relevance to astronomy were discarded from the database. The incidence rate of irrelevant articles in the 30% of the database examined by hand supports the contention that the resource is 90% reliable. In practice, a user almost never encounters an irrelevant article since the "small world" aspect of the clustering keep results on-target.

Images come from two sources. Astronomy Picture of the Day (APOD) content is downloaded daily from the APOD web site (http://apod.nasa.gov). The content is then mirrored on our web site. We have obtained special permission from Robert Nemiroff to provide APOD content for Teach Astronomy users. Teach Astronomy is an educational website with no advertising so it falls under APOD's educational mission. We are not the creators or the owners of any of APOD so author credit is given on each article page along with a link to the original page on the APOD site. As of early-2016, there were 5,500 unique APOD images, after filtering for a modest number of duplicates where the most recent version is utilized. More recently we have added a second image resource, based on the Astropix web site (http://astropix.ipac.caltech.edu), hosted by JPL/Caltech's Image Processing and Analysis Center (IPAC). Astropix is under development and we are collaborating with Robert Hurt and his team to incorporate images from the currently operating NASA Great Observatories (HST, Chandra, and Spitzer) plus selected ground-based observatories. The APOD index is incremented daily as we add the most recent image, and the Astropix index is subject to periodic updates as new batches of images are released by IPAC. Clustering of these resources adds value by allowing users to find closely related content, in addition to searching by a keyword. Although the captions are shorter than a typical article, clustering the images is a robust and reliable process.

The last three types of content are video clips, podcasts, and news stories. All video lecture clips on Teach Astronomy are hosted on a YouTube page (https://www.youtube.com/user/astropedia). The nearly 1,200 videos are organized into 29 playlists based on subject matter; they are also available as a free download on iTunes U within the University portal. The filming was done using a good quality video camera with a green screen as background so that various astronomical backgrounds could be added in the editing phase, along with titles and transitions. Although the "talking head" format does not have the visual interest of TV or film treatments of astronomy, the goal of making video clips was more modest: to provide a modular resource that could easily be streamed to someone using a portable device with Internet access (see the next section). The podcasts are derived from the 365 Days of Astronomy project that started with the International Year of Astronomy in 2009. Podcasts have continued daily, and over 2000 are now available. By agreement with the project, we formed an index from the transcripts so podcasts could be clustered in the same way as articles and images. Content is downloaded daily from the 365 Days website (http://cosmoquest.org/x/365daysofastronomy). The podcasts are also linked directly from the 365 Days website. The final aspect of the site is astronomy news stories from the previous week. The Science Daily web site (http://www.sciencedaily.com) provides an RSS feed for their Space & Time category that contains mostly astronomy items. Using this RSS feed, we provide links to the most recent 25-30 articles from this category. We do not host the content so it is not subject to clustering.

5. Mobile Development

Although Teach Astronomy was designed to be viewed on a desktop web browser, the growing importance of handheld Internet devices led us to address this audience, who might be interested in the option to learn about astronomy “any time and any place.” The increasing market presence of smartphones gives the academic community an exciting new outlet from which to propagate information in a quick and easily accessible format both to students and informal learners (Cook, 2011). The mobile information environment is evolving very rapidly (Purcell et al., 2011). Two smartphone operating system platforms have taken the lead in the market over the past few years, Google’s Android and Apple’s iOS. Figure 3 shows world-wide quarterly smartphone sales by operating system based on data from Gartner, Inc. Smartphones running the Android operating system account for 80% of all smartphone sales in the last year. Apple’s iPhone, the only smartphone on the market using iOS, accounts for a little over 16% of all smartphone sales in the last year. The rise of smartphone sales in recent years gives us the opportunity to enter the mobile market and make Teach Astronomy accessible on these portable devices. Smartphones are becoming increasingly powerful and are now a valuable resource for general daily information. Their impact on formal education is yet to be measured or fully understood, but is likely to be profound (Shin et al., 2011).

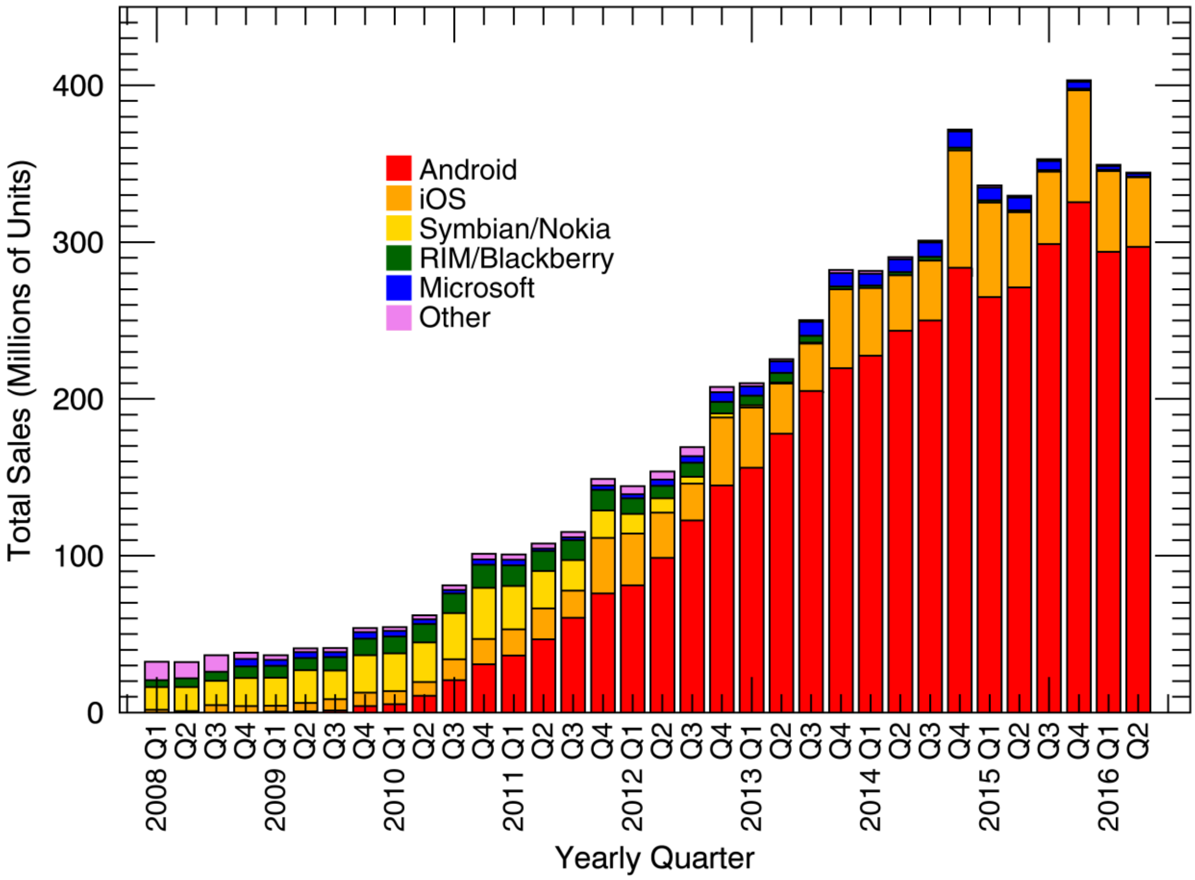

Figure 3. Smartphone sales**.** Smartphone sales per quarter have increased by a factor of nearly nine since 2008. Android and iOS clearly dominate the market. (Data from Gartner, Inc., 2008-2016)

For mobile device deployment we need to be accessible to all major smartphone platforms. The Wikimap is a Flash based tool. Not all smartphone devices support Flash, notably the iPhone, and with the recent decision by Adobe Systems Incorporated to cease development of the Flash Player to work in-browser with new smartphones, it is clear that we will need to make our tools accessible on mobile devices without the use of a Flash interface (Winokur, 2011). Eventually, we will convert the functionality of the Wikimap from Flash to HTML5, which is becoming the new standard for multimedia web delivery using the browser. Meanwhile, the method by which we return search results from a Lucene index allows us to do this using stylesheets for different smartphone browsers. Search results on smartphone browsers are automatically displayed using the "list view" option. The list view displays search results in an ordered list rather than in a radial Wikimap. Functionality is similar to the Wikimap wherein the user can click on a search result to find the closest related items to that query. Icons to view content of a search result are displayed to the right of each result. Figure 4 shows the mobile device layout of Teach Astronomy.

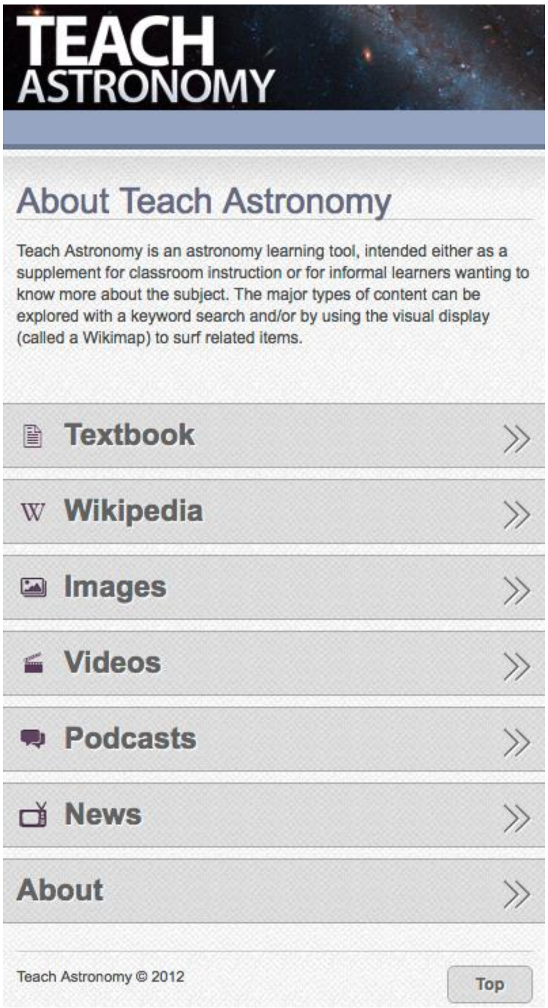

Figure 4. Teach Astronomy in mobile. Mobile device layout of Teach Astronomy displaying content in list view.

In the 2014 EDUCAUSE Center for Analysis and Research (ECAR) Study of Students and Information Technology, 86% of undergraduate students report owning a smartphone and 47% report owning a tablet device, up by 10 and 16%, respectively, from the previous year (ECAR, 2014). Among the students who own a tablet device, most use the device for academic purposes such as emailing professors or students and checking grades, and the number one use for smartphones in class is looking up information (ECAR, 2013). A recent Harris Interactive survey conducted on behalf of the publisher Pearson indicated that college students are very positive in their views about tablet computers, with 80% believing that tablets can transform the way students learn and 60% believing that tablets will replace textbooks in the next five years (Pearson, 2013). The significant interest by students in the tablet market gives us another outlet for Teach Astronomy. Development for tablets, netbooks, and Internet capable e-readers is similar to development for smartphones. The Teach Astronomy web site will automatically detect which browser is being used and apply a style sheet accordingly. The default view for tablets is the list view. Delivering astronomy content, and educational content in general, for handheld devices is a good strategy since the market is booming; Gartner reports 968 million smartphones sold in 2013, and over 1.24 billion smartphones sold in 2014 (Gartner, 2015). The limitations of the small format display are unlikely to deter people from studying or from learning about science on the devices that have become so essential for everyday life.

Acknowledgements

We acknowledge Adrienne Gauthier, Jon Trump, Jared Gabor, Erika Offerdahl, Ingrid Daubar, Pamela Gay, and Jared Males, who have made valuable contributions over the past three or four years by editing Astropedia articles, and Bill Hartmann for his co-authorship of the textbook that that was the original source material for the original articles. Author acknowledges support from the Howard Hughes Medical Institute via grant 4215580 for developing online educational tools. We thank the University of Arizona for partial support of this project. Enthusiasm and cooperation of Bob Nemiroff (*Astronomy Picture of the Day*) and Nancy Wilkinson (*365 Days of Astronomy*) in sharing their resources is much appreciated. Author is grateful to his colleagues at the University of Arizona for testing the web site and giving useful feedback. Some of this work was carried while Author was in the productive setting of the Aspen Center for Physics and while on sabbatical in the Department of Astrophysical Sciences at the University of Princeton.

References

Alexa Internet, Inc. The top 500 sites on the web. (2015). http://alexa.com/topsites. Accessed 15.09.15.

Cook, J., Pachler, N., and Bachmair, B. (2011). Ubiquitous Mobility with Mobile Phones: A Cultural Ecology for Mobile Learning. *E-Learning and Digital Media, 8,* 181-186.

EDUCAUSE Center for Analysis and Research. National Study of Undergraduate Students and Information Technology. (2013). https://net.educause.edu/ir/library/pdf/ERS1302/ERS1302.pdf Accessed 14.06.20.

EDUCAUSE Center for Analysis and Research. Study of Undergraduate Students and Information Technology. (2014). http://net.educause.edu/ir/library/pdf/ss14/ERS1406.pdf Accessed 15.09.15.

Gartner, Inc. Gartner Says Worldwide Smartphone Sales Soared in Fourth quarter of 2011 with 47 Percent Growth. (2011). http://www.gartner.com/it/page.jsp?id=1924314 Accessed 13.07.30.

Gartner, Inc. Gartner Says Annual Smartphone Sales Surpassed Sales of Feature Phones for the First Time in 2013. (2014). http://www.gartner.com/newsroom/id/2665715 Accessed 14.06.19.

Gartner, Inc. (2016). Gartner Says Five of Top 10 Worldwide Mobile Phone Vendors Increased Sales in Second Quarter of 2016. Gartner Research. http://www.gartner.com/newsroom/id/3415117. Accessed 18.09.16.

Giles, J. (2005). Internet Encyclopedias Go Head to Head. *Nature, 438,* 900-901.

Author. (2012). Astronomy in the United States: Workforce Development and Public Engagement. In A. Heck (Ed.), *Organizations, People, and Strategies in Astronomy I* (pp. 77-91). Venngeist.

Author, Author, Author, Srinithan, A., Author, Ganesan, N.K., and Guvenen, B.C. (2013). A New Online Astronomy Resource for Education and Outreach. Concept and Content. *Astronomy Education Review,* 12(1).

Lih, A. (2009). *The Wikipedia Revolution: How a Bunch of Nobodies Created the World's Greatest Encyclopedia.* New York: Hyperion Books.

Pearson Education. Pearson Student Mobile Device Survey 2013. (2013). http://www.pearsoned.com/wp-content/uploads/Pearson-Student-Mobile-Device-Survey-2013-National-Report-on-Grades-4-to-12-public-release.pdf Accessed 14.06.06.

Purcell, K, Rainie, L., Rosenstiel, T., and Mitchell, A. How Mobile Devices are Changing Community Information Environments. (2011). http://www.pewinternet.org/2011/03/14/how-mobile-devices-are-changing-community-information-environments/ Accessed 14.06.06.

Saleh, A.A.M. (2011). Technology and Architecture to Enable the Explosive Growth of the Internet. *IEEE Communications Magazine, 49,* 126-132.

Shankar, K. (2012). A Review of Scientific Collaboration on the Internet. *The Information Society, 28(3),* 197-199.

Shin, D.-H., Shin, Y.-J., Choo, H., and Beom, K. (2011). Smartphones as Smart Pedagogical Tools: Implications for Smartphones as U-Learning Devices. *Computers in Human Behavior, 27(6),* 2207-2214.

The Wikimedia Foundation. Wikimedia Report Card. (2015). http://reportcard.wmflabs.org Accessed 15.09.15.

Winokur, D. Flash to Focus on PC Browsing and Mobile Apps; Abode to More Aggressively Contribute to HTML5. (2012). http://blogs.adobe.com/conversations/2011/11/flash-focus.html. Accessed 12.07.20.

Xiaojing, G. (2012). Data Integration for the World Wide Web. *Transaction Series on Engineering Science and Technologies, 1,* 142-144.