

TECHNIQUES

Topics in Communications

STC

Society for
Technical
Communication

Spring 2015

IN THIS ISSUE

2 Communicating with Assistive Technology

Paige Boyle

5 Visualizing Your Work: Tools for the Technical Communicator

Ines Wingert

8 Unethical Authorship: When Ghostwriting Crosses the Line

Michelle Tirado

11 The Technical Writer-Project Manager: Preparing for a Hybrid Position

Tiffany Jackson

13 Can I Help You ?

Katrina Myers

15 The Story Within Documentation

Tara Neita

17 Establishing Ethos Through Anonymity

Emma Furness

19 Technical Writer or Documentation Specialist — Which Job Title Best Reflects the Role of a Technical Writer?

Caitlin Wolanek

21 Blueprints to the Screen

Jeff Pellico

23 Entropy and the Complexity of XML Schema

Jack Sandberg

COMMUNICATING WITH ASSISTIVE TECHNOLOGY

By Paige Boyle

Computer technology dominates our world. Some think it is for the better, while others think the opposite. As with everything, its development has advantages and disadvantages. But for one group in particular—individuals with physical or mental impairments—computer technology has been beneficial and life-changing.

Computer technology has opened up a world of opportunities for people with disabilities—hearing, visual, motor, and cognitive disabilities, to name a few. For instance, individuals with motor disabilities—cerebral palsy, quadriplegia, arthritis, multiple sclerosis—communicate more effectively thanks to technological advancements, such as voice recognition software, adaptive keyboards, oversized trackball mice, and head wands. People with hearing disabilities communicate with sign language, captions, and transcripts. Individuals with visual disabilities communicate with Braille, screen readers, transcripts, and audio descriptions. People with cognitive disabilities communicate with a variety of technologies, including narrated multimedia, conceptual information, simplified text, and non-distracting layout. These types of technologies are commonly referred to as “assistive technology.”

In recent years, an increasing awareness has spread around the globe about different types of disabilities, culminating in what is known as universal design. With universal design, developers and designers strive to take all individuals into account when developing or designing a product or building. In terms of computer technology, companies such as Microsoft, Adobe, and Apple took universal design into account and designed their own assistive technologies for their products, providing more opportunities for a wide range of customers. Companies that build their own assistive technology into their products allow individuals to not only use the products more easily, but they also eliminate the stress of having to purchase a separate product.

Microsoft has paid attention to accessibility challenges for more than twenty years. Its first development, Access Pack for Windows NT, allowed people with motor and hearing disabilities to navigate the computer more easily. For instance, Access Pack gave people the ability to move the mouse with the keyboard. It also optimized the keyboard for those who use one hand or a wand to type. Nearly a decade later, Microsoft came out with Windows Vista, the first product to include the Ease of Access Center, which offers various accessibility features, including screen magnification, speech recognition, screen reader support, enhanced touch capabilities, and visual substitutions for sounds. (Windows 8 offers the Ease of Access Center in the sign-in screen so that people can have direct access to it.) With Microsoft’s screen magnification, people who are visually impaired can magnify the text so they can read it. Narrator, the company’s screen reader, helps people who are blind navigate their computers. Visual substitutions give people who are deaf, hard-of-hearing, or not close enough to their computers the ability to detect sound. Enhanced touch capabilities, such as touch screens, assist individuals with motor disabilities to select options that are traditionally activated by a mouse or keyboard. Speech recognition (i.e., voice recognition software) allows people with motor disabilities to navigate the computer by speaking commands into a microphone. Speech recognition is recommended for people with clear-sounding speech; however, individuals with shaky voices may find speech recognition frustrating.

Starting with Office 2010, Microsoft Word includes an add-in feature to save documents as DAISY XML files. DAISY XML files can then be converted to DAISY Talking Books, which can be read by several reading programs. PowerPoint 2010 introduced an add-in feature for closed captions in video and audio files. Also with Office 2010, Microsoft incorporated the Accessibility Checker feature, which ensures that Word documents, Excel spreadsheets,

In recent years, an increasing awareness has spread around the globe about different types of disabilities, culminating in what is known as universal design.

and PowerPoint presentations are accessible to people with disabilities. The checker scans documents and points out areas that may not be accessible to all individuals. This feature allows authors and document designers to fix their documents before publishing and distributing them.

Adobe is another company that embraces assistive technology. Reader and Acrobat are familiar accessible products. They allow people who use screen readers to access and read the text in PDF documents. However, document designers should pay more attention to accessibility when designing their PDF documents. For example, if document designers use figures in their documents, they should add alternative text. Alternative text allows people with cognitive and visual disabilities to read what the images are as well as what purpose they serve. It also helps people who simply want to read image descriptions rather than look at the image itself. If a browser does not load the image for one reason or another, then alternative text is useful for that purpose, too.

Another way to make PDF documents more accessible is by tagging the document. According to WebAim, “PDF tags provide a hidden structured, textual representation of the PDF content that is presented to screen readers” (“PDF Accessibility,” 1). To facilitate making content accessible, Acrobat X and XI contain two accessibility checks, Quick Check and Full Check. WebAim recommends doing the Full Check, as the Quick Check only ensures that a document has tags. It is important to note that tagging PDFs is just one step out of several to make content accessible. In addition to supporting screen readers, Adobe Reader supports magnified text, reading form fields, auto-scrolling, and saving in other formats (Word, HTML, RTF, and XML). Reader and Acrobat also use a text-to-speech tool called Read Out Loud, which reads an opened PDF document. Read Out Loud is not a screen reader; it is simply a reading tool strictly made for PDF documents.

Apple’s inclusion of assistive technology into its computers and iOS devices is perhaps the most notable. Among its features are VoiceOver, FaceTime, Siri, Dictation, and AssistiveTouch. VoiceOver is a screen reader that helps the blind or visually impaired to read and navigate their computers and iOS devices. Similar to Skype, FaceTime is a video-calling tool built into Apple’s computers and iOS devices. It was designed to help people who are deaf or hard-of-hearing communicate via sign language or other visual gestures. Siri and Dictation help people with both visual and motor disabilities—as well as people with no disabilities at all—navigate their iOS devices (iPhone, iPad, and iPod Touch). With Siri, users speak commands or ask questions into an iOS device, such as “Call Mom,” “Text Dad,” “Email John,” “Open my music folder,” “What’s the weather like today?” and “What is the 2015 MLB schedule?” Dictation permits a user to speak messages into an iOS device. For instance, a user who wants to text a friend the message “I’ll see you soon” just presses the microphone on the iOS keyboard and speaks the message.

Clearly, assistive technology has dramatically evolved over the last two decades. Individuals with disabilities now have more options than ever before to communicate and receive information. This article covers just a fraction of those options. The bottom line is this: The technology is out there, and it is time-saving, cost-effective, and, most importantly, it enhances the quality of life of people with—and without—physical or mental impairments.

References

“Accessibility for People with Disabilities.” Microsoft Corporation. 1997. http://techsys.vetmed.wsu.edu/gettingresults/APXA_1.HTM.

“Accessibility in Microsoft Office 2010.” Microsoft Corporation. 2015. <http://www.microsoft.com/enable/products/office2010/>.

Individuals with disabilities now have more options than ever before to communicate and receive information.

- “Adobe Acrobat Accessibility FAQ.” Adobe Systems Incorporated. 2015. <http://www.adobe.com/accessibility/products/acrobat/faq.html>.
- “History of Microsoft Commitment to Accessibility.” Microsoft Corporation. 2015. <http://www.microsoft.com/enable/microsoft/history.aspx>.
- “Motor Disabilities: Types of Motor Disabilities.” WebAim. Last modified Oct. 12, 2012. <http://webaim.org/articles/motor/motordisabilities>.
- “PDF Accessibility: Acrobat and Accessibility.” WebAim. Last modified Feb. 18, 2014. <http://webaim.org/techniques/acrobat/acrobat>.
- “The Access Pack for Microsoft Windows NT.” Microsoft Corporation. 1997. http://techsys.vetmed.wsu.edu/gettingresults/apxa_4.htm.
- “Types of Assistive Technology Products.” Microsoft Corporation. 2015. <http://www.microsoft.com/enable/at/types.aspx>.

VISUALIZING YOUR WORK: TOOLS FOR THE TECHNICAL COMMUNICATOR

By Ines Wingert

Technical communicators are trained to think carefully about the visual elements of their work: layout, fonts, headers, and graphics. Over time, your skill with these concepts grows; however, you may overlook the importance of how you visualize documents, much to the detriment of your potential productivity.

This article covers common tools that technical communicators can embrace to better visualize their documents in all stages of production—from computer monitors and desk arrangements to Post-it notes and index cards. By optimizing your digital and physical workspaces, you can get more done and make fewer mistakes.

TECHNOLOGY

Widescreen Computer Monitors

Today's large and high-resolution computer monitors can be a great boon to technical professionals; a single purchase and quick installation can bring years of added productivity. “High Definition” (HD) or “1080p” widescreen monitors have a resolution of 1920 x 1080 pixels and come in generous yet affordable sizes: Many 24” models cost less than \$200. HD monitors offer the technical communicator more possibilities than smaller screens, such as the ability to display two documents or pages side-by-side in a fully legible size. This allows layout specialists to view both the recto and verso sides of publications; editors can view a marked-up copy alongside a new working draft. The time and error savings can be amazing.



Figure 1 - Widescreen HD Monitor

Those who are willing to pay more money to be on the cutting edge of technology can realize even more of the benefits of size and resolution. You can view your entire “tool chain” of applications on the screen at once with 27” and even 30” displays with nearly twice HD resolution (WQHD, at 2560 x 1440 pixels) or the newest “4K” screens (at a whopping 3840 x 2160 pixels), virtually eliminating the time spent switching between windows and applications. When designing a manual, for example, you could browse an entire image library of product graphics and drag them directly into the proper places in a large document, all in one continuous motion.

Multi-Monitor Configurations

Given the low prices of today's computer monitors and the clear benefits of having a large workspace, many professionals are discovering the benefits of using two HD monitors simultaneously. Most desktop PCs support this with little hassle, and you can even choose a variety of monitor configurations; most prefer a side-by-side horizontal layout, though others prefer a horizontally-stacked configuration or a hybrid side-by-side orientation that allows for one side to be a “desktop and applications” monitor and the other to be a vertically-oriented, dedicated “document layout” monitor.

In addition to the vast expanses of workspace that dual monitors offer, there are other benefits as well. Dual monitors can be configured to act as one continuous workspace or as two separate ones. For example, clicking “maximize” on an Excel spreadsheet on the right-hand screen would expand it to fit perfectly within the bounds of that monitor, while all other activities on the left monitor would continue normally. Some technical communicators who specialize in instructional design use this feature to view video demonstrations on one monitor while producing procedural documents from the video on the other monitor.



Figure 2 - Horizontal Dual-Monitor Configuration (left) and Hybrid Dual-Monitor Configuration (right)

Mind-Mapping Software

If computer monitors help to increase and enhance your digital workspace, “mind-mapping” tools are used to increase your ability to visualize and represent ideas. Software tools like Mindjet and Microsoft Visio enable you to quickly release often-vague ideas from your mind and bring them into the reality of the computer screen. You can then easily manipulate and expand on them by creating shapes, directional arrows, common objects (computers, furniture, documents), and popular chart types by simply dragging and dropping. Mind-mapping software gives you the power to model and share ideas too complex for even the world’s greatest geniuses.

For technical communicators who often work with information that must have a logical structure, mind-mapping software may become an essential part of the planning and design phases of the production cycle. Perhaps the most tangible benefit for technical communicators is being able to send stakeholders a quick visual representation of how a document or information structure would look; this avoids the common situation of having to throw out hours of work if a full draft turns out to be poorly received.

SIMPLE OFFICE TOOLS

Some of the most effective tools for visualizing your work are sitting in your home or office right now. Note pads and white boards may come to mind, but more advanced and lesser-known visualization and planning strategies like card sorting and Post-it planning offer unique benefits, too.

Card Sorting

Packs of three-by-five index cards can be a technical communicator’s best friend. Each card is large enough to hold a useful amount of information (on either one or both sides) yet is small enough to be sorted, arranged with other cards, or spread out over a large area to develop or reveal structure and patterns.

Card sorting sounds simple, but people have used it to plan billion-dollar projects and solve complex crimes. You can use it to force information into new structures, such as hierarchical rankings, categories, or flows. Web designers, for example, may use card sorting to develop navigation menus; document designers may use index cards to plan document sections in layers of headers and sub-headers.

Use card sorting to help you plan and visualize structure, sort and organize ideas, and reveal unique connections between otherwise chaotic batches of information.

Post-it Planning

Post-it notes are self-adhesive, square pieces of paper that come in a variety of colors. They are similar to index cards, yet you can easily reposition them; their ability to stay affixed to a surface makes them uniquely excellent for long-term iterative planning.

If your document or information architecture is in a state of change or planning, use Post-its to represent ideas like buttons, document sections, or concepts; then move them around until you are satisfied with the result or find a hidden connection. You will find that the benefits of Post-it planning will “stick” with you.

CONCLUSION

No matter their specialty, technical communicators fundamentally deal with information. By adopting tools and processes that allow you to better visualize that information, either by offering a wider perspective or by making it easier to experiment with and examine the objects within it, you can magnify all of your existing professional skills, get work done faster and with greater quality, even see entirely new possibilities.

UNETHICAL AUTHORSHIP: WHEN GHOSTWRITING CROSSES THE LINE

By Michelle Tirado

Most writers will consider ghostwriting at least once during their careers. Many may even go on to earn a decent living by writing articles and books that they will never receive credit for. However, just as there are good and bad ghosts, there are good and bad ghostwriting opportunities. Separating the two is the sometimes blurry yet ever present ethical line.

In the non-technical genres, ghostwriting is generally perceived as an acceptable practice. People have stories and messages to share—whether they are novels, memoirs, or magazine articles—but they may lack the time or the writing skills needed to put their stories into publishable words. Thus, they turn to freelance writers or employees to pen these manuscripts for them. Often, there is little to no effort to hide the real author of a work; in fact, sometimes the name of the ghostwriter is made public by the writer or the person who paid the writer. Even when the ghost goes undetected, what harm comes of it? The ghostwriter receives compensation, and the reader enjoys a well-organized and grammatically sound composition that is based upon information from the person who got the byline.

But in scientific fields, where it is often referred to as ghost authorship or honorary authorship, ghostwriting is more ethically spooky. And it is probably at its spookiest and most pervasive with medical journals. A common scenario goes something like this: A pharmaceutical company with a new product retains a medical communications firm to create a publications plan, which involves identifying target journals; the communications firm assigns a freelance writer to draft a paper from data produced by a scientist who works at the pharmaceutical company; the communications firm hires an academic—preferably a key opinion leader—to lightly edit the paper; the academic places his or her name on the paper as an author and then submits it to a target journal; the paper goes through the peer-review process and is published (Moffatt 2013).

One of the most high-profile ghost authorship cases involves Merck's Vioxx, a painkiller that the pharmaceutical company marketed and sold for medical uses other than what it was approved for by the FDA (Wilson 2011). In 2004, Merck pulled the product from the market after evidence emerged indicating the drug doubled the risks of heart attack, stroke, and death (Wilson 2011). An investigation that followed revealed the company, with the help of ghostwriters, produced dozens of research studies on Vioxx, some based on clinical trials, to bolster the benefits of the drug in the treatment of other conditions and then "lined up prestigious doctors to put their names on the reports before publication" in medical journals such as *JAMA* and *Neuropsychopharmacology* (Saul 2008). Before Merck recalled Vioxx, it had been prescribed to some 25 million Americans; and after the recall, approximately 27,000 people harmed by the drug filed lawsuits against the company, which it settled for \$4.85 billion in 2007 (Wilson 2011).

According to Moffatt and Elliott, what makes ghostwritten medical articles unethical and even dangerous are the undisclosed conflicts of interest, such as the payment that an academic receives for the honorary authorship and the profit-driven pharmaceutical company that writes the check to the academic, and the harm these articles can cause to public health if they mislead physicians about the risks and benefits of drugs (2007). Then there are the effects on the collective knowledge within the medical community, which journal articles help build and maintain (Solomon 2007). As Sismondo succinctly puts it, "Ghost writing and honorary authorship are not in and of themselves scientific problems, though they become so when they shape science to meet particular interests" (2007, 1429).

That ethical line regarding authorship has been drawn and spelled out in the codes of conduct

and author guidelines of professional associations and journal publishing companies. More than 600 medical journals have adopted the International Committee of Medical Journal Editors' criterion for authorship, which is as follows:

- Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; and
- Drafting the work or revising it critically for important intellectual content; and
- Final approval of the version to be published; and
- Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Why, despite these ethical codes, does ghostwriting continue to plague the scientific community? While some specifically blame the codes and guidelines – that they do not go far enough – and a lack of enforcement (Moffatt 2013), others point their fingers at society's overall perception of the practice. Although ghostwriting may raise an eyebrow or two in mainstream publishing, it is generally allowed. Undoubtedly, this acceptability has made inroads in the scientific realm. On its proliferation in medical journals, Bosch and Ross write, "Ghostwriting flourishes because it is perceived as a slight, easily comprehensible moral failing, rather than as unethical" (2012, 324).

"Ghostwriting flourishes because it is perceived as a slight, easily comprehensible moral failing, rather than as unethical."

The writers who take on scientific ghostwriting assignments are not blameless; therefore, they must also bear some of the responsibility in curbing the practice. It may be difficult to turn down a medical article-writing gig that pays \$90 to \$120 per hour (Moffat and Elliott 2007) or a potential job promotion, but writers should consider the potential ethical implications of their work. They must be aware of professional codes of ethics in their professional community as well as in the scientific community that they will be serving. When in doubt, they need to turn on their moral compass and ask themselves: "Who might get hurt?"

References

- Bosch, Xavier, MD, Ph.D., and Joseph S. Ross, MD, MHS. "Ghostwriting: Research Misconduct, Plagiarism, or Fool's Gold?" *American Journal of Medicine* 125, no. 4 (2012): 324-6.
- "Defining the Role of Authors and Contributors." International Committee of Medical Journal Editors. <http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html>.
- Moffatt, Barton and Carl Elliott. "Ghost Marketing: pharmaceutical companies and ghostwritten journal articles." *Perspectives in Biology and Medicine* 50, no. 1 (2007):18-31.
- Moffatt, Barton, Ph.D. "Orphan Papers and Ghostwriting: The Case against the ICMJE Criterion of Authorship." *Accountability in Research: Policies and Quality Assurance* 20, no. 2 (2013): 59-71.
- Saul, Stephanie. "Merck used ghostwriters and misrepresented data on Vioxx, article says." *The New York Times*, Apr. 15, 2008. <http://www.nytimes.com/2008/04/15/business/worldbusiness/15iht-15merck.12018748.html?pagewanted=all>.
- Sismondo, S. "Ghost Management: How Much of the Medical Literature Is Shaped Behind the Scenes by the Pharmaceutical Industry?" *PLoS Med* 4, no. 9 (2007): e286. doi:10.1371/journal.pmed.0040286.

Solomon, David J. "The role of peer review for scholarly journals in the information age." *Journal of Electronic Publishing* 10, no. 1 (2007): 32.

Wilson, Duff. "Merck to Pay \$950 Million Over Vioxx." *The New York Times*, Nov. 23, 2011. <http://www.nytimes.com/2011/11/23/business/merck-agrees-to-pay-950-million-in-vioxx-case.html>.

THE TECHNICAL WRITER-PROJECT MANAGER: PREPARING FOR A HYBRID POSITION

By Tiffany Jackson

The ever-changing role of the technical writer has expanded to include project management. Proof is in the growing demand for job candidates who have an education or experience in both technical writing and project management. Whether it is through university degree programs, professional associations, or on-the-job training, there are opportunities for technical writers to gain the knowledge and skills required to fulfill these new duties.

Technical communication and project management are more alike than different. Both positions entail audience analysis, persuading readers with logic, and conveying clear messages. More than likely, technical communication students will manage projects in the workplace or serve as a project manager in some capacity, such as managing their own documentation projects.

Project management involves organizing all aspects of a project to ensure each component is completed for the purpose of the project's main goal, which is the implementation of an initiative. This requires frequent communication with all business owners and subject-matter experts in the form of meetings, phone calls, and e-mail correspondence. The project manager must obtain status reports from impacted business areas, track updates on project plans, and distribute status reports weekly to project team members and others involved in the project. Planning is a critical component in executing a successful project, and meeting deadlines is necessary to move to the next phase. A technical writer uses a similar methodology to complete structured documentation.

The role of the project manager has evolved over the past decade from its previous administrative category. The project manager is a communicator because the role may involve taking complex details and processes and condensing them verbally to project team members. The project manager essentially gathers all of the impacted business areas' opinions, feedback, and concerns during various stages of the project and works to understand the organization's goals.

How equipped are technical writers to take on project management duties? In one survey (Whiteside 2003), conducted more than a decade ago, 33 percent of newly employed technical communication graduate students reported that their project management skills needed to be enhanced. This is largely due to assuming the ownership of a project and not being familiar with project plan management, scheduling, and deployment dependencies. But perhaps today's technical communication students who are entering the workforce are better prepared for project management functions.

Several schools now offer technical communication project management courses, including the University of California (San Diego and Riverside), East Carolina University, and even here at Minnesota State University, Mankato. The Project Management Institute, an international association for project management professionals, provides various project management training programs and continuing education courses. PMI also offers several certification programs, such as its Certified Associate in Project Management certification, which allows entry-level project managers to demonstrate their knowledge of the fundamentals of project management. The certification is available to individuals who have a high school diploma and 1,500 hours of project management experience or twenty-three hours of project management coursework at the time of the exam.

If a technical writer is unable to enroll in a project management course or complete a

How equipped are technical writers to take on project management duties?

certification program, on-the-job training will suffice and may arguably be the best teacher. Technical writers are often tasked with owning content management systems and other repositories for policies and procedures accessed by staff. The ownership of these systems is a natural progression for the technical writer, as he or she may write the content displayed within it, upload content to the content management system, and communicate related updates to staff. This coordination involves project management and may require technical writers to work with IT and other business partners.

Technical writers that possess project management skills are in high demand. Currently, there are over 1,500 job announcements on reputable job websites from employers seeking candidates for technical writer-project management hybrid positions with titles such as Project Management-Technical Writer, Project Management Technical Writer, Technical Writer Project Manager, and Writer and Project Manager. Requirements include the ability to author technical content in content management systems; writing business requirements, specifications, and training materials; and guiding project and program management activities. These activities span managing schedules, budgets, risks, changes, opportunities, and resources to maintaining comprehensive project plans.

Flexibility is key for new and experienced technical communication professionals interested in these types of hybrid positions. Technical writers must be open to functioning in more than one capacity and understand that not every assignment will be the same.

References

Brandel, Mary. "The New Project Manager." *Computerworld* 40, no. 15 (2006):43, 46-47.

Hart-Davidson, William. "On Writing, Technical Communication, and Information Technology: The Core Competencies of Technical Communication." *Technical Communication* 17, no. 3 (2001):145-148.

Whiteside, Aimee. "The skills that technical communicators need: An investigation of technical communication graduates, managers, and curricula." *Journal of Technical Writing and Communication* 33, no. 4 (2003): 303-31.

CAN I HELP YOU?

By Katrina Myers

When I started writing help files, I worked with marketing teams to assess the demographics of our end users. We knew the age range, socioeconomic status, and the percentage of men vs. the percentage of women who composed our customer base. I learned how to write task-based topics and I talked with end users to identify the best phrases for useful indexing. I followed the “three click” rule and researched the best format for fast content delivery. I thought I was creating content that end users would find truly helpful. Imagine my surprise when I found that many end users do not even know what application help is.

I wanted to know more about how people approached application help. I decided to initiate my own small-scale survey. I chose six individuals, each with a different computer-use background, and I talked with them about application help.

| Name/Age | Career | Software Used | Do you use application help? |
|------------|------------------------|--|---|
| Brad, 32 | Military | Microsoft Office and proprietary military programs | “Yes, five percent of the time; but mostly I ask a coworker or use a search engine to find help.” |
| Bernie, 59 | Underwriter, Insurance | Microsoft Office, social media, and proprietary insurance programs | “You mean the chat window? Oh, you mean the help desk guy who fixes my computer?” |
| Olivia, 28 | Online Comic Author | Microsoft Office, social media, and graphic tools such as GIMP | “Are you kidding? No, I ask questions in online forums.” |
| Kerry, 44 | Massage Therapist | Microsoft Office, social media, and website design | “No, I ask my kids.” |
| Emily, 22 | College Student | Microsoft Office and social media | “Sometimes I try, but I get frustrated and ask my friends instead.” |
| Keegan, 26 | Computer Programmer | Primarily open source/ LINUX user and some online gaming | “I use the command line to access help files or I use websites and forums.” |

The general consensus is that help is not helpful. With all of the guidance already provided in books, classes, and seminars regarding how to write good help, I have to wonder, “Why?”

Almost everyone I interviewed turns to coworkers, family, or search engines for help; no one uses the application help regularly, if at all. When I asked for more specifics, I heard exactly what I expected: “The help does not answer my question most of the time, or it assumes I know things that I don’t know and I have to click multiple topics before I can understand the answer, or I get so lost in hyperlinks that I give up.”

As a technical writer, I wondered what I could do. I cannot anticipate every question an end user might have. Yet, search engines provide these answers much of the time. That means the answers are there, and I should be able to provide them. Maybe application help should be

more like a search engine? Wait, I may have something. I thought about some of the interview comments and started to form some “real-world” suggestions.

Olivia told me that she gets confused about the strangely-worded answers in application help. Technical writers know the importance of analyzing their audience, but demographics are not enough. You need to observe and interact with your audience and understand the types of questions and vernacular that your audience uses. I fear that many writers have allowed themselves to focus on single-sourcing standards, blinding themselves to what end users need to relate to material. One reason people prefer search engine answers is that the answers are written in simple terms. This led me to thoughts expressed in another interview.

Keegan, a computer programmer, mentioned that search engines use heuristic learning models. In layman’s terms, that means search engines track the words used in a query and the successful results. The search engine continues to track these queries and results to learn how and what information to present. How can you do that in help files? I suggest communicating with your help desk. You should have access to bug-tracking tools; if not, you can talk with help desk representatives. Many online help source tools include usage tracking and feedback options. Are you using these tools? All of these options allow you to identify not just the questions end users need answered, but also the insight to the phrases you need to use in your topic titles. Help is not useful if it becomes stale. Writers need to mimic search engines; you need to learn from your end users and update continuously.

Emily and Bernie both offered another important tip during their interviews. Even if you create wonderful help, previous experience has taught your end user not to bother. As my informal survey shows, most computer users skip the application help and go straight to a coworker or a search engine. If you want people to see your application help, ensure that customer training sessions include instruction for using help files and encourage trainers to use the help during class. Also, encourage help desk representatives to guide end users through the process of finding the answer within the help files. Customers will be pleased to learn that they can find answers for themselves without having to wait on hold, and the business profits as the number of help desk calls decreases.

So, what did my informal research teach me?

1. Build and test your help file content on real end users, not just demographic representations.
2. Use tools to gather end user comments and FAQs (Frequently Asked Questions).
3. Use information gathered from end users to constantly improve and revise your application help.
4. Work with company training and service departments to promote the use of the application help.

Writers will never be able to predict every question an end user may pose, but it is important to understand that help file maintenance is not restricted to feature updates. Writers need to make a point to continue monitoring and refreshing application help files, because good application help can improve the quality and sales of a product while raising the level of customer satisfaction.

THE STORY WITHIN DOCUMENTATION

By Tara Neita

For quite some time, there has been a great divide between creative writing and technical communication. In the academic realm, we typically pursue either one track or the other, not both. Looking further, however, we can argue that both writing styles benefit each other. More specifically, creative writing can provide a unique aspect to technical documentation in regards to audience and style.

To compare, technical writing gives rise to our logical elements. Its purpose is to provide information and persuade the reader to act in a detailed, yet concise and clear manner. Structurally, technical writing is very rigid with strict guidelines such as document specifications, format, and style.

Creative writing, on the other hand, brings forth our imaginative elements, where the reader focuses on the effectiveness of prose. Creative writing's goal is to promote full immersion across multiple levels. Its complex and often ambiguous text is meant to entertain its readers and evoke multiple meanings.

Although they have their differences, technical and creative writing require a great amount of skill and talent. Each written word is given a great deal of thought. Notably, both technical and creative writing have an identifiable audience and a clearly defined purpose. It is a matter of interpreting audience and purpose that differentiates the two writing styles. Arguably, applying aspects of creative writing could change the way we see technical writing. In this case, the generation of a story could benefit technical documentations. We could make technical documentations more engaging for the user by making them more personalized, yet still generic enough for the unique aspects that make technical writing what it is.

AUDIENCE

As a creative author, I give much thought about the characters I develop. I ask, "What are my character's thoughts, actions, problems, questions, and hesitations, and how will they affect the progression of the story?" One question writers often need to decipher is, "If something happens, how will the character overcome it?" We can apply the same strategy for storytelling development to writing technical documentations.

As the writer, it is best to view the world from the eyes of the character—or in this case, the end user—and anticipate those needs. In the technical writing domain, the writer needs to sift through copious amounts of facts, figures, statistics, and background information gathered through intensive research. The writer then needs to develop what would become the important pieces. First, the writer must clearly understand who the intended audience is in order to determine what is considered important.

Once the writer discovers the audience, the story can begin.

NARRATIVE STYLE

Storytelling is the essence of creative writing. The story guides the reader through a journey while developing the character and witnessing the successes and failures of the protagonist's struggles. The same works for technical writing. The exception is that the struggle occurs only within an instant of time. The user needs help to overcome an issue (to learn how to do a task, be persuaded to act, etc.), and it is up to the guiding words of the technical writer to get through it.

Applying aspects of creative writing could change the way we see technical writing.

Style is a topic of discourse in any writing community. Style in creative writing is all about the way the prose is written—the way the fluidity of ink on paper permeates the audience. Creative writers pay special attention to word choice, tone, and rhetorical devices so that the author’s personality can resonate throughout the piece. In technical writing, style connotes the way information is conveyed. It needs to be universal and get the point across without sounding monotone.

The application of a narrative style can guide the technical writer to choose and appropriately arrange words to help the user through a process. To start, the writer asks, “How can I convey all of this information in a manner easily understood by my readers?” Simply put, there is a problem and a solution, and the writer’s mental roadmap of directions is the only aid for the user.

The writer can draft a sequence of steps, or a series of events within a story, that will get the user one step closer to his or her goal. Effective diction can adequately present information that may have otherwise left readers confused.

Creativity and technicality infiltrate each other’s worlds.

THE BALANCE

Creativity and technicality infiltrate each other’s worlds. The traditional approach may need unconventional flare. The chaotic, anarchic web of information may need the structure and discipline of a prominent cornerstone. The fresh perspective of the audience and the application of a new narrative style create a mutual benefit. The creativity learns from the technical, and the technical develops the story within documentation.

REFERENCES

- “Can Technical Writing Be Creative?” Writing Assistance, Inc RSS. June 30, 2011.
 Accessed April 3, 2015. <http://www.writingassist.com/newsroom/can-technical-writing-be-creative/>.
- “Creative, Business and Technical Writing.” Write Wise Communications, LLC.
 Accessed April 3, 2015. <http://www.writewisecommunications.com/courses.html>.
- Patel, Idrees. “Creative Writing vs. Technical Writing.” Writers Treasure. Accessed April 3, 2015. <http://www.writingassist.com/newsroom/can-technical-writing-be-creative/>.

ESTABLISHING ETHOS THROUGH ANONYMITY

By Emma Furness

Electronic word-of-mouth (eWOM) marketing is a business tool to encourage consumers to purchase goods and services. These are the reviews on product web pages or message boards exclaiming how wonderful a product is or complaining about how poor a product is. Sites like Yelp are dedicated to eWOM. Familiarized with word-of-mouth (WOM) marketing, that is product recommendations by family and friends, we comfortably rely on the WOM of other online users (Lopez and Sicilia 2014, 29). We trust that reviewers have the highest stakes in honesty with little evidence. But unlike our close, real-life social circles, we can't know who offers us recommendations online. As technical writers and business writers, we attempt to persuade our audiences while remaining largely anonymous on behalf of a company or organization. At the same time, we also make up an audience of online consumers who research product and service recommendations. How can individuals and companies invoke ethos while remaining anonymous, or even while misrepresenting themselves?

The good news for the writer is that it can be done—and quite easily, at that. Through a phenomenon called truth bias, people are convinced by online reviews by purportedly real-life people (Ott, Choi, Cardie, and Hancock 2011, 313). Truth bias is the tendency for people to believe information that is presented to them, even when the source of that information is unfamiliar. A study at Cornell University tested truth bias in online reviews of hotels. Some reviews were real and some were false, but all were created for the study (Ott, Choi, Cardie, and Hancock 2011). A computer was programmed to detect language that might indicate deception. Human subjects competing to determine whether the reviews were true or fake were confident they would do better than the computer; however, all of the humans performed considerably worse than the computer program.

The bad news for the business writer is that consumers also suffer from negativity bias, or the tendency of negative information to have more impact on decisions than positive information (Chen and Lurie 2013, 464). Because people are more motivated to shield themselves from pain and disappointment than to achieve satisfaction, negative reviews still have greater effect than positive reviews (Park and Lee 2009, 65). However, a recent study found that simply adding temporal verbiage in the positive eWOM message can make it more powerful than a negative eWOM message (Chen and Lurie 2013, 463).

The bad news for the consumer is that (1) we tend to believe things rather than disbelieve them, even when there is no great reason to do so; (2) we tend to overestimate how accurate we are when detecting deceptive language; and (3) we can be tricked by a few strategic keywords into believing something more strongly than we would have otherwise, with no real basis for doing so. Furthermore, when we imagine someone else's experiences, we view them as more strongly positive or negative than when we experience it ourselves (Monga, Chen, Tsiros, and Srivastava 2011, 32). Keeping all of this in mind, the eWOM sender is most likely a complete stranger, with unclear motivations for publishing the review, whether positive or negative. We are used to traditional WOM, and it's reasonable to trust a sister when she recommends a particular cellular service provider or a friend when he recommends a certain grocery product. Is it reasonable to trust a complete stranger who offers a recommendation?

The good news for the consumer is that this phenomenon isn't new, even if the mode is modern. Historically, many authors and writers were anonymous or deliberately misrepresented their identity. Many proclaimed doctors and inventors have pushed products and services successfully before being called charlatans and quacks, if ever they were. For most of history, knowing much about anyone who was not directly part of one's community was difficult. While the ethics of

Through a phenomenon called truth bias, people are convinced by online reviews by purportedly real-life people.

misleading eWOM marketing is certainly questionable, it isn't new. We can't know with certainty which online reviews and recommendations are sincere, but we can make educated guesses based on the information we're given, and we can be conscious of our biases as consumers.

References

- Chen, Zoey and Nicholas H. Lurie. 2013. "Temporal Contiguity and Negativity Bias in the Impact of Online Word of Mouth." *Journal of Marketing Research* L: 463-76.
- Lopez, Manuela and Maria Sicilia. 2014. "Determinants of E-WOM Influence: The Role of Consumers' Internet Experience." *Journal of Theoretical and Applied Electronic Commerce Research* 9(1): 28-43.
- Monga, Ashwani, Haipeng Chen, Michael Tsiros, and Mona Srivastava. 2011. "How Buyers Forecast: Buyer-Seller Relationship as a Boundary Condition of the Impact Bias." *Springer Science + Business Media* 23: 31-45.
- Ott, Myle, Yejin Choi, Claire Cardie, and Jeffrey T. Hancock. 2011. "Finding Deceptive Opinion Spam by Any Stretch of the Imagination." *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*: 309-319.
- Park, Cheol, and Thae Min Lee. 2009. "Information Direction, Website Reputation and eWOM Effect: A Moderating Role of Product Type." *Journal of Business Research* 62: 61-67.

TECHNICAL WRITER OR DOCUMENTATION SPECIALIST – WHICH JOB TITLE BEST REFLECTS THE ROLE OF A TECHNICAL WRITER?

By Caitlin F. Wolanek

As a student in the field of technical communication, I often wonder what the official title of my future dream job will be: “technical writer” or “documentation specialist?” While technical communication is a vast field, encompassing many different areas of information and content formats, the roles of each profession overlap and sometimes parallel each other.

While the titles of “technical writer” and “documentation specialist” sound like different career paths, the roles of each of these positions are similar. Both focus on making technical information more accessible for users and both meet the needs and advance the goals of an organization. This article presents an explanation of which job title most accurately reflects the duties involved in carrying out the tasks of a technical writer.

The most common job title for writing technical content is “technical writer,” but does it accurately reflect the range of learned capabilities and required duties that a technical writer is expected to possess and accomplish? Think about it from a descriptive viewpoint. The word “technical” denotes the image of complicated information created for skilled readers; and the word “writer” restricts technical writers’ capabilities to simply writing. While technical writers compose information for specific, highly skilled specialists and professionals, that is not all they do.

Accentuating other learned skills that accompany writing provides the most valuable assets to a technical writer’s professional image. One could display skills such as a thorough knowledge of several software programs, research skills, collaboration experience, product management experience, and experience with graphic design and document design. Although many technical writers could fulfill roles other than being a writer, they often are not given the chance because these roles require skills that are considered to be outside of the sphere of technical writers. Organizations often do not expand the responsibilities of technical writers beyond the scope of writing because the “technical writer” job title limits the scope of their perceived abilities to writing.

Rather than use the title of “technical writer,” I would argue the title of “documentation specialist” offers a more encompassing description because it reflects the wide range of abilities and talents technical writers have learned through their extensive education. The title of “documentation specialist” is used to describe professionals who also work with technical content, subject-matter experts, and various software programs. Generally, a documentation specialist is considered a technical professional with a wide range of responsibilities within an organization. The primary roles of documentation specialists are research, design, documentation, and the development of new types of products as well as the analysis of competitive products. Documentation specialists draft and edit documents containing technical information and coordinate and manage projects. Although it sounds a lot like a technical writer, the documentation specialist’s role encompasses more abilities, giving credit to technical writers because they really do so much more than just write technical material.

In summary, technical writers translate complex content into a format users can understand; therefore, they bridge the gap between the experts and the audience. However, developing technical material requires research and collaboration skills; proficiency in a number of software programs; the ability to take on the roles of a document designer or a photographer; and the ability to make educated decisions on the organization, style, word choice and grammar, and visuals of a document.

Accentuating other learned skills that accompany writing provides the most valuable assets to a technical writer’s professional image.

Understanding and applying knowledge of technical communication and being the user's advocate are critical to both job titles. Overall, technical writers accomplish more than just putting pen to paper—or fingers to keyboard. They specialize in much more than putting technical jargon into plain language or editing. While the term “technical writer” seems to have limited the skills many technical writers have spent years to learn and apply, the “documentation specialist” job title is a start in reflecting the range of abilities necessary to be highly proficient in the diverse profession of technical writing.

BLUEPRINTS TO THE SCREEN

By Jeff Pellico

A screenplay must be adaptable.

It exists purely as the precursor to films, shows, and various video productions to transcend stories to an audiovisual level. While screenplays may tell stories, they are more concerned with the telling than the stories. That is where the screenplay diverges from most creative literature; the screenplay's primary function is to be adapted. By design, a script takes on a form similar to a user manual, becoming documentation used to inform and instruct the design of its eventual audiovisual counterpart—i.e. the motion picture.

As with any technical document, communication is king. By breaking down a short excerpt from Graham Moore's *The Imitation Game*, we can see how communicative design has been built into the identity of the screenplay in order to ease the barrier between translating from script to screen.

BLACK.

ALAN TURING (V.O.)
Are you paying attention?

INT. ALAN TURING'S HOUSE - DAY - 1951

Within the opening four lines, we are met with a series of curiosities: scattered alignments, technical jargon, and an ugly typeface, each with a purpose.

Although generally unused by most writers, the unattractive Courier fonts find shelter in screenwriting because of their fix-pitched spacing between each character. Committal to 12-point Courier allows each page in a script to equate to around one minute of runtime in its filmic counterpart. The reader can estimate a total runtime by quickly glancing at the total page count.

Aligned right, capitalized, and taking the spot of where screenplays often begin with a "FADE IN," the author instructs the film to begin with a black screen. The right alignment indicates that this is a transition, or the lack of one. Notice that there are no further instructions on camera transitions; instead, the author relies on a "slug line" to imply the transition to a new scene.

The slug line—seen bolded at the bottom and aligned left—indicates the location of a scene. With new locations comes an implicit cut from one image to another. If a screenplay consists of hundreds of scenes, explicitly writing a transition such as "CUT" before every new scene would total hundreds of lines that provide no worthwhile information to the reader.

INT. ALAN TURING'S HOUSE - DAY - 1951

A HALF-DOZEN POLICE OFFICERS swarm the Manchester home of mathematics professor Alan Turing.

ALAN TURING (V.O.)
Good. This is going to go very quickly now. If you are not listening carefully, you will miss things. Important things. You're writing some of this down? That's good.

INSIDE ALAN'S HOUSE: There's been a break-in, and the house is a mess - someone has given it a pretty thorough once-over.

Building on a theme of brevity, Moore keeps his initial description (or action lines) down to a sentence. On a visual level, the opening line offers no indication that the setting takes place in 1951; rather, by introducing the date in the slug line, Moore avoids using a separate line for the direction "SUPERIMPOSE: 1951," saving himself some space. While it may seem insignificant to

save one line, consider that the screenplay already runs for 118 pages and ran much longer before editing. If a page amounts to a minute, and time is money, then a page saved is a penny earned.

Returning to Moore's opening action lines, a single, brief sentence also keeps the scene moving. Moore opts for one sentence instead of a staccato of fragmented lines, using both the brief descriptions and the dialogue (presented in a voice over) to build a slow burning urgency in the scene. As opposed to opening on a still image, the writing instructs a quiet start that shoots out into hectic sequences.

Jumping further in, Moore switches to an active voice: Scenes quickly cut between one another, emphasizing the actions taking place. The sharp transitions indicated by a hyphen—starting and ending each sequence—use a phone call to bind the cuts into a singular idea, which keeps the reader from becoming overwhelmed or lost in transit.

```
A CONSTABLE PHONES IN the robbery to police headquarters -

INT. POLICE HEADQUARTERS - DAY

- At headquarters, a RADIO GIRL RECEIVES the information and
passes it to an assistant for delivery to the detectives on
duty -

INT. MI-6 - RADIO OPERATORS' ROOM/HALLWAYS - DAY

- While in London, a RADIO OPERATOR in a dark room far below
Victoria Street INTERCEPTS THE MESSAGE -
```

Separate locations are chunked apart and given their own identity, acting as sign posts and orienting the readers. These sign posts allow readers to instantly recognize that they are snapping between scenes. The audience, experiencing the audiovisual feedback from the film, would not struggle with understanding the switch between scenes. The readers, however, are given only touches of imagery so that the story can keep moving, leaving them to visualize on their own. They are riding a roller coaster, not a tour bus.

The experiences of the readers and a film's audience are difficult to compare, as both groups act off of different stimuli. Still, the design set out by the screenplay must go beyond strictly capturing the audiovisual elements at play. The goal is to feign style as close to the filmic experience without explicitly outlining how a film should be captured.

The Imitation Game, directed by Morten Tyldum, stays in line with the design set out in Moore's screenplay. The film's runtime comes in four minutes shy of the script, which is understandable because extensive dialogue tends to dislocate the one-page-a-minute rule of thumb. More so, the runtime indicates the ease of adaptation that both pieces could align so close together. In all measures, Moore was able to use the systems of communicative design to construct a comprehensive manual for building a film based on the life of Alan Turing.

For an even further dissection of how screenwriting techniques communicate their design, watch a film and afterwards flip open its script to see how the production fared. Try to look for all the quiet systems in place, working away to communicate not just a story, but a way for that story to be told.

References

Moore, Graham. The Imitation Game. Film Script. 2014. The Weinstein Company. <http://twcguilds.com/assets/downloads/ScreenplayTIG.pdf>.

ENTROPY AND THE COMPLEXITY OF XML SCHEMA

By Jack Sandberg

When we single-source information in an XML file, we want to make it available quickly in the form that customers want it, whether it be print, web, or mobile. Generally, that means we must moderately granulate it and transmit it to the customer through the web in compressed form. While more detailed granularity provides flexibility for delivery and structure for authoring, the concomitant structural complexity adds to the effort of managing the content and increases the time required for compression and transmission. Quantifying the complexity of XML files may help us to do the best we can for our users by appropriately setting the granularity (Rockley and Cooper 2012). In this paper, I briefly discuss granularity, complexity, and entropy and then calculate the entropy for each of two XML files that have the same content but much different granularity.

Granularity

Granularity refers to the degree to which a collection of information is deconstructed into individually tagged components that users can access and manipulate with a query. Figure 1 shows an XML file with low granularity, along with its schema. The information has been deconstructed enough so that the individual lines are available independently for reference or display. This allows a certain amount of flexibility when designing and displaying the identification portion of a resume or a city directory. Figure 2 displays the same information set in a much more granular file. (Although the second file is remarkably larger, it contains less content because the punctuation that appears in the smaller file would be introduced by the XSL transform of the larger file.)



```
<identification>
  <name>Gaius J. Ceasar</name>
  <address>50 Apian Way NE</address>
  <citystatezip>Rome, Italy 00001</citystatezip>
  <phone>555-555-5555</phone>
  <email>gjcaesar@forumtel.com</email>
</identification>
```

```
<xs:complexType> <xs:sequence>
  <xs:element name="name" type="xs:string"/>
  <xs:element name="address" type="xs:string"/>
  <xs:element name="citystatezip" type="string"/>
  <xs:element name="phone" type="xs:string"/>
  <xs:element name="email" type="string"/>
</xs:sequence></xs:complexType>
```

Figure 1 - File with low granularity

The higher granularity of the second file allows us to reference, query, and independently display the smaller individual components; thus, it provides us with a great deal of creative license when we make the information available to users. However, that granularity makes it more difficult to manage the information. Picture an information set with 1,000 identifications recorded in these two XML files. Authoring, editing, and reading the more granular XML file demands greater effort and attention. We want to strike a compromise between the two examples of the XML file (Rockley and Cooper 2012).

Compression

Contemporary transmission of XML files goes beyond simple text compression. XML conscious compression separates the structural information from content; hence, tags and attributes are isolated, compressed, and transmitted separately from the bare-bones compressed content. Furthermore, some compression techniques include the schema, which makes it possible to query the transmitted compressed XML file. This is essential for mobile applications to run efficiently. The time and digital processing power required to orchestrate the separation of content from structure hinder delivery of the highly granular file compared to the minimally granular file. This is not simply because the larger file has six times the number of lines of code and more than four

times the number of elements. It is because the more granular file is more complex. Understanding how to quantify and compare complexity of XML files can enable us to properly assess the type of compression that would be most appropriate (IBM 2011).

```

<identification>
  <name>
    <last>Caesar</last>
    <first>Gaius </first>
    <middle></middle>
  </name>
  <address>
    <number>50</number>
    <street>
      <street_name>Apian</street_name>
      <street_type>Way</street_type>
      <street_location>NE</street_location>
    </street>
    <city>Rome</city>
    <state>Italy</state>
    <zip>00001</zip>
  </address>
  <phone>
    <area_code>555</area_code>
    <phone_number>
      <exchange>555</exchange>
      <index>5555</index>
    </phone_number>
  </phone>
  <email>
    <e_address>gjcaesar</e_address>
    <at>
      <company>forumtel</company>
      <suffix>com</suffix>
    </at>
  </email>
</identification>
<!-- SPACE-->
<identification>
  <name>Gaius J. Caesar</name>
  <address>50 Apian Way NE</address>
  <citystatezip>Rome, Italy 00001</citystatezip>
  <phone>555-555-5555</phone>
  <email>gjcaesar@forumtel.com</email>
</identification>
</xs:element name="identification">
  <xs:complexType><xs:sequence>
    <xs:element name="name">
      <xs:complexType><xs:sequence>
        <xs:element name="last" type="xs:string"/>
        <xs:element name="first" type="xs:string"/>
        <xs:element name="middle" type="xs:string"/>
      </xs:sequence></xs:complexType>
    </xs:element>
    <xs:element name="address">
      <xs:complexType><xs:sequence>
        <xs:element name="number" type="xs:integer"/>
        <xs:element name="street">
          <xs:complexType><xs:sequence>
            <xs:element name="street_name" type="xs:string"/>
            <xs:element name="street_type" type="xs:string"/>
            <xs:element name="street_location" type="xs:string"/>
          </xs:sequence></xs:complexType>
        </xs:element>
        <xs:element name="city" type="xs:string"/>
        <xs:element name="state" type="xs:string"/>
        <xs:element name="zip" type="xs:string"/>
      </xs:sequence></xs:complexType>
    </xs:element>
    <xs:element name="phone">
      <xs:complexType><xs:sequence>
        <xs:element name="area_code" type="xs:integer"/>
        <xs:element name="phone_number">
          <xs:complexType><xs:sequence>
            <xs:element name="exchange" type="xs:integer"/>
            <xs:element name="index" type="xs:integer"/>
          </xs:sequence></xs:complexType>
        </xs:element>
      </xs:sequence></xs:complexType>
    </xs:element>
    <xs:element name="email">
      <xs:complexType><xs:sequence>
        <xs:element name="e_address" type="xs:string"/>
        <xs:element name="at">
          <xs:complexType><xs:sequence>
            <xs:element name="company" type="xs:string"/>
            <xs:element name="suffix" type="xs:string"/>
          </xs:sequence></xs:complexType>
        </xs:element>
      </xs:sequence></xs:complexType>
    </xs:element>
  </xs:sequence></xs:complexType>
</xs:element>

```

Figure 2 - File with the same content as in Figure 1, but with higher granularity

Information, Complexity, and Entropy

The minimally granulated file has six elements, each of which has a declaration in the schema. The only element type is “string.” On the other hand, the highly granular file has twenty-four elements, some of which are doubly nested, with corresponding complex declarations in the schema and more than one element type declaration. Clearly, we judge the larger file to be more complex simply by inspection. Specialists have tried to attach numerical value representations to the different aspects of an XML file and its schema, combine them, and thereby express a calculated complexity in a single quantitative measure. One way this is done is by determining the entropy of a file’s schema (Basci and Misra 2008).

Entropy is a concept from physical chemistry that quantifies the intensity or richness of the energy stored in a system. From a science-fiction perspective, the universe will reach its end in a high-entropy death when all of its energy has been dissipated—evenly spread out. In statistical terms, entropy represents disorder, as in two sets of marbles being sorted (low entropy) or completely intermixed (high entropy). An archetypical teenager might have a bedroom with high entropy.

As applied to information systems, entropy measures the amount of information contained in a source. It is an appropriate concept for XML documents because the content in a large, complex XML document may appear disordered and unconnected, while in a different sense appear to have more order because of the structure built around the granularity. Entropy takes these aspects into account. In a general sense, the more complex a file, the more entropy it contains (Basci and Misra 2008).

Calculating Entropy

Here I illustrate the calculation of the entropy of the above two XML schema, following the procedure outlined in Basci, et.al. (2008). Shannon applied the concept of entropy to represent the

amount of information contained in a source (1948). A highly simplified expression of his entropy-information calculation looks like this:

$$\text{Entropy} = - \sum p \log_2(p)$$

p is the probability that a character occurs in the source
Entropy is the amount of information per character

For the XML files, each element is assigned to a class depending on how many child elements and how many parent elements it has. Then the entropy is calculated from the probabilities over all elements that each class of element occurs:

$$\text{Entropy} = - \sum (\text{class prob}) \times \log(\text{class prob})$$

The calculation parameters and results are presented in Table 1 and Table 2.

Table 1. Parameters and results for small XML file

| Element Class | Members | Number | Probability | Entropy Contribution | |
|-------------------|---|--------|-------------|----------------------|------|
| 0 parent, 5 child | identification | 1 | 1/6 | $-(1/6) * \log(1/6)$ | 0.43 |
| 1 parent, 0 child | name, address, citystatezip, phone, email | 5 | 5/6 | $-(5/6) * \log(5/6)$ | 0.22 |
| | | | | Entropy per element | 0.65 |
| | | | | Total entropy | 3.90 |

Table 2. Parameters and results for large XML file

| Element Class | Members | Number | Probability | Entropy Contribution | |
|-------------------|---|--------|-------------|--------------------------|------|
| 0 parent, 4 child | identification | 1 | 1/24 | $-(1/24) * \log(1/24)$ | 0.19 |
| 1 parent, 3 child | name, street | 2 | 2/24 | $-(2/24) * \log(2/24)$ | 0.30 |
| 1 parent, 5 child | address | 1 | 1/24 | $-(1/24) * \log(1/24)$ | 0.19 |
| 1 parent, 2 child | phone, phone_number, email, at, | 4 | 4/24 | $-(4/24) * \log(4/24)$ | 0.43 |
| 1 parent, 0 child | last, first, middle, street_name, street_type, street_location, number, city, state, zip, area_code, exchange, index, eaddress, company, suffix | 16 | 16/24 | $-(16/24) * \log(16/24)$ | 0.39 |
| | | | | Entropy Per Element | 1.5 |
| | | | | Total Entropy | 36 |

Conclusion

The complexity of my highly granulated and structured XML file is apparent at a glance. To quantitatively compare the complexities of two different XML files that have the same content, I calculated the entropies using a simplified application of Shannon's equation, following Basci, et.al. The highly granulated file has twenty-four elements, and the less granular file has six. It is smaller by a factor of four. However, the entropy of the less granular file is smaller by a fact of about eight. This demonstrates how the entropy calculation captures the intricacies of the XML schema of the larger file. Entropy gives us a sense of the structural weight of an XML file, entirely apart from its content, that must be carried in its storage and transmission.

References

- Basci, Dilek, and Sanjay Misra. "Entropy Metric for XML DTD Documents." ACM SIGSOFT Software Engineering Notes 33, no. 4 (2008). <http://www.balisage.net/Proceedings/vol8/print/Ruellano1/BalisageVol8-Ruellano1.html>.
- IBM. 2011. Investigate state-of-the-art XML compression techniques. <http://www.ibm.com/developerworks/library/x-datacompression/index.html>.
- Rockley, Ann, and Charles Cooper. 2012. *Managing Enterprise Content: A Unified Content Strategy*. 2nd. Berkeley: New Riders.
- Shannon, C. E. "A Mathematical Theory of Communication." *The Bell System Technical Journal* 27 (1948): 379-423, 623-656. <http://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=6773024>.

TECHNIQUES STAFF

Managing Editors

Paige Boyle
Tara Neita
Michelle Tirado
Caitlin Wolanek

Layout Editor

Alicia Campbell

Editorial Staff

Tiffany Jackson
Amanda Jones
Katrina Myers
Jeffrey Pellico
Ines Wingert

Faculty Advisor

Lee Tesdell

You may only reprint material in TECHNIQUES with permission. Credit must be given to the author and a copy of the reprint must be sent to the faculty advisors.

A member of the Minnesota State Colleges and Universities System. Minnesota State Mankato is an Affirmative Action/Equal Opportunity University. This document is available in alternative format to individuals with disabilities by calling the Department of English at 507-389-2117 (V), 800-627-3529 or 711 (MRS/TTY).

Contributors

Paige Boyle
Emma Furness
Tiffany Jackson
Katrina Myers
Tara Neita
Jeffrey Pellico
Michelle Tirado
Jack Sandberg
Ines Wingert
Caitlin Wolanek

Cover photo credits:

MNSU Image Library #: 734FA08, 512SP07, 673SP08, 969SP10

TECHNIQUES is published by the Minnesota State Mankato student chapter of the Society for Technical Communication (STC). For information about the STC student chapter, please contact Lee Tesdell at lee.tesdell@mnsu.edu or visit www.stc.org.

