Crossing the Objective-Subjective Divide in Information Space Organization

Stephen Davies, Roger King

University of Colorado, Boulder
{stephen.davies, roger.king}@colorado.edu

Abstract

People mentally organize the information they encounter, forming their own subjective interpretation (categories, principles, relationships) of objectively agreed-upon facts. Many tools are available to help them record such interpretations, from pen and paper to word processors to next-generation note-taking tools. It is nearly always the case, however, that the personal theories and observations they record refer only indirectly to the entities described in the source documents. When an investigator reads a Web page about something, for example, and then records a note about the same topic, the only connection between the two is one of convention: perhaps the same character strings are used to refer to the same entities. This connection is tenuous at best, and could be improved by combining Semantic Web technology with a personal information management tool that maintained explicit, unambiguous links between the objective and subjective realms.

Keywords: Information space organization, personal information management, Semantic Web, semantic network, note taking

1. Introduction

Bring together a half dozen experts in any particular field – say, some aspect of European history – and two things are certain. The first is that there will be a tremendous amount of shared knowledge in the group, and to a degree of detail quite impressive to the layman. These experts will all know the same events and places and people – hundreds of them that the average man has never even heard of. The experts will all be familiar with the important researchers in the field, and aware of each one's specialty and affiliation. And they will share a specialized vocabulary of strange and mysterious terms unknown to anyone outside their own circle.

The second certainty is that despite all of this agreed upon knowledge, these experts will have widely different views on the subject. Each will have their own opinions and pet theories; each will stress certain events and think others less important; each will favor different explanations of reigns and revolutions. When the name of a certain king is mentioned, one of the experts will immediately think of two of that monarch's contemporaries who were directly involved in setting his policy; while another participant will instead be reminded of three other kings from different eras who faced similar challenges. One perceives a relationship, causal or coincidental, between two events; another discards the notion. The way each researcher relates all this information, and organizes it in his or her mind, is unique. One reason is that those seated at the discussion table have different backgrounds and different biases, and it is these differences which lead to fruitful dialogue.

These, then, are the two kinds of knowledge present in the room: the facts shared by all, which allow the discussion even to take place, and the differing interpretations of those facts, which are unique for each one present. We might term these two areas of knowledge "objective" and "subjective." The objective realm consists of those items that are on the table for everyone to see – they are the indisputable building blocks without which nothing whatsoever could be understood. The subjective realm, on the other hand, is composed of an individual's perceptions and preferences. It is a different arrangement of the building blocks for each viewer.[1]

All knowledge, it seems, has these two components. We constantly encounter new information throughout our lives, and accumulate

---

[1] Note that *objective* is not here synonymous with *true*. Several centuries ago, "the earth is flat" was a "fact" in the objective realm that could be discussed and referred to, even though we now know it to be false. Similarly, "Zeus," though most today would consider the name to refer to a nonexistent being, is nevertheless a reference to a conceptual entity that can be discussed and related to other entities. Hence we can think of the World Wide Web as containing facts from an "objective realm" while by no means asserting that all of its information is true.

more and more objective knowledge. But there is something else, too: in order for our mind to retain the information, to make sense of it and interpret it in light of what else we know, we form cognitive structures to organize it all.[6, 7] We associate some facts with others and form groups of related entities. Soon, when we hear mention of the French Revolution, our minds immediately evoke those of the Bolsheviks and of the American Colonists; when we read that a famous leader was born in 1926, we cannot help but place his or her childhood in the context of the Great Depression. Our minds do this automatically, it seems, and philosophers and linguists have debated for decades as to how such knowledge is actually represented internally. But no matter the particulars, it is clear that the subjective realm is at *least* as important as the objective for our understanding. It is the organizing structure, full of clusters and categories and perceived relationships, that gives the objective facts context and larger meaning. It is what helps us form conclusions, assimilate new facts, and make sense of our world.

2. Materializing the subjective realm

Most of the time, our subjective realm remains ethereal. It is a product only of the mind, and we seldom think about it explicitly at all: it is only the lens through which we view information in the objective realm as we encounter it and consider it. The trouble with this situation, however, is that it relies exclusively on our biological memory to retain not only the facts, but our conclusions and the structure we have built to embody them.

The natural solution, known since the dawn of humankind, is to take notes as we learn. We can choose physical pen and paper, or an electronic word processor and hard disk. As we read document after document, we will record the salient information in a distilled form so that we can return to it later, mixing the prominent objective facts with our own conclusions. This is roughly satisfactory, and is where most people stop. But consider two problems with this technique:

1. The medium of expression we use to record the notes is alien to our natural knowledge representation. Some linguists may dispute this, but it seems clear that after reading a chapter on the American Civil War I do not literally have the *sentence* "Ulysses S. Grant became the commanding general of the Union Army in October 1863" in my brain.[2] That sentence does not even come into being until I choose to articulate my knowledge to others, or jot it down on paper. Once it is read and parsed and

accepted, the English form goes away. Hence taking notes and reviewing them later involves a rather painful conversion process in both directions: from mental impression to natural language and back again. It seems worthwhile to explore the possibility of actually recording the mental impression without the awkward intermediate form.

2. They offer no hypertext capability whereby one idea can "link" to another. Surely our minds are full of such links: depending on context, John F. Kennedy could be compared with other Cold War presidents, or other assassinated statesmen, or other figures from political dynasties. But unless we maintain an explicit list for every conceivable category, and strive rigorously to keep them all up to date, we cannot manage these relationships.

To address this last point, we can certainly consult the world of hypertext systems. We could create a document or a page for every concept in our minds, and then use hyperlinks to express the interrelationships we see. This is essentially the approach taken by advanced note-taking systems like MindJet's Mind Manager[1] and Eastgate Systems' Tinderbox[5]: reduce the size of a mental concept (a "topic" in Mind Manager's jargon, or a "note" in Tinderbox's) down to something very small, typically just a phrase, and allow them to be arbitrarily linked. This permits the full range of expression that our minds demand, where any concept can be associated with any other concept. Thus the second problem, above, is seemingly solved. And if we can sufficiently break down our knowledge into a fine enough granularity, we have made great strides in addressing the first problem as well: sentences are few and far between in our subjective materialization, replaced by a loose semantic network that presumably better corresponds to our mental impression.[8]

3. The objective-subjective divide

There remains, however, one problem, even for advanced tools such as Mind Manager. It is subtle and conceptual in nature, but it has very practical ramifications. And that is that we have created a chasm between the objective and subjective realms.

Consider what happens when a user browses an information space, such as a group of Web pages, in order to gain knowledge. She scans page after page, following links, searching for relevant data. Along the way she encounters many bits of useful information which help her better understand the domain and to construct her own mental model of it. Now if she wants to externalize some of this information, rather

than relying on her own memory, she will inevitably jot down a note. This may be on physical paper, in an electronic text editor, or using a sophisticated tool like Tinderbox. But in any of these cases, she will write phrases that contain words referring to the entities she is reading about. Suppose she writes:

Lincoln influenced by both Unitarians and C. Finney.

The pages she is reading are conceptually *about* a particular 19th century leader named Abraham Lincoln and those who impacted his worldview. The notes she is composing are about the same things. Yet the only link between them is tenuous: it depends on a human reader properly interpreting potentially ambiguous terms and drawing the right connections himself. Certainly even a sophisticated note-taking tool is unaware of the connection: to it, "Lincoln" is just a string of letters, and has been completely severed from its source. And to a human reader, it might well signify the capital of Nebraska, or a type of luxury automobile. Thus even though what is really happening is that a viewer is exploring an objective realm of real facts and forming her own subjective framework to organize and explain those facts, none of this is visible to the tools she is using – either to the Web browser, which displays only human readable character strings, or to the note-taking tool, which accepts the same kind of data. And there are certainly no guarantees of semantic equivalence that either tool could possibly take advantage of.

The basic problem here is that even though the user is trying to organize the information space, the tools are not letting her do exactly that. Rather, she is *simulating* that process by composing English phrases in one tool that contain words bearing some resemblance to those in another tool. All correspondences are up to her, or whoever she shares her notes with, to draw on their own.

4. Bridging the divide

Great gains could be made if we could bridge this divide by more closely correlating the whole architecture with the human thought process. Currently, the Web (an objective realm, say) is mostly comprised of free text that describes important entities. People's notes (subjective realms), scattered in their various forms, are also comprised of free text that describes those entities. Both approximate, but partially miss, the truth. It would be a great advantage if both realms were actually *comprised of entities,* rather than merely free text that indirectly described those entities. Then the objective and subjective realms would be speaking the same unambiguous language that allowed one to refer directly to the other.

On the objective side, this is exactly the aim of the Semantic Web initiative. Berners-Lee, et. al's vision for the Semantic Web[4] is one in which a web page is no longer merely of interest to human readers, but contains hidden, machine-processible markup that rigorously describes the entities the page mentions. Hence an individual's home page, say, will not be merely an attractive display that reflects upon its owner's life, but a first-class object of a universally agreed-upon type and with properties specified by a standard ontology. Agent software can then operate on such a page to form reliable deductions and perform automated tasks.

If the Semantic Web becomes what its founders intend, then the objective realm we use daily will undergo a substantive change. The information will
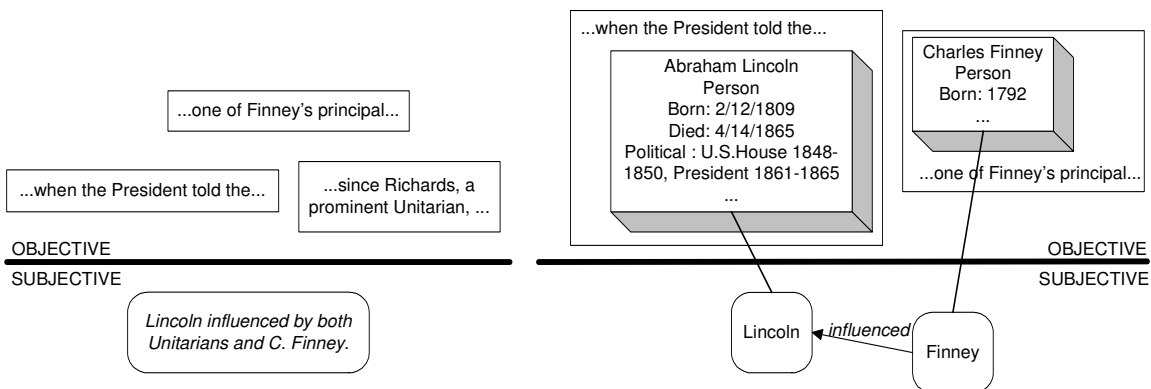

Figure 1. The left diagram shows today's state of the art. A user browses an objective space, learning about some topic, and then creates free-form notes to reflect their own subjective interpretation and organization of the facts. The correspondence between the entities referred to in both the sources and the notes is fragile, based only on linguistic clues. The right diagram shows a substantial improvement: Semantic Web pages are annotated with objective information about the entities, and a user's note-taking tool can refer directly to the entities themselves.

not become richer, but it will become more precise. In addition to billions of pages that describe people and events in elaborate prose, there will be available a staggering array of *objects*: formally defined entities about which strict assertions are made. In effect, these assertions become the "facts" in the objective realm, expressed unambiguously and formally. Each conceptual entity, referred to now on a thousand different pages and with a dozen different noun phrases, will have one and only one globally unique identifier.[3]

What is needed, then, is a tool that allows individual users to build their own subjective frameworks that refer directly to those objective entities. Rather than writing English sentences which contain imperfect labels for objective data, users could drag and drop Semantic Web objects directly into their workspace. These objects could be annotated with the user's personal notes. They could be classified and arranged according to the individual's perception. New relationships could be drawn between them and between purely "subjective" objects of the individual's own making. Over time, the user's workspace would be a crisp reflection of her own mind, depicting precisely how she chooses to organize the objective information she has encountered.

Such an information management tool on the subjective side, coupled with the Semantic Web on the objective side, offers the following advantages:

- The objective entities are unambiguously defined. No longer is there any question of who "C. Finney" was or which particular group of Unitarian thinkers supposedly influenced Lincoln. The facts that the user has organized are actually pointers into a space of globally unique identifiers which can be accessed to resolve uncertainties.
- The subjective space is tied to the objective space through these identifiers, making it easy to access the "agreed upon" facts about any particular entity. Mind Manager and Tinderbox offer an approximation to this function through their ability to attach URLs to any topic or note. But these links function as interesting cross-references rather than the unique, definitive source of coded information about the entity.
- Everyone who uses such a tool to express their subjective perceptions will be using identical identifiers to refer to identical entities. This makes it easy to share and compare subjective views. If one Civil War expert wants to browse a colleague's subjective view to understand her theory, he doesn't have to search through large amounts of free text for key words like "Grant" or "Fredericksburg" to get his bearings. He can simply say, "find the object with this URI

wherever it appears" and be certain that he has found everything. One could even imagine a "diff" function that compared two subjective views, showing the objects that each had in common and perhaps highlighting in two different colors the distinct organizational patterns, categories, and links of the two views.
- The kinds of agent programs that are intended to operate on Semantic Web data could be applied to the subjective realm as well. All of the objective facts and relationships are immediately accessible to the tool through the URIs of the workspace's entities, and so the bridge between subjective and objective is seamless.

4.1. Collaborative possibilities

We have also taken a significant step towards true collaboration. Not only is my own subjective realm now captured in a form eminently compatible with the objective realm, but it is compatible with everyone *else's* subjective realm as well. I can better share and compare and collaborate, as alluded to above. And automated tools could potentially process all of this data – objective and subjective – to find important global trends in the way people view information.

This allows us to bridge the objective/subjective divide in quite a different way. Consider that as consensus builds within a group, statements that were once in various thinkers' subjective realms begin to migrate to the objective realm. The more people begin to universally see things the same way, the more that aspect of the subjective world *becomes* objective. And the fact that the representations of the two worlds are so closely aligned would allow this transition to easily take place. One could imagine a process in which what one person asserts can be compared and shared and assimilated by others, and eventually offered up to the community at large as a candidate for admission to the objective realm.

5. A word about bookmarks

Now some may claim that this bridge between the two realms already exists, and is in fact employed by millions of users every day: the ubiquitous Web browser "bookmark" (or "favorite.") When browsing objective information on the Web, users can quickly store the URL of any page of interest, thereby enabling easy return. URLs are globally unique, guaranteeing that any two users bookmarking the same page will be using the same identifier. And such bookmarks can be organized in custom ways by each user; either in a "favorites" folder hierarchy, or by including the URLs inline in a set of notes, perhaps even attaching them to

topics or notes in a Tinderbox-like tool. Doesn't this feature give us exactly the ability we seek?

The answer is an emphatic "no." There are two reasons that today's bookmarks are insufficient for this task:

1. As noted previously, a bookmark is a reference to a *page*, not to a conceptual entity. It is a pointer to syntactic, not semantic information. True, there are examples where a Web page corresponds roughly to an underlying, real-world entity – a personal home page may serve as a rough stand-in for a specific human being, for instance; or the welcome page for a company's Web site might be considered to refer to that business. But this involves considerable guesswork on the part of human users. If we are to capture knowledge, we want to cleanly and crisply refer to the elements of that knowledge, not to unreliable approximations of such.

2. Even more importantly, a bookmark is not nearly *precise* enough for our purposes. It allows one to create a reference only to an entire document, when in fact a document may contain paragraphs of information describing all sorts of people, things, events, and relationships. We have much higher aspirations than merely to point to a few large chunks of unplowed knowledge in cyberspace. This leaves the user with the burden of revisiting those chunks later to re-read, re-parse, and re-interpret them in order to form their mental impression once again. Instead, we want to point directly to the smaller, more elemental bits of information they contain.

Another way of stating this is to realize that our goal is primarily to capture and organize *knowledge*, not knowledge *artifacts*. I read civilwar.com because I want to learn more about the Civil War, not because I want to learn about civilwar.com. And when we take the time to process an information source, our minds emerge with much more than an opaque chunk called "www.civilwarhome.com/shilohbattle.htm." Instead, we emerge with knowledge about the document's *contents*: information about Grant, Buell, Beauregard, the battle of Shiloh, Owl Creek, infantry and cavalry units, sieges, ravines, churches, and hills. To properly capture this knowledge, we need to do much better than merely aim a crude pointer towards the blob of text from which it was originally derived. We need to liberate and expose the many different entities that a text describes so that our knowledge representation can be in terms of those entities, using nimble and precise links to refer to the accepted concepts of a domain and to organize them in personalized ways.

6. Conclusion

The objective realm consists of the facts upon which everyone presumably agrees. Isn't it reasonable, then, that these entities should have agreed-upon identifiers that everyone can use to refer to them? And isn't it reasonable that when someone organizes part of the objective realm by building their own subjective framework, that they use these identifiers directly?

It is not the purpose of this paper to outline a specific tool or to prescribe any particular user interface paradigm. There is room for much creativity here, and it is the work of researchers in the immediate future. The only intent is to call to attention the fact that Web users currently have very few ways of actually referring to the entities described by the sources they read so that they can properly organize their own knowledge space.

But if we could break through the syntactic barrier and let users manipulate objective entities directly, rather than hint at them imprecisely through erratic note-taking, we have achieved a real gain. We could build tools that let us catalog the entire breadth of our knowledge in an integrated framework, so that the information we process throughout our lives does not fade over time, but can be readily consulted and retrieved in a form very like what we see in our minds.

7. References

[1] MindManager(R) by MindJET LLC. Test version V5.2.344, 2004.

[2] Anderson, J. R., <u>Cognitive Psychology and Its Implications</u>, 3rd Ed. New York: W.H. Freeman.

[3] Berners-Lee, T., Fielding, R., and Masinter, L., "Uniform Resource Identifiers (URI): Generic Syntax." Network Working Group: Request for Comment 2396: 1998.

[4] Berners-Lee, T., Hendler, J., and Lassila, O., "The Semantic Web," in *Scientific American*, May 2001.

[5] Bernstein, M. "Collages, Composites, Construction," in *Proceedings of the Fourteenth ACM Conference on Hypertext and Hypermedia.* Nottingham, UK. August 26-30, 2003.

[6] Kintsch, W., <u>Learning, Memory, and Conceptual Processes</u>. John Wiley & Sons Inc., 1970

[7] Sowa, J. F., <u>Conceptual Structures: Information Processing in Mind and Machine</u>. Addison-Wesley Publishing Company, 1984.

[8] Woods, W. A., "What's in a Link: Foundations for Semantic Networks," in <u>Readings in Knowledge Representation</u>, R.J. Brachman and J. Levesque, Editors. 1985, Morgan Kaufmann.