What You (Fore)see is What You Get: Thinking About Usage Paradigms for Computer Assisted Text Analysis

John Bradley
King's College London
john.bradley@kcl.ac.uk

Abstract

Humanities Computing (HC) has been in existence since the beginning of computing, and was originally created as an attempt to transform aspects of humanities scholarship – a goal that it retains today. It would seem, however, that to date this central aim has not been achieved. In this paper we argue that HC might be more influential if it moved its operations closer to traditional scholarly methods. To illustrate the issue, we describe four usage paradigms for computing in humanities research. We argue that the conduit model is the one in the minds of most scholars who use computing to support their research today – a model which supports established practice but fails in important ways to take advantage of much of what computing could offer. In contrast, the HC community is much more interested in two other models (labelled here as the transformation and markup models) both of which, it is argued, do not connect well with the activities traditionally associated with the majority of humanities scholarship. We therefore propose a fourth model, the object manipulation model, which we believe is able to take better advantage of the possibilities computers provide to research than the conduit model and is, like it, more closely based on actual majority scholarly practice. We also mention that other models (such as those arising from the work of Druker and McGann) are also possible, and worthy of serious consideration within the HC community.

KEYWORDS: Mental models, Text analysis, Scholarly practice.

I have enjoyed close involvement with humanities computing (HC) since 1977 and this paper arises out of some personal reflections that I have developed over the last ten years about the role of the computer in the humanities. It arises out of the, let us say, ambitions of humanities computing that has driven it from the very beginning: an evangelistic wish to

affect humanities research not only of those who already have the HC bug and find existing methods useful to them, but also those who do not use computing in any way that is terribly relevant to their research interests. I want to influence the type of scholar who was described in the *Face of Text* conference as "our colleague down the corridor."

The pioneers of HC, such as John B. Smith (see Smith 1978) and Rosanne Potter (see Potter 1980), seemed to believe firmly in the great potential for the computer in textual analysis and criticism. They saw it as a way to transform textual criticism so as to make it more objective and to eliminate what Potter called "acts of sorcery." Now, twenty or thirty years on, we see that this goal has not been achieved – if that goal is measured by the number of humanities scholars who have adopted this view for their own research. Instead, as Potter herself has written: "many computer critics have written themselves out of the range of their natural audiences." (Potter 1989, xviii).

More recently, the rise of the World Wide Web, and the arrival of the Internet into pretty well every academic office in western academia has encouraged many again to expect great things from technology. See Wulf 1995 as a perhaps typical view. Less then ten years on, however, we find the expectations surrounding the World Wide Web and the appearance of digital libraries are more measured. See the comments written in a report sponsored by the American Council on Library and Information Resources in 2001:

> While digital resources are becoming more visible in the humanities, use of these resources by scholars remains limited. … Digitization projects are bringing texts, data sources, sound, and images to the scholar's desktop; however, the functions on which research in the humanities depend are neither well understood nor well supported by librarians. (Brockman et al, 2001)

This feeling of disappointment is strong enough that we find members of Canada's TAPoR project conducting a survey to explore how computing seems to have affected scholarship, and to get some sense of what might be further needed. On the question of computing tools to support research the response was interesting:

> In general, respondents believe that they need text analysis

> tools, although not complex tools, and are not happy with the tools that are currently available. Somewhat surprisingly, over 50% did not know about commonly available tools such as TACT, WordCruncher and Concordancer. The one most highly used was TACT but few found it useful. In addition to our list of about ten tools, participants added another two dozen tools that they employ in their work. These included tools such as the Wordsmith Tools as well as common Microsoft Office products such as Word and Access. (Siemens et al, 2004)

So 50 and 60 years on it still feels as if the HC community has not had the breakthrough that it still seems to want and expect. It seems that most humanities researchers find existing HC tools do not support the research they actually do. Perhaps it is time to step back and re-examine the mental models (a set of assumptions) we all bring to the question of what the computer can do to support research. The computer is a very abstract machine, and my premise here is that existing models both consciously and unconsciously exclude certain aspects of how computing might conceivably help. Hence, it is my intention to draw attention to the models we find in the minds of computing humanists, and to contrast them with the models held by our non-computing-humanist colleagues down the hall. As we shall see, the same project might exhibit one model for developers and a different one for users.

In this paper I will be discussing four models, although there are doubtless others you could think of. I will start by suggesting that for most of our humanities colleagues, the model that they apply to computing for their research is what I call here the *conduit model*. In contrast, within the humanities computing community much work fits into two other models which I have dubbed the *Transformation Model* and the *Markup model*. Furthermore, these are two models that many in the HC community use to place computing within the humanities, and they represent models we'd always hoped our colleagues down the hall would also adopt. Finally, in order to suggest other ways of thinking about computing and humanities scholarship I describe yet another model – also well established in computing in general, but seemingly missing from much thinking about computing and scholarship: what I call the *Object Manipulation Model*. Furthermore, this last model can be clearly related to some of the issues that Brockman *et al* suggest in their report are missing from established HC.

The Conduit Model

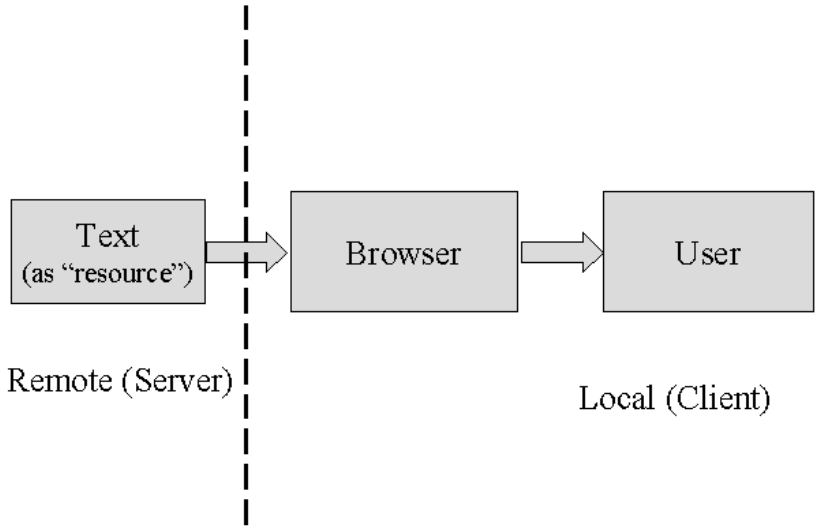

Figure I: The Conduit view

As the word "conduit" suggests, this model focuses on the ready availability of seriously scholarly digital materials that we believe the WWW delivers (or at least will deliver sometime in the future) to almost everyone's desktop. The user "opens the tap" and the stuff flows out whenever s/he wants it – and to the individual's desktop as well. A characteristic phrase associated with this view is the description of digital objects as "digital resources" – the *resource* term came into common usage at about the same time as the WWW became widely available. Indeed, this particular view is now so widespread that almost every digitization project I hear about today is virtually equated to "web editions" and "web publishing". In the UK, of course, we have, in the conference *Digital Resources for the Humanities* (DRH), a conference that focuses on project work that is largely centered on these kinds of scholarly tools. In parallel with this was the development of "digital libraries" as a subfield in computing science, with – at least for the past 10 years or so – serious research money directed to support new thinking in this area.

Figure I provides a very simple schematic view of the conduit model, and draws attention to certain aspects of it. First, the overall model, introduced by the WWW as we see it today, is almost completely

centred on the *client-server* approach. The resources are "out there" on a remote server machine. The user is "here," and can select what s/he wants to look at with his/her browser. Once selected, the object flows down the Internet pipe to appear on the user's screen to be viewed. Conventional wisdom says that WWW resources are more readily usable when material is presented in small screen-size chunks – so the digital resource is often conceived of more like a printed reference work than, say, an monograph. Each individual user will consult only a part, and in an order directed by their interests.

There are now quite a number of WWW-oriented resources that are potentially useful to conventional scholarship. I will mention two quite different examples here. The *Early English Books Online* (EEBO) is a partly commercial project through which its developers intend to put online *images* of about 125,000 books published in English through the age of Spenser and Shakespeare. Some texts are presented in transcription, but the primary goal of EEBO seems to be to make available a large set of textual materials as images. The images are good enough for many scholarly purposes (those that would, for example, be served by microfiche in the past), and the scholar can ask for them to be packaged up in a PDF file so that s/he can retain the complete work on his/her own computer. However, it is clear that it is expected that this is *all* the end user will do with the EEBO materials – view and perhaps print them. We can see what the UK's Arts and Humanities Data Service (AHDS) thought was the principal benefit of EEBO for institutions in the UK from this quote provided by a post-graduate student user in the AHDS newsletter:

> Once you get a taste of what research can be like with EEBO you want more. It transforms how you work. I can work at 2am. I can scribble on my printouts. I'm not restricted by library opening times. I've cut my transport costs and time. It's simply more efficient. (Leon 2004)

Jerome McGann's influential *Rossetti archive* (McGann 2000) is, perhaps, at first glance a striking contrast to EEBO. Unlike EEBO, which seems to be to be inspired by a librarian's perspective of a resource, the *Archive* is a scholarly project with ambitious and subtle scholarly goals. McGann has thought a lot and deeply about the implications of the creation of digital resources. However, let us take a moment to look at the archive from the perspective not of McGann and his colleagues at the Uni-

versity of Virginia, but of the *user* of McGann's archive in his/her office.

The "affordances" – what it allows the user to do – provided by the users of the *Rossetti* are described informally in the archive's introduction (http://jefferson.village.virginia.edu:2020/intro.html):

> The user of the Archive is therefore not simply given access to linked and organized sets of digital images and alphanumeric texts of Rossetti's own works. These primary materials have been integrated into a critical and scholarly environment of great depth and complexity.

One has to parse this sentence with some care to see its significance from the perspective of this paper. In fact, the user *is* just given *access* to this rich and sophisticated set of materials in the archive. The aspect of the archive that differentiates it from the more modest scholarly goals of the EEBO is *not* in the more complex interaction the user can have with the material, but in the more complex nature of the of the materials provided. In spite of the sophisticated structure designed by the archive's editors, end *users* of the archive still use it in the context of the conduit model. They see the Archive largely as the digital equivalent of a fine scholarly printed work and they use it in that way. They find material of interest to their research; they view the primary source material available from the archive, they read the commentary prepared by the Rossetti scholarly team. Perhaps they print out what they want and treat the printout as they would a photocopy: file it; perhaps write on it. In the end, their experience of the computing *potential*, let us say, of this material is not very different from printed resources.

Much of the user experience of this material comes out of the set of functions built into the software that most strongly defines the conduit model – the web browser. The browser makes certain actions easy, it makes others at least possible, and provides little or no assistance for still others. The classic model of interaction with the browser is "point and click" saying "I want to see *this*." The browser software makes it easy to print a document that has been thus fetched, and makes it possible (although perhaps not quite so easy) to save a personal copy. Support for independent "surfing" through the web to find materials of interest is not really provided by the browser at all, but by search engines provided on some remote web server machine.

More importantly, as a scholarly tool the browser provides little

sense of *state* (perhaps only a favourites list) that can develop over time and that could represent work in progress for the browser user. Since, surely, scholarly research is found centrally in the development of a more and more complex, interconnected, and richer "state" of understanding in the researcher's mind, this means that the browser only supports this essential aspect of scholarship in a very limited way.

The Transformation Model

The conduit model dominates how our colleagues have thought in recent years about computing in their research. However, there have been two strikingly different, and incompatible, models that have driven humanities computing for decades. Unfortunately I believe that neither model really touches on the research lives of most non-computing colleagues in the way that the conduit model could.

Humanities computing began with a role for the computer that broadly corresponds to a transformation model. In this model (shown in figure II) the text is, as it was in the Conduit model, an *object of study*. However, whereas in the Conduit model the only role of the computer is to deliver material to the user, here the researcher uses the automaton nature of the computer to transform the given text in some way or other. The transformed results are viewed by the researcher who hopefully sees new things of interest in the text that perhaps have not been seen before.

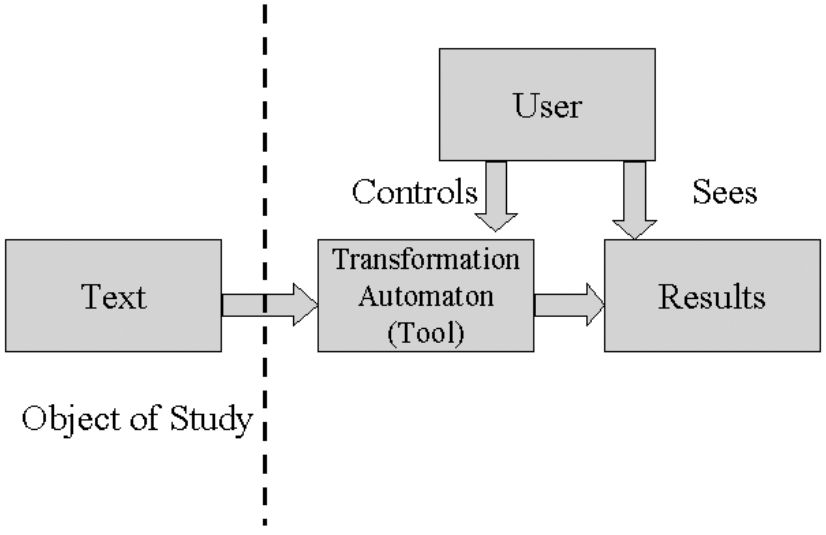

Figure II: The Transformation Model

The transformation is provided through a program "tool" (one of the words that often implies a transformation model at work) that is designed to perform some automatic transformation. Sometimes, a set of related transformations has been packaged together into a single package (a toolkit) by the program developers. The user of the toolkit then has the ability to apply several of the provided transformations, either serially or in parallel, to his/her text. See (Ott 2000) for a description of what is considered by many in HC to be an exemplary toolkit – TUSTEP.

The KWIC display is perhaps the best-known example of a result from the transformation model. Of course, however, more radical transformations are possible. TACT, for example, could produce a distribution display which showed how the use of a word was distributed through a text. Statistically based transformations are perhaps more radical still – think, for example, of John Burrows's influential work based on the frequencies of the most common words appearing in major 19 and 20th century writers. There is a good example of this in Burrows 1992.

It is important to realize two things about the classic HC transformation model. First, the transformation must be carried out entirely by the computer without the need for human intervention. It is in this way that pioneers such as Potter believed that the computer contributed to the objective nature of the analysis. The human user's wishes or prejudices did not affect the result. Furthermore, in the classic HC transformation model, the transformation operation is something the user fully understands. For the KWIC display, for example, the user understands completely what the computer has done to generate it, and a KWIC display could be done by hand without the aid of the computer, but of course at the cost of considerably more time and effort. Transformations such as the distribution display or Burrow's statistical graphs work the same way. In both cases, they are most convincing when we fully understand the computing (in Burrows's case the statistics) behind them.

Still, in spite of the fact that we could do the transformation we ask the computer to do, the results provide a new context for viewing the text. The KWIC readily draws our attention to the context in which a word occurs and it is from seeing at a glance this context for all instances of a word that we hope to see something we would have missed otherwise. A transformation such as the distribution graph uses a transformation to facilitate the visualisation of some phenomena. The eye-catching nature of the result is meant to stimulate the mind to new thoughts about the underlying material.

Finally, if the computer can carry out the transformation quickly, one often finds oneself using the transformations in an experimental way – not always sure, in advance, what we are going to find, but directing our requests based on some degree of informed guessing about what is likely to be fruitful.. So, even though the transformation does not represent something that we couldn't do "by hand," it is this speed of response when the computer does it that is important, and we can do more experimenting as a result. There is a role of serendipity – the user may not know at the beginning whether results will be interesting (or perhaps how it will be interesting) until s/he sees it.

Unfortunately, from the point of view of our colleagues down the hall, the transformation model is centrally based around the use of the computer as an automaton, and there is great difficulty in developing (perhaps, indeed it is impossible) transformation rules to control an automaton that are sophisticated enough to relate to much humanities scholarship. Perhaps because it is straightforward to pull text apart into word tokens and word forms, word based transformations tools abound, but unfortunately actual word forms don't directly relate to issues that are central to many scholars. This kind of work is relevant only to certain researchers, since the majority of critics don't see their analysis is based on aspects of the text readily caught by a computer automaton.

The Markup Model

Perhaps in part as a response to the limitations of expression in the transformation model, some in the HC community began to think that perhaps through markup the researcher could use the computer to formally record on a text aspects of that text that *were* of interest to him/her. Out of these considerations the Markup model arose.

The Markup model has both a long and, with the development of the massive and profoundly influential *Text Encoding Initiative* (TEI) (Sperberg-McQueen and Burnard 1994), an extremely important history. Anyone whose scholarly work focuses on the preparation of primary texts for scholarly use should surely be aware of the TEI and its implications. Furthermore, the TEI provides a way (or sometimes a framework in which a way can be found) of expressing many kinds of analyses that one would want to make about a text.

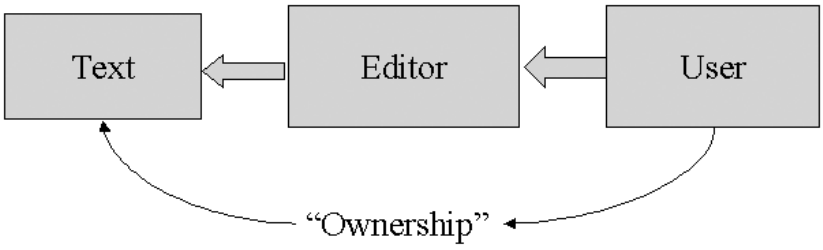

Figure III: The Markup Model

Figure III shows the simple schematic that characterizes the Markup model, and it is useful to contrast this with both the conduit and transformation models shown in figures I and II. Perhaps the most important difference is that the arrows point the other way, and this results in a substantially different relationship between the user and his/her text. In the Markup model the researcher feels a sense of ownership of the text, and the materials that s/he is adding to it.

Whereas the characteristic piece of software for the conduit model is the *browser*, the characteristic piece of software for the Markup model is the *editor*, and it is often the editor's affordances that shapes the Markup model. Leaving aside the many important arguments about non-hierarchical markup that have enlivened the humanities computing community for many years, we will focus in this article on markup as expressed in what is the current standard for such things: the Extensible Markup Language (XML).

Figure IV shows a bit of a transcript in TEI XML that we at King's College London are preparing as a part of our work on the digital edition of the *Durham Liber Vitae* (DLV) (a project headed by Prof David Rollason, University of Durham), as viewed in one of the very popular XML editors *oXygen*. The DLV markup is very dense and perhaps in some ways characteristic of the complex markup needed to represent the results of the kind of intensive study the project is applying to this text.

There are a couple of things to note that constantly affect the work of the marker/user. First, the XML editing software is "aware" of the principles of XML and emphasizes the XML tags, their attributes and their essential hierarchical relationship. The user's attention is drawn, particularly in this dense markup, to the tags and their technical relationship to each other rather than what the tags represent. For a project of this kind this issue becomes a serious one since the project team has to glean sig-

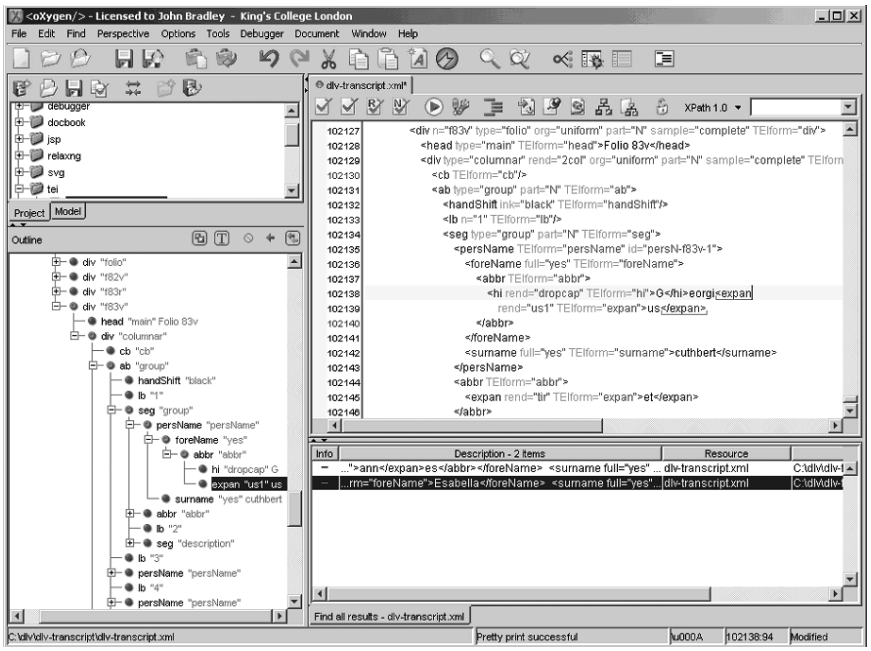

Figure IV: An XML Editor Screen: oXygen

nificance from the markup that is hard to grasp when the tagging is so rich. The markup, rather than the things about the text it represents, begins to take over the visual field of the user.

The second thing to note is that the editor software (as does XML itself) prioritizes the document ordering of the materials – the order in which they are presented as one works through the text from the beginning to the end. This is fine when the concepts the researcher is working with are naturally thought of in document order as well. If, however, the items of interest are not usefully thought of in document order the fit is not quite so natural. Connections can be expressed between objects through links in XML, but editors such as *oXygen* do not provide any serious support for this kind of task. See (Bradley 2005) for an extensive discussion of this issue as it applies to the DLV, where the researchers wanted to work with materials and concepts centered on the DLV text some of which were well represented in document order, but others of which were not.

I don't think that the desire to work with aspects of a text that are both document and non-document ordered is that unusual. It is perhaps surprising, therefore, to note that although there has been considerable debate with the Markup Model community about the problems caused by XML's single-hierarchy containment model, there has been far less dis-

cussion about the problems caused by XML's preference for document order. This suggests to me that the orientation of those discussing markup theory and related issues have been primarily centered on using markup to express things that relate comfortably to the document order. Projects such as the preparation of digital editions of text work in this way, and indeed represent the bulk of projects that use XML and TEI as the primary representation of their materials. For them, the prioritization of document order is not perhaps so much of a problem.

Another aspect of the markup model relates to ownership of the materials of study. As shown in figure III, the Markup model assumes that the researcher uses markup to make assertions about the text – the tags make whatever the researcher wishes to say about the text explicit, and the markup that expresses these things is embedded within the text itself. For this reason, there is a history of viewing TEI markup as a task that should be done by the scholar him/herself. See, for example, this quote from an article written by one of the gurus of TEI and XML, Michael Sperberg-McQueen (highlighting added by me):

> A recent survey of literary scholars by the TEI work group on literary studies (Fortier 1991), for example, evoked several responses which denied that individual interpretations of a text (grammatical construction, thematic analysis, even explicit markup of orthographic sentences) should be expressible in a markup scheme. Such interpretative and analytic information, the respondents claimed, should be the task of the scholar, and not that of the encoder. **That the scholar and the encoder might be the same person seems not to be a universally shared assumption**; these respondents appear to assume a final division of labor between scholar and encoded which would, taken literally, absolve researchers from any responsibility for the quality, intelligence, or utility of the encodings they create -- a worrisome state of affairs. I propose here to hold scholars to a higher standard and make them responsible for the quality of their work. (Sperberg-McQueen 1991, p. 36)

Now, there *is* something fundamentally true and important in Sperberg-McQueen's remark, and the emphasis that markup can be an

activity that seriously engages the scholarly mind is an important one to make. However, this is also a statement made from a technical necessity; in XML there is no satisfactory way to readily enforce and then support the separation of tagging that represents one person's assertions about a text (say, the editors of the edition) from those that are unique to a particular scholar's view. This separation is surely essential – although many scholars rely on high quality editions prepared by someone else and shared over the entire scholarly community, their kind of interpretive work in fact layers new individual responses to that text on top of that shared textual base. For these researchers, XML and the markup model does not deal very satisfactorily with this separation between the common textual base and the personal interpretation.

In summary, then, the "text ownership" nature of the markup model seems to go against the work practices of many scholars – unlike the "conduit model." For certain kinds of work the clear separation between the text – the studied – and the study is actually highly desirable.

Object Manipulation Model

We now come to the fourth model I proposed at the beginning: the "**Object Manipulation Model**." I don't propose this because I think, out of these four models, we have a complete taxonomy of models, or anything grand like that. Instead, I mention it because it seems to suggest a radically different role for the computer in the analysis task than those proposed above, and one that seems to have been little exploited. There are other significantly different models – see Johanna Druker and McGann's work on the Ivanhoe game (Druker 2003) for a model that focuses on a different set of issues related to textual criticism.

The Object Manipulation Model is inspired, perhaps surprisingly, by the work of Douglas Englebart, whose computing work on his *Augment* system (Englebart 1962) way back in the 1960s and 70s explored for the first time many ideas such as the mouse (invented as a part of this project) that are now so commonplace that most of us cannot think of computers without them. However, as the name of Englebart's project suggests, he was not primarily interested in GUIs and user interface issues, but in how the machine could help people *augment their natural intelligence* to solve difficult problems. He recognized even then that computers could support humans in the work of developing models of the processes they are interested in. He believed that the computer, in the process, enhances

what humans could deal with in their own minds, without taking away the researcher's involvement in the material that is being studied.

What makes Englebart's core ideas hard to see is that they seem to involve no grand new way of doing things. Unlike the evangelical orientation of both HC's Transformation or Markup models, it isn't meant to consciously revolutionise the way we tackle questions; instead it is meant to support the ways we already use naturally. In his otherwise rather dry 1962 formal report about *Augment* Englebart has included an unusual element; a fictional description of a user (Joe) talking about how the system works and how it affects how he does his job:

> Joe begins, "You're probably waiting for something impressive. What I'm trying to prime you for, though, is the realization that the impressive new tricks *all* are based upon lots of changes in the *little things* you do. This computerized system is used over and over and over again to help me do little things – where my methods and ways of handling little things are changed until, lo, they've added up and suddenly I can do *impressive* new things." (Englebart 1962, p. 83)

So, what are the "little things" that scholars do when doing their research, and can any of them be usefully assisted by the computer? One aspect of research would be notetaking, and the gradual transformation of materials originally recorded as notes into a new research idea, and then into a scholarly article that describes it. The study by Brockman *et al* gives some examples:

- Many scholars produce extensive marginal notes on photocopies or personal copies of materials they read for their research – some write directly on the copy, others attach adhesive notes to the text.
- Many scholars use word processing programs to some degree for digesting or transcribing notes and for sketching out preliminary ideas in conjunction with their research reading.
- Each scholar has his or her own way of integrating the digital and handwritten research materials.

Brockman *et al*'s article points out something quite striking – that a researcher's digital materials almost always first go into paper before the analysis task really begins. Obviously, the conversion into paper format before personalising it throws away the digital advantage and the kind of flexibility that digital materials provide. Furthermore, many scholars use the word processor to re-enter and organise notes even though it is clearly a clumsy tool for this kind of task. The word processor focuses on words and layout rather than the manipulation and organisation of notes. Perhaps a tool that is more like a "note processor" rather than a "word processor" is the right one? A note processor would be designed using Englebart's Augment approach; supporting the user and helping them to mentally organise the notes they collect in the same way that a word processor doesn't write the text, but leaves all the important business of the content in the hand of the user.

Figure V shows schematically some aspects of this object manipulation model as it might apply to the process of annotation, note taking and organisation. Here, the text sits as a kind of common property, as it were, of a community – not really owned by anyone in the sense that the markup individual owns his markup on the text. It shares this characteristic with both the Conduit and Transformation models. Furthermore, the texts in the figure will not only be primary texts, but perhaps secondary resources as well, and in this way, the model is again more like the Conduit model, and less like the Transformation model.

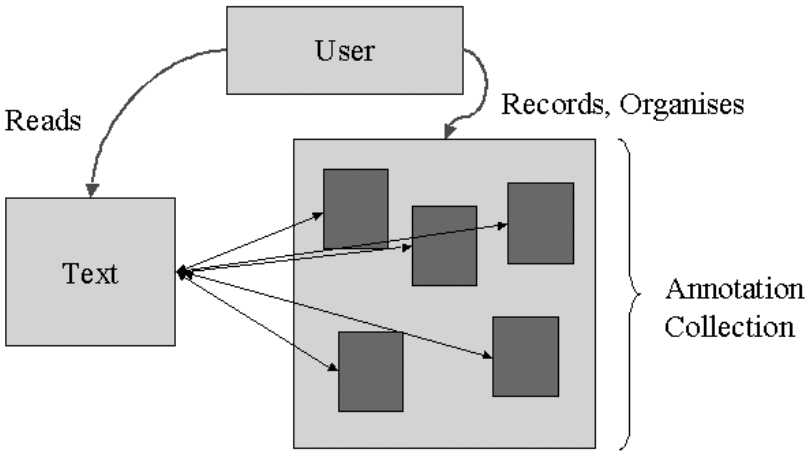

Figure V: Object Manipulation Model: notes about texts

A number of our projects at KCL have what might first appear to be an "annotatory air" about them. The *Durham Liber Vitae* project, mentioned earlier, shows an annotation character in the aspect of the project that wishes to associate historical persons with names in the manuscript. We found a complex formal structure that related materials together (described in Bradley 2005), and the making of this structure formal and explicit strengthened the project. However, like the *Rossetti Archive* the resulting object ended up being, from the point of view of the users, an example of a complex conduit model object. Only the project team members could own the person/annotation material, create and modify it, and change the structure if a better one emerged at some point.

To support annotation and note manipulation for a resource user rather than the resource builder required a significantly different approach, and we began to see the issues more clearly when we undertook an informal *use case analysis* of annotation and note manipulation. First, an annotation would need to be created, and attached to the current spot in the text when the user thinks s/he wants to create one. Questions still to be resolved here involve what is meant more formally by the "current spot". Annotations (like writing in the margin) must be visible each time the user returns to view the same spot in the source. Tools like the University of Virginia's INote (INote 1998) provided one interesting approach to these issues, but did not really take the annotation issue much further than this. A second set of use cases revolved around helping the note creator organise a set of notes into a structure that helped him/her mentally structure and organise them. Here tools that can inspire some thinking comes from the Social Sciences – Nud*ist, NVivo and Atlas.ti (see discussions about Nud*ist and Atlas.ti in the context of research in Alexa 1999 and Kelle 1997). Finally, a set of use cases emerged that dealt with the organising and presenting of digital annotations as public objects. Here, the work of computer scientist Carol Marshall who examined some of the significance of handwritten annotations in books (Marshall 1998), and explored the implications on research oriented reading when a tablet computer with annotation support is used (Marshall et al 1999) was useful.

At KCL we have begun exploring some of the issues of annotation support for research users in the Online Chopin Variorum Edition (OCVE) project (OCVE 2004). Space here does not permit me to describe the work in any detail (although there is some presentation of it in Bradley 2004), but during Phase II of the project we expect that the preliminary work done there to support user annotation and note organisation will be

expanded significantly.

Summary and Conclusions

Established ways of thinking about computing and research in the humanities as developed by the Humanities Computing community have, while producing some great successes, not much affected the research processes of most in the humanities. In this article I have shown that the established models of thinking about computers shown by most in the humanities, and that used by many of the HC community do not overlap well. Perhaps it is time to take a step back and re-evaluate these models in light of an understanding of how humanities research is primarily conducted. Based on this work, we will be in a better position to develop a model of the role of computers that does more to support humanities research. While the Object Manipulation model I described above provides one possible alternative, work by others such as Druker and McGann on things like the Ivanhoe game provide other exciting possible models.

Works Cited

Alexa, M and Zuell C (1999). *A Review of Software for Text Analysis*, in series ZUMA Spezial, Band 5, ZUMA: Mannheim, Germany.

Bradley, John (2003). "Finding a Middle Ground between 'Determinism' and 'Aesthetic Indeterminacy: a Model for Test Analysis Tools" in *Literary and Linguistic Computing* 18:2. Oxford: Oxford University Press

---(2004). "Highlighting the Past: Annotations of historical texts to support Humanities Scholarship", a presentation given at University of Kentucky, May 2004. Presentation powerpoint online at http://www.cch.kcl.ac.uk/legacy/staff/jdb/papers/kentucky.pdf (Referenced April, 2006).

---(2005) "Documents and Data: modelling materials for Humanities research in XML and relational databases" in *Literary and Linguistic Computing* 20:1 pp 133-51. Oxford: Oxford University Press.

Brockman, William S., Neumann Laura, Palmer , Carole L., Tidline, Tonyia J. (2001). *Scholarly Work in the Humanities and the Evolving Information Environment*, a report from the Council on Library and Information Resources. (December 2001).

Burrows, J. F. (1992). Tiptoeing into the Infinite: Testing for evidence of National difference in the Language of English Narrative. In S. Hockey & N. Ide (Eds.), *Research into Humanities Computing 4 Selected papers from the ALLC/ACH Conference Christ Church*, Oxford (Vol. 4). Oxford: Clarendon Press.

Digital Resources for the Humanities website, at http://www.drh.org.uk/ (referenced November 2004).

Drucker, Johanna (2003). "Designing Ivanhoe" in *Text Technology* 12:2, McMaster University.

Durham Liber Vitae Project, The. Project web site (2003). Online at http://www.kcl.ac.uk/humanities/cch/dlv/ referenced November 2004.

Early English Books Online. (pub date not provided), Oxford: Chadwyck-Healey. Online at http://eebo.chadwyck.com/home (referenced November 2004).

Englebart, Douglas C. (1962). *Augmenting Human Intellect: A Conceptual Framework*. Menlo Park CA: Stanford Research Institute Report. Now online at http://www.bootstrap.org/augdocs/friedewald030402/augmentinghumanintellect/AHI62.pdf (referenced November 2004)

Inote: An Image Annotation Tool in Java (1998) online at http://www.iath.virginia.edu/inote/ (referenced November 2004).

Kelle, U. (1997) 'Theory-Building in Qualitative Research and Computer Programs for the Management of Textual Data' in *Sociological Research Online*, 2:2 online at http://www.socresonline.org.uk/2/2/1.html (referenced November 2004).

Leon, Pat. (2004) "Early English Books Online: The Holy Grail of online

resources?" in *Arts and humanities data server Newsletter*, Autumn/ Winter 2004.
Marshall, Catherine C. (1998). "Towards an Ecology of Hypertext Annotation". *Hypertext* 98, pp 40-4).
Marshall, Catherine C, Golovchinsky, G., Price, M., Bill Schilit, W. (1999), "Introducing a digital library reading appliance into a reading group" , *Proceedings Digital Library Conference* 1999.
McGann, Jerome (2000). *The Complete Writings and Pictures of Dante Gabriel Gossetti: A hypermedia research archive*. IATH University of Virginia. Online at http://www.rossettiarchive.org/ (referenced, November 2004).
Online Chopin Variorum Edition Pilot Project. (2004) Online at http://puffin.cch.kcl.ac.uk/ocve/ (accessed November 2004).
Ott, Wilhelm (2000), 'Strategies and Tools for Textual Scholarship: the Tübingen System of Text Processing Programs (TUSTEP)' in *Literary & Linguistic Computing*, 15:1 Oxford: Oxford University Press pp. 93-108.
Potter, Rosanne., (1980) "Literary Criticism and Literary Computing: The Difficulties of a Synthesis", *Computers and the Humanities* 22 (1980) 91-97.
---. ed. (1989) *Literary Computing and Literary Criticism: Theoretical and Practical essays on Theme and Rhetoric*. Philadelphia, 1989.
Siemens, Ray.G. "Disparate Structures, Electronic and Otherwise: Conceptions of Textual Organisation in the Electronic Medium, with Reference to Electronic Editions of Shakespeare and the Internet." *Early Modern Literary Studies* 3.3 / Special Issue 2 (January, 1998): 6.1-29 <URL: http://purl.oclc.org/emls/03-3/siemshak.html>.
Siemens, Ray.G. ,Toms, Elaine, Sinclair, Sefan, Rockwell, Geoffrey, Siemens, Lynne (2004), *The Humanities Scholar in the Twenty-first Century: How Research is Done and What Support is Needed*, abstract in program for ALLC/ACH 2004 conference: Gothenburg, 2004.
Smith, John B. (1978) "Computer Criticism" reprinted in Potter, Rosanne G. (ed) *Literary Computing and Literary Criticism: Theoretical and Practical Essays on Theme and Rhetoric*, Philadelphia: University of Pennsylvania Press, 1989.
Sperberg-McQueen, C.Michael. (1991), "Text in the Electronic Age: Textual Study and Text Encoding, with Examples from Medieval Texts", *Literary and Linguistic Computing* 6:1.
---.and Bernard, L., *Guidelines for Electronic Text Encoding and Interchange (TEI P3)*, ALLC, ACH, ACL 1994
Wulf, William (1995). "Warning: Information Technology Will Transform the University". *Issues in Science & Technology* 11 (Summer): Dallas: University of Texas pp. 46-52.