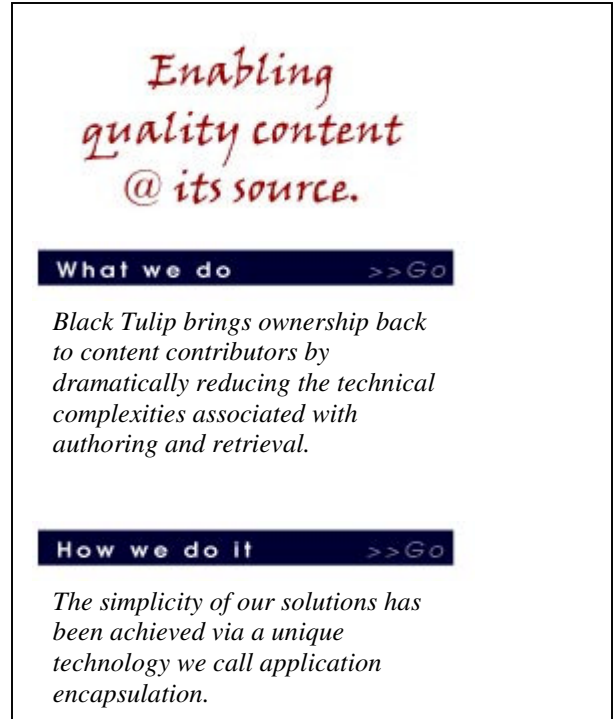


Why Automate Context Search?

Webster's definition of context is: *The part of a text or statement that surrounds a particular word or passage and determines its meaning.*

How do we determine the context of a particular word in a document? We need to read the document and interpret what is the context for each content area. This historically has been a tedious manual task and thus has been done only for specialized needs. Today, context search engines still require the user to determine what context you will attach to documents. As documents are saved, a pop-up window requires the publisher to fill in a "meta-data" abstract for that document, to be saved in a database that will allow the document to be retrieved through a search tool. Besides the time consuming process of creating "meta-data", the more significant limitation of this methodology is the lack of consistent interpretation of context between human interpretation and the creation of "meta data". The reality is that different people will input different "meta data" for the same document. This meta-data is now called 'context'. Thus, the lack of consistent interpretation of a document and its stored meta-data remains both a flawed and time-consuming process. The result is ineffective search and retrieval within a desired context, and unhappy users. The marketplace has thus resorted to the extensive use of keyword only searching which is the brute force approach, giving volume results that must be interpreted by the user. The true solution to unstructured text retrieval is to automate the creation of meta-data to allow context searches, and store them in a database so that any updates to the document will automatically update the meta-data. By alleviating the need for manual action of "meta data" creation, the process becomes consistent and the manual time spent is completely eliminated. Meta data is stored more accurately, and the resultant searching is far more efficient for the user. The ability to automate this meta-data creation process is both significant and difficult. How can this be accomplished?

By examining a document we can reveal the layers that build contextual meaning. At the highest level, the title of a document indicates the first level of context. Each Chapter title also indicates the context for the specific chapter. Looking closer, each paragraph header or title indicates a more specific context of the content within. Even within



*Enabling
quality content
@ its source.*

What we do >>Go

Black Tulip brings ownership back to content contributors by dramatically reducing the technical complexities associated with authoring and retrieval.

How we do it >>Go

The simplicity of our solutions has been achieved via a unique technology we call application encapsulation.

paragraphs, context can be identified in sections of content.

To better explain the levels of context, we can analyze this simple example. At the highest level of document title, we can see that the overall context is "Enabling quality content @ its source". At this level the rest of the document would now be content. At the next level we begin with the section starting "Black Tulip brings ...". This paragraph is in context of "What we do" and the paragraph starting with "The simplicity of ..." is in context of "How we do it".

In this simple example we found three instances of context within three sections of content. This example is a simple one, but as documents increase in length and volume, manual interpretation is time consuming, inconsistent, and unreliable and much of the true context can easily be lost. The problem here is then compounded as documents are revised and new context is added but not recognized. Only by automating this process can use of context searching of unstructured documents be a truly valuable tool for any organization.