

## Enterprise Search: What You Must Know about Information Retrieval and the “Google Effect”

© Stephen E. Arnold, Postal Box 320, Harrod's Creek, KY 40027  
Email: [sa@arnoldit.com](mailto:sa@arnoldit.com) – Voice: 502 228 1966



**ARNOLD  
INFORMATION  
TECHNOLOGY**

*Postal Box 320  
Harrod's Creek  
Kentucky 40027*

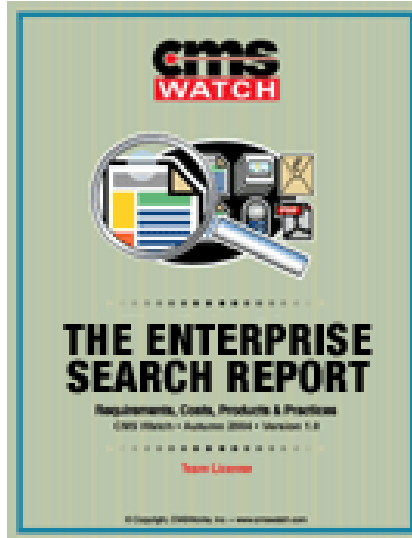
**Management consulting and  
strategy**

“The Google Legacy” became  
Available earlier this month  
Order at [www.infonortics.com](http://www.infonortics.com)

“The Enterprise Search Report”  
2nd edition now out. Order at  
[www.cmswatch.com](http://www.cmswatch.com)

Additional information at  
[www.arnoldit.com/sitemap.html](http://www.arnoldit.com/sitemap.html)  
Contact: [sa@arnoldit.com](mailto:sa@arnoldit.com)

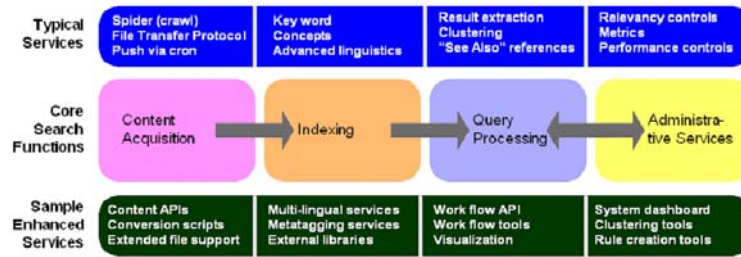
Limited Access Document



# 3

## The Parts

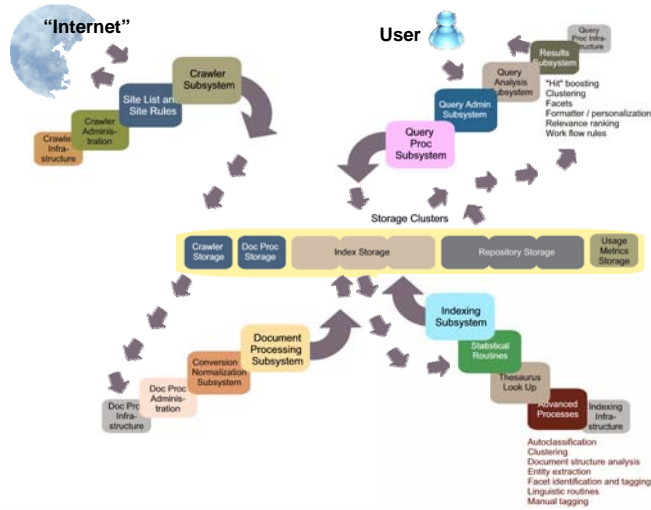
## How It Works: Search “Sandwich”



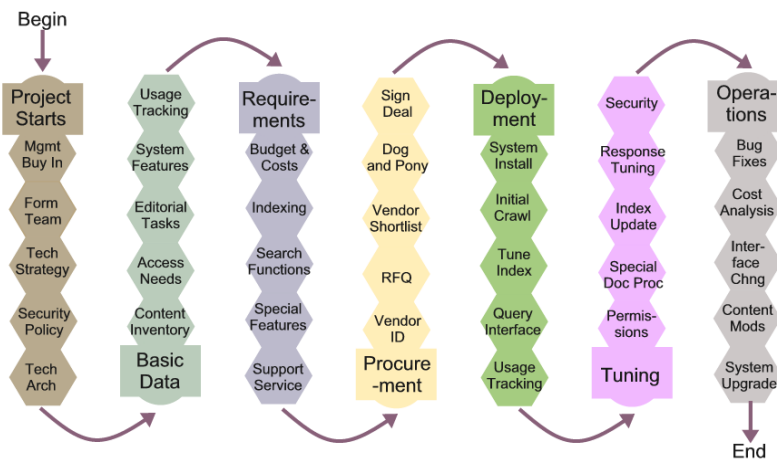
## Search... Multiple Subsystems

- **Spider:**
  - Locate content and changed content
  - Copy it to the search system
- **Index:**
  - Identify words
  - Extract / assign metadata
- **Query processor:**
  - Parse user's query
  - Process link for stored search
- **Administration**

# Search: One Machine or 10,000



# Search Itself: Not Well Understood

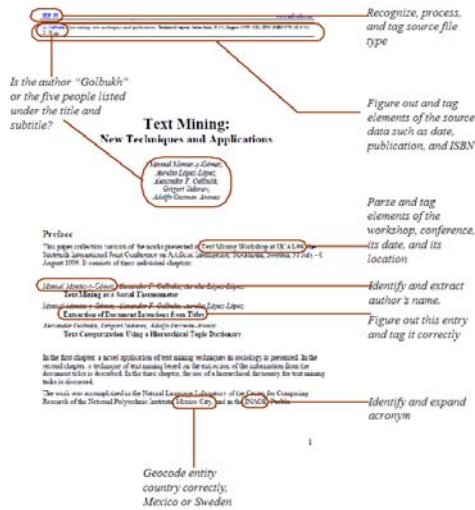




To This...

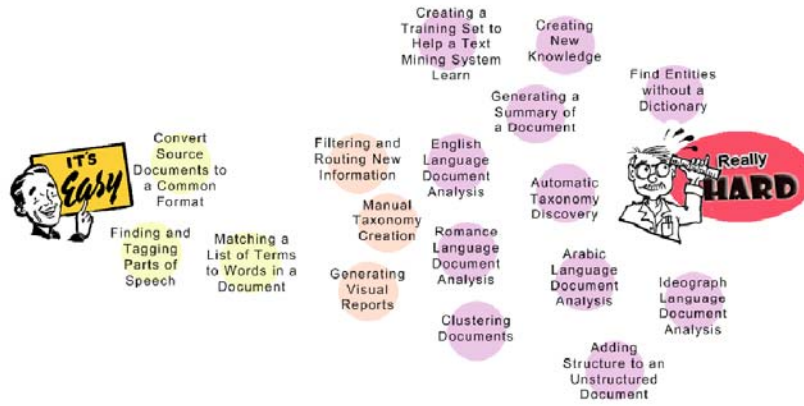
```
C:\PROGRESS\WRK\MSMQTest\ProductPriceTest.xml - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites History Mail Print Edit Discuss
Address http://C:\PROGRESS\WRK\MSMQTest\ProductPriceTest.xml
<description>urn:schemas-biztalk-org:BCPSoftware:ProductPriceUpdate</description>
</document>
</manifest>
</header>
<body>
  <ProductPriceUpdate xmlns="urn:schemas-biztalk-org:demo/ProductPriceUpdate.biz">
    <Transmission>
      <store_code>15596</store_code>
      <transmission_date>2000-05-23</transmission_date>
    </Transmission>
    <batch>
      <batch_no>000987</batch_no>
      <effective_date />
      <week_no>01</week_no>
      <transmission>1</transmission>
    </batch>
    <ProductPrice>
      <ean_codes>5010067301509</ean_codes>
      <Product>
        <product_code>136790</product_code>
        <tuc>03010067333107</tuc>
        <description>SILV SP CASTER SUGAR</description>
        <product_size>1KG</product_size>
        <short_description>SUGAR 1KG</short_description>
        <retail_units>10</retail_units>
        <VAT_code>A</VAT_code>
        <commodity_group>0</commodity_group>
        <supplier_code>10003</supplier_code>
      </Product>
    </ProductPrice>
  </ProductPriceUpdate>
</body>
Done My Computer
```

Search ... No Longer Enough





# What's Easy? What's Hard?



**Search is not yet a  
slam dunk**