



Search in the Wind

Heather Hedden

Taxonomy Manager, First Wind

Background of Heather Hedden, the Accidental Taxonomist

- Journalist > trade journal abstractor/indexer > controlled vocabulary editor
- Taxonomist career path:
 - self-employed at Hedden Information Management
 - taxonomist at enterprise search software startup Viziant
 - taxonomy consultant with Earley & Associates
- Teach online continuing education course at Simmons College GSLIS
- Wrote *The Accidental Taxonomist* (Information Today Inc., May 2010)
- Taxonomy manager at First Wind working on enterprise document management

Thus, the Accidental Taxonomist has stumbled into a new industry.

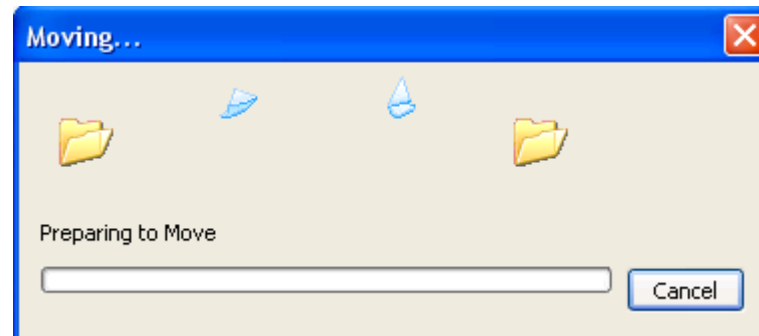
About First Wind

- An independent wind energy company focused on the development, financing, construction, ownership and operation of utility-scale wind projects
- Based in Boston
- 7 operating projects in four states: Maine, New York, Utah and Hawaii
- 4 new projects in various stages of construction
- Capacity to generate up to 504 MW of energy every hour
- About 225 employees in offices spread across the country



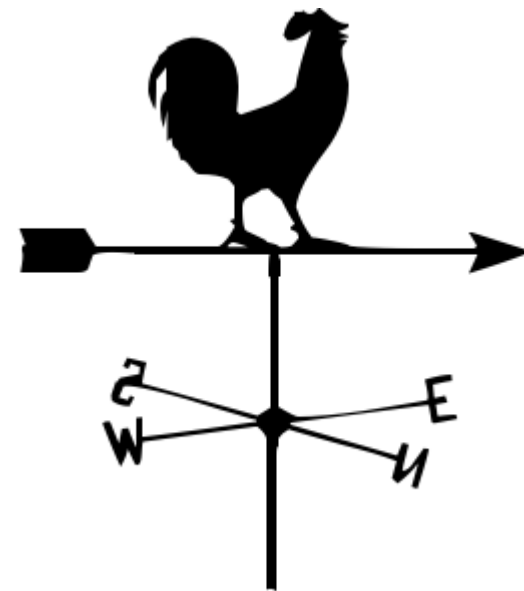
Challenge: Findability in a Moving Target

- It's difficult to find documents as they are blowing around.
- Documents in an enterprise tend to move around.



Challenge: Findability in a Moving Target

- As wind is unpredictable and inconsistent, so is the location and movement of documents.



Challenge: Findability in a Moving Target

- Even a content management system (CMS) or SharePoint cannot always hold documents in structured folders.

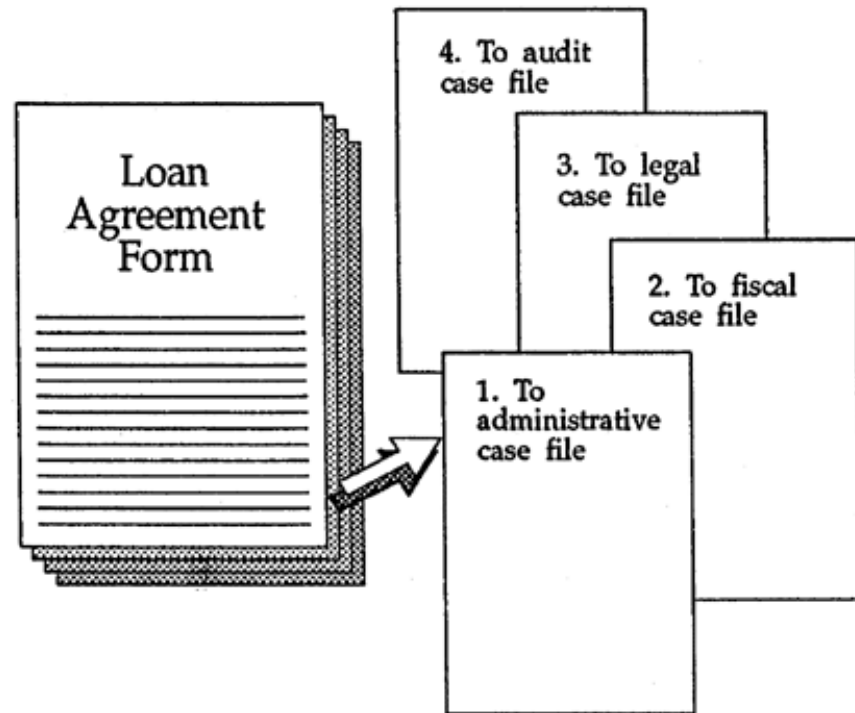
Because:

- There are multiple ways to classify.
- People like to move documents around to where they are convenient: sometimes copying, sometimes simply moving (without telling others).



Challenge: Findability in Hierarchical Folders

- In some departments, people file documents in some folders and in other departments, people file similar documents differently.
- The same document in multiple folders might not be the same version.
- A pre-set hierarchical folder structure might not be followed.



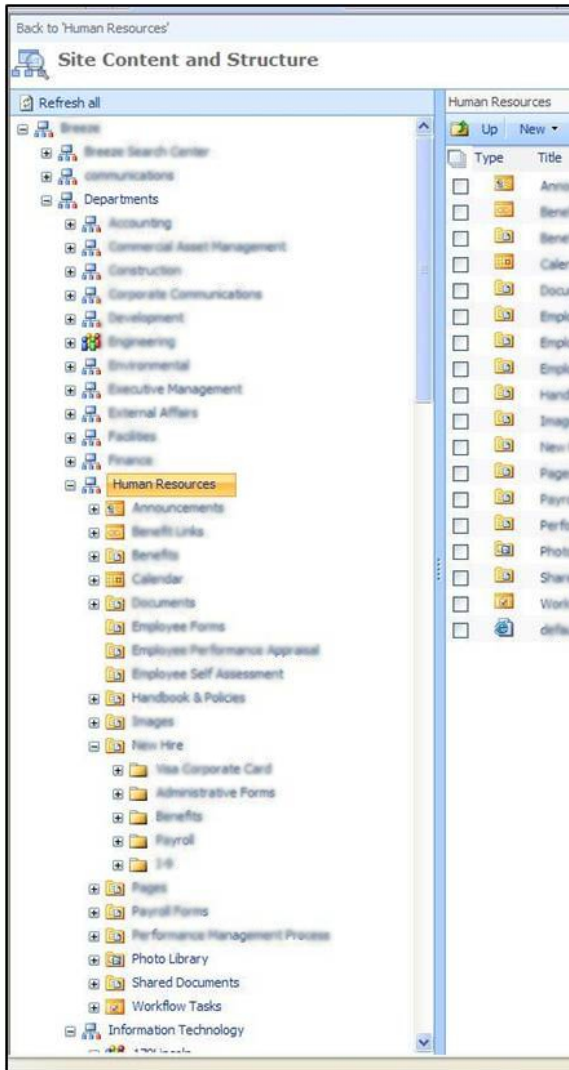
Challenge: Findability and Metadata Tagging

Metadata of descriptive terms can overcome the problem of inconsistent filing of documents in folders.

- However, enterprise employees adding documents to the system will not consistently apply metadata topic tags.
- A taxonomy for manual tagging or indexing requires willing and trained human indexers/catalogers.

Update Metadata	
Title	CuCMS v1.2
Creator	ALEXANDER VIGDOR
Keywords	
Description	An exhaustive reference manual for CuCMS Version 1.2, including chapters dedicated to installation, customization, use and technology references.
Hide Advanced metadata	
Publisher	
Contributor	
Date (yyyy-mm-dd)	2003-10-16
Type	Text
Language	English
Rights	
Relation	
Coverage	
Format	text/xml
Identifier	/docs/manual1_2/index.xml
Update Metadata	

Challenge: Findability in a Moving Target



- Hierarchical folders are important.
- Manual tagging is important.
- *But* don't count on such means alone.
- **Search** may be needed to find the documents that are moving around, inconsistently filed, or lacking metadata tagging.



Search and Taxonomy

A **taxonomy** (controlled vocabulary) enhances search:

1. A controlled vocabulary gathers synonyms, acronyms, variant spellings, etc.
 - Documents not missed due to use of different words (e.g. **Renewables** or **Alternative energy**)
 - Greater recall
2. A search restricted on the controlled vocabulary retrieves concepts, not just words.
 - Documents excluded for mere text-string matches (e.g. **Transmission** for electricity, not for automotive parts)
 - Greater precision

Search in the Wind and in Enterprise Content

- Like wind, just because you cannot see the desired content, does not mean it is not there.
- Like wind (with other conditions), too much of it can impair visibility.
- A well-designed and well-positioned search system, like the turbines, catches the most.



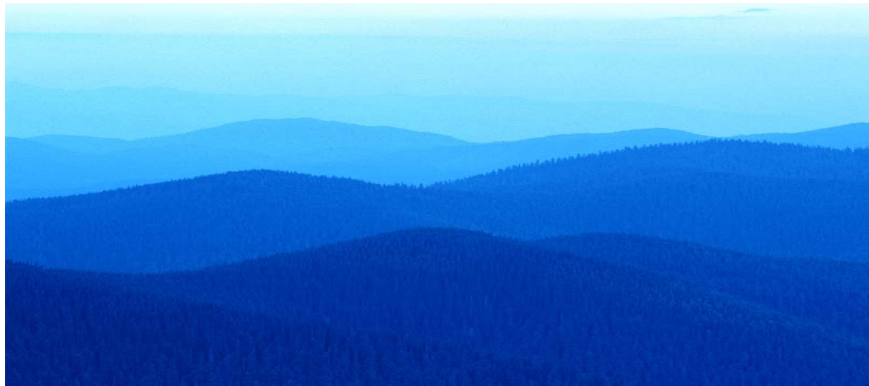
Both Wind Energy and Enterprise Search

1. Get more at higher elevations
2. Choose the right technology to capture the most
3. Consider the local community
4. Monitor and adjust
5. Follow through with the interconnection/interface to deliver to the user
6. Implement governance and compliance
7. Invest for the long-term

1. Get More at Higher Elevations

Better results are at higher elevations.

- *Elevate* the content for search engine capture by applying metadata to content items with a taxonomy.
 - Controlled vocabulary improves precision and recall.
 - A hierarchical or faceted taxonomy display helps guide the user.
- If using file path folders, don't go too many levels deep.



2. Choose the Right Technology

Auto-classification using taxonomies could be either:

- Machine-learning (algorithm-based) and training documents
- Rules-based categorization

For auto-classification options, consider:

- The types of documents and nature of the text
- The uniformity of document types
- File formats and metadata
- Existing sample documents for training if using machine learning
- The type of classification desired



3. Consider the Local Community

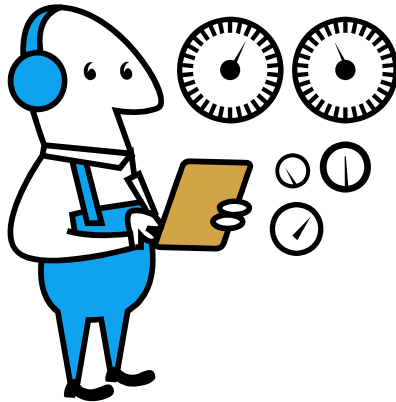
- Create custom taxonomy(s) for your users.
- Off-the-shelf are not appropriate for enterprise taxonomies.
- Interview the stakeholders to determine their needs.



4. Monitor and Adjust

Taxonomies for auto-classification can and should be tuned.

- If based on training documents, adjust the set by adding, removing, and perhaps even editing training documents.
- If based on rules, check and edit the rules.
- In all cases, add or delete synonyms (nonpreferred terms) as needed.



5. Interconnect/Interface for End-user Delivery

- The user interface should offer various features for varied users:
 - Search box
 - Hierarchy browse
 - Boolean searching options (and/or choices)
 - Advanced search
 - Limiting and/or sorting by facets
- Remember the point of serving your users.



6. Governance and Compliance

- Authority is needed.
 - Who owns the taxonomy, manages it, and responds to issues.
- Rules must be followed in taxonomy maintenance and usage.
 - Just as folders morph, so could the taxonomy.
- Create and follow-up on policies that support auto-classification.
 - Standardized filenames, document titles, headings, parent folders support search.



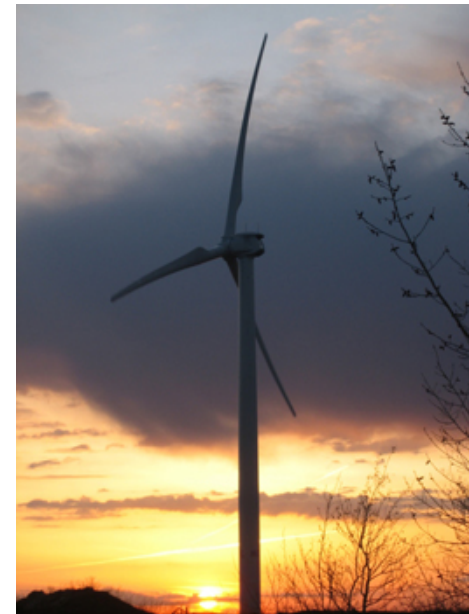
7. Long-term Investment

- Content management and findability strategy is an important capital investment.
- Time and money will be saved in the long-term, when time is not lost in searching for difficult-to-find documents.
 - Invest in a good search/auto-classification system.
 - Invest in hiring a taxonomist or consultant.
 - Be prepared for the process to be thorough and lengthy.



Conclusions

- In the enterprise, documents/content tends to:
 - Move around
 - Get put into folders inconsistently
 - Lack sufficient metadata indexing
- Enterprise search technologies can help.
- Taxonomy further enhances search in precision and recall.
- Search systems require:
 - A customized taxonomy
 - Tuning to the taxonomy
 - A full-featured user interface
 - Governance
 - Investment of time and money





firstwindSM

CLEAN ENERGY. MADE HERE.

Heather Hedden
Taxonomy Manager
First Wind

179 Lincoln Street, Suite 500
Boston, MA 02111

O: 617.960.9596

C: 617.875.0393

hhedden@firstwind.com

www.firstwind.com