TeamSite Taxonomy

Overview
As organizations seek ways to efficiently leverage the ever-increasing amount of information available internally and on the web, it is becoming crucial to develop methods that make content “intelligent”. By making content intelligent, organizations can find and repurpose content to relevant audiences automatically. This ability is key to obtaining maximum value from search, personalization, syndication, and portal applications. Autonomy TeamSite Taxonomy, a native extension of Autonomy TeamSite, automates the process of making content intelligent by creating rich, descriptive metadata from articles, blogs, emails, social content, audio, video, documents, or web pages. TeamSite Taxonomy automatically categorizes content items and then indexes them so that they can be linked to other relevant information. This next-generation content intelligence solution enables content creators to categorize their content interactively and allows developers to create procedures for recognizing and classifying large volumes of content automatically.

Powered by Autonomy’s Intelligent Data Operating Layer (IDOL), TeamSite Taxonomy can suggest appropriate metadata for content based on IDOL’s conceptual understanding of information. Through either a semi- or fully automated approach, TeamSite is able to add intelligence to content for use later in run-time search, personalization, syndication, and portal applications. TeamSite Taxonomy lets you rapidly utilize pre-built taxonomies that can be combined with your corporate taxonomies or easily customized to meet company- and industry-specific requirements. Each out-of-the-box taxonomy is based on industry standards, and built using the same conceptual analysis that provides the highest level of accuracy available through Autonomy IDOL. The range of pre-built taxonomies includes:

- Biotechnology
- Civil / Criminal Procedure
- Defense
- Electronic, Mechanical and Structural Engineering
- Emergency Services
- Epidemiology
- Financial Services
- Homeland Security
- Human Resources
- Information Technology
- Petrochemical
- Pharmaceutical
- Sales & Marketing

A Unique Approach
Autonomy takes a unique, holistic approach to web-based taxonomy generation that strikes the optimum balance between automating processes and support for manual refinement, meaning automation is not an either/or choice.

By forming a conceptual understanding of content associated with one or more websites, IDOL automatically generates taxonomies and instantly organizes data into a familiar child/parent taxonomic structure. IDOL automatically identifies names and creates each node based on an understanding of the concepts within the data set as a whole. It eliminates laborious data-entry work, creating enterprise level taxonomies with unmatched speed.

Datasheet
Highlights
- Jump-start classification initiatives with tested and proven pre-built taxonomies
- Easily customize industry-standard taxonomies to achieve specific business objectives
- Analyze over 1,000 document types
- Reduce overheads and increase productivity
- Decrease manual processes with a technology that automatically adapts to changing uses of language
- Benefit from the system’s integration to over 400 repositories
Autonomy's data agnostic classification capabilities enable organizations to build taxonomies based on any type of data: unstructured, semi-structured, structured, audio, and visual. Conceptual operations can be used as starting points for the process, which allows clusters to be automatically broken down into taxonomies. Alternatively, content can be used to find similar information about a subject, which can then be broken down into a new taxonomy.

Autonomy's understanding of concepts is based on the entire corpus of content analyzed by IDOL. Therefore, as new vocabulary and terms are introduced, or their meaning changes, they will automatically be updated in the system, without the need for the taxonomy to be manually updated. An example is the change in meaning of the term “tweet” after the wide adoption of Twitter. With its original meaning being a “weak chirping sound”, the phrase took on a new meaning overnight. Autonomy IDOL was able to immediately understand the change in meaning and see that it was being used in relation to social media, and not in relation to a chirp or bird. This meaning-based technology eliminates problems that can derive from a purely manual approach to classification:

- **Descriptive Inconsistency:** Without a built-in, conceptual understanding of information, content can be tagged inconsistently and without strict adherence to corporate taxonomy
- **Interoperability of Tagging:** If two organizations are going to interoperate and apply the same meaning to the same tags, they have to explicitly agree upon their classification schemes in advance
- **Scale:** As the number of tags increases, so does the cost of labor and the likelihood of misclassification through human error
- **No Learning Ability:** Definitions are static once entered into the system. They are not able to learn as words change their meaning or as new words are added to the language

TeamSite Taxonomy can be configured to automatically apply metadata to assets in both a semi-automated and fully automated manner. The fully automated mode provides organizations with a robust, efficient, and rapid mechanism for applying metadata to either current or legacy content, incoming syndication feeds, or disparate corporate assets. TeamSite itself leverages TeamSite Taxonomy automation within workflows to assign metadata to assets without necessarily involving humans in the process.

**The Best of Both Worlds**

It is also essential, however, to offer classification experts the freedom to edit and refine taxonomies. Autonomy offers a sophisticated array of controls to facilitate the manual maintenance of enterprise taxonomies. TeamSite offers a comprehensive management interface for subject matter experts — the people who know and use taxonomies — and knowledge engineers — the people responsible for the controlled vocabularies used within the enterprise. It enables cross-functional collaboration between these experts, allowing improvements to be made in real-time so that the enterprise can adapt in response to changing business requirements.

The intuitive user interface provides familiar features such as drag-and-drop of documents and categories across the taxonomy, and query tools that quickly locate documents and keywords to define categories. Onscreen visualization of document distribution lets editors and publishers track, assess and improve the assignment of documents to categories. In addition, Autonomy enables existing taxonomies to be imported, including URLs, file paths, Autonomy Taxonomies, as well as corporate and other third-party taxonomies.

**About Autonomy**

Autonomy Corporation, an HP Company, is a global leader in software that processes human information, or unstructured data, including social media, email, video, audio, text and web pages, etc. Autonomy’s technology manages and extracts meaning in real-time from all forms of information, both unstructured and structured, enabling companies to leverage their data assets. Autonomy’s product portfolio helps power companies through enterprise search analytics, business process management and OEM operations. Autonomy also offers information governance solutions in areas such as eDiscovery, content management and compliance, as well as marketing solutions that help companies grow revenue, such as web content management, online marketing optimization and rich media management. Autonomy’s solutions are used by more than 25,000 customers including 87 of the Fortune 100, 10 of the top 10 financial services firms, 75% of the global 100 law firms, 9 of the top 10 pharmaceutical companies and many government agencies. Over 400 of the world’s leading technology companies embed Autonomy’s technology in their products. Autonomy also owns the largest private cloud of diverse data, with 31 Petabytes of information.

Please visit [www.autonomy.com](http://www.autonomy.com) to find out more.