


Inxight Software, Inc.

U.S. | 500 Macara Avenue, Sunnyvale, CA 94085
 Phone: 408.738.6200 | Fax: 408.738.6203

Europe | Centaur House, Ancells Business Park, Ancells Road
 Fleet, Hampshire GU51 2UN U.K.
 Phone: +44 (0) 1252 761314 | Fax: +44 (0) 1252 761315

PRODUCT DATASHEET

Inxight SmartDiscovery provides the fastest and most accurate entity extraction to automatically identify and categorize named entities in electronic text. This allows quick and efficient discovery and retrieval of relevant information, enabling better informed business decisions.

Inxight SmartDiscovery™ entity extraction

Intelligent, Automatic Extraction of Relevant Entities from Text Data

Business success is tied directly to the speed of access and effective use of relevant information. Yet, most companies have no way of knowing what portion of their information assets is relevant to a given situation thus they have no way of effectively using these assets. How does a company identify the “who”, “what”, “where”, and “when” discussed inside email, computer files, presentations, and analyst reports without opening each individual file? Inxight SmartDiscovery’s entity extraction solves this problem by automatically identifying and categorizing named entities in text data sources to speed the discovery and retrieval of relevant information. This enables companies to make better-informed business decisions.

Entity extraction — Discover the true value of your unstructured information

Entity extraction automatically creates meta-data reflecting what is *inside* a company’s files. It quickly and efficiently discovers the “who”, “what”, “where”, and “when” of each document to create metadata that is consistent and germane to each search request. The metadata generated by entity extraction is always relevant because it comes from the content itself! This contextual, content-driven metadata provides consistent and accurate information for use in information retrieval, routing, categorization, and data mining tasks — regardless of storage location or format.

SmartDiscovery’s entity extraction, unlike other solutions, requires no training or tuning. This powerful text-analysis tool automatically identifies twenty-five named entity types — out-of-the-box, including people, places, companies, dates, measurements, currency figures, and email addresses. It can also be extended to recognize customer-specific entity types such as SKU numbers or project names.

Leveraging Inxight’s deep linguistic understanding, entity extraction is able to discover entities based on patterns in text. It provides the ability to identify a company name the *first* time it appears in a data set without reliance on customer-provided lists or training sets.

SmartDiscovery’s out-of-the-box entity extraction capabilities enable you to:

- **Review** content and route documents appropriately, without reading the text.
- **Find** all references to products and people in customer service logs and emails, automatically and regardless of variations in format.
- **Create** link analysis or business intelligence applications that identify and monitor trends and events associated with people, places, dates, and companies.
- **Enhance** any data categorization solution by identifying the entities most relevant to a document.
- **Enable** precision searching by honing in on named entities.

Only SmartDiscovery with entity extraction provides this valuable metadata creation capability within a complete information discovery solution.



Inxight SmartDiscovery's entity extraction — Identifies “who”, “what”, “where” and “when” within unstructured documents

Without training sets or manually created rules, SmartDiscovery entity extraction generates metadata that is always relevant and enables discovery of novel entities. By normalizing and grouping together different variants of the same entity, SmartDiscovery removes the guesswork from search, data mining, and link analysis. High-level knowledge of what data your documents contain speeds information retrieval and alleviates doubt about your information. All this adds up to a compelling return on previous IT investments.

The screenshot displays the Inxight SmartDiscovery interface. The search results are filtered by 'merger'. A red box highlights the 'entity filter' section, which lists various entities categorized by type:

- Company / Organization:** Adobe Systems Inc (3), America Online (27), Aventis S.A. (22), Baker-Hughes Inc (7), Bank of America (16)
- People:** Gabriel Bell (2), Jordan Black (2), Jessica Butler (3), Abigail Coleman (4), Jose Diaz (5)
- Place:** Canada (1), Chicago (2), Clayton (2), Cleveland (1), Columbia (4)

Below the categories, there are radio buttons for 'must contain any' (selected) and 'must contain all', along with a 'filter' button.

SmartDiscovery entity extraction helps discover the key players in your unstructured data.

At Inxight, we are leveraging more than 20 years of Xerox PARC (Palo Alto Research Center) research as well as 70 patents, to focus exclusively on addressing the information discovery requirements of our customers. All of this research is evident in SmartDiscovery.

About SmartDiscovery

Inxight SmartDiscovery is the only complete enterprise solution for information discovery. Its comprehensive set of advanced text analysis tools includes search, entity extraction, fact finding, categorization, and visualization. SmartDiscovery's comprehensive metadata creation and search capabilities, for applications ranging from information retrieval to data mining to trend analysis is powered by Inxight's industry-proven linguistic analysis capabilities.



SmartDiscovery includes the following entity extraction functionalities:

Extraction and Discovery

- Extracts more than 25 entity types from documents and classifies them by types, such as companies, people, dates, places, measurements, organizations, currencies.
- Discovers new entities using linguistic and textual clues, based on SmartDiscovery's deep linguistic understanding of your data – both the content and the context.

The proposed merger between Mega, Inc. and CNA Systems, Incorporated, has been postponed, Mega CEO Joe Smith said in an analyst call. "CNA's 1st quarter revenue dropped by 32%, and they lost 23 million dollars," Smith explained. CNA Systems sources blame weak sales in China. CNA shares (CNAI) fell 47 percent to \$9.84 on May 12, the first trading day after the announcement.	
Company	Mega, Inc., CNA Systems, Incorporated
Date	May 12
Person	Joe Smith
Person Position	Mega CEO
Currency	23 million dollars, \$9.84
Measurement	32%, 47 percent
Country	China
Noun Group	proposed merger, analyst call, 1st quarter revenue weak sales, first trading day

Variant Identification and Grouping

Recognizes variant names as different forms of the same entity giving accurate entity counts as well as the location of *all* appearances of a given entity. For example, SmartDiscovery recognizes that the word "Smith", in the example, refers to the "Joe Smith" identified earlier and therefore groups them together as aliases of the same entity.

Canonical Form	Variant Forms	No. of Appearances
Joe Smith	Smith, Mr. Smith, J. Smith	2
CNA Systems, Incorporated	CNA, CNA Systems, CNAI	5

Normalization

Normalizes entities such as dates, currencies, and measurements into standard formats, taking much of the guesswork out of the metadata creation, search, data mining, and link analysis processes.

Entity Found	Normalization
May 12	05/12
\$23 million	23000000 USD
CNA Systems, Incorporated	CNA Systems, Inc.

Relevance Ranking

Ranks extracted entities with relevance-scores reflecting their importance to the document as a whole.

Customization

Can be extended to identify and extract custom entity types, such as project names, SKU numbers, drug names/composition, chemical names, or patent numbers.

About Inxight

Inxight Software, Inc., is the leading provider of enterprise software solutions for information discovery. Using Inxight solutions, companies can better discover, retrieve, and connect with the estimated 85 percent of corporate information not housed in a database or other structured format. Inxight is the only company that provides a complete, scalable solution to enable information discovery in all major languages. Customers include Air Products, Computer Associates, Factiva, HP, IBM, LexisNexis, Reuters, SAP, SAS, Thomson and Xerox. The company has offices throughout the United States and Europe. For more information, visit www.inxight.com or call 408-738-6299 (in the U.S.) or +44 (0) 1252 761314 (in Europe).

Copyright © 2004 Inxight Software, Inc. All rights reserved. Inxight is a registered trademark and Inxight SmartDiscovery is a trademark of Inxight Software, Inc. All other products are trademarks of their respective companies.

Technical Specifications

Operating Systems

- Microsoft Windows 2000, XP
- Sun Solaris 2.8

Language Modules

- Arabic
- Chinese (Simplified)
- English
- Farsi (Persian)
- French
- German
- Spanish

Contact Inxight about the availability of other languages.

Pre-defined Entity Types

Entity types can be customized and new categories may be added. Pre-defined categories vary by language.

- Address
- City
- Company
- Country
- Currency
- Date
- Day
- Financial Index
- Holiday
- Internet Address
- Month
- Measure
- Noun Group
- Organization
- Percent
- Person
 - Position
 - Given Name
 - Family Name
 - Suffix
- Phone Number
- Place
 - Regions
 - Political
 - Geographic Types
- Product
- Social Security Number
- State
- Time
- Time Period
- Vehicle
 - Make
 - Model
 - Color
 - VIN
 - License Plate
- Year