

## Industry Indexes and Information Infrastructure

### EXECUTIVE SUMMARY

The WAND® Industry Indexes comprise a comprehensive library of information about products and services traded worldwide. The Industry Indexes use state-of-the-art retrieval technology based on a data-rich electronic thesaurus and integrated knowledge management system.

The hub of the system is the WAND®-proprietary thesaurus of product and service types. These types are organized as a controlled vocabulary with a hierarchical structure and synonymous terminology. Types are also cross-referenced to a variety of international classification schema commonly used by industries worldwide. Principal among these schema are the Harmonized System (HS), the Standard Industrial Classification (SIC) and the North American Industry Classification System (NAICS).

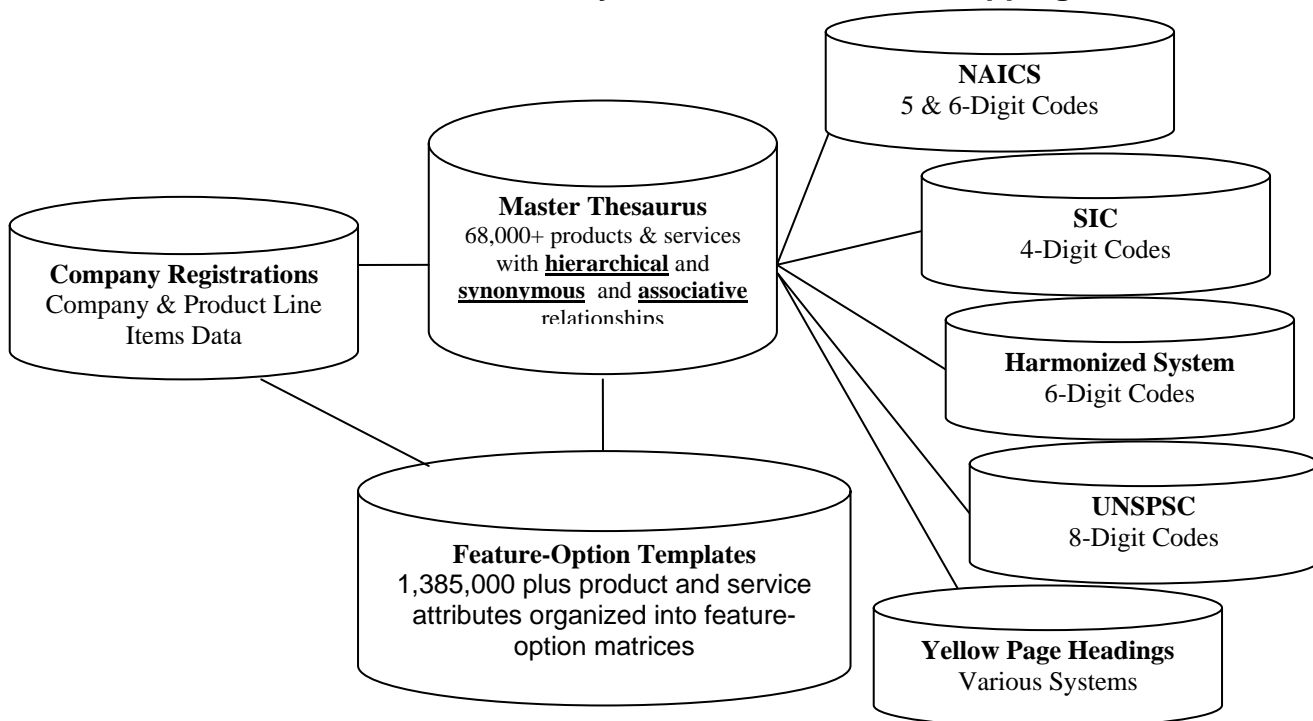
The thesaurus sits on top of a massive WAND®-proprietary database of detailed product and service attributes (technical characteristics). Attributes are organized into matrices of features and options providing an infrastructure for capturing and retrieving product information in a unified way. The thesaurus and attributes database facilitates user-friendly searching and accurate recall of those vendors and suppliers who precisely match highly specific procurement requirements.

A team of professional lexicographers, indexers and parametric developers has been working on building WAND®'s thesaurus and attributes database since 1995. The base build of the database was completed in 2001 and a continuous effort to revise and expand the database is now ongoing. The project team also uses numerous consultants and experts drawn from each industry sector to review terminology and attributes prior to certifying each Industry Index. The WAND®-proprietary data set represents a major barrier to entry in terms of time and cost for any competing information provider.

### WAND®'s INFORMATION INFRASTRUCTURE

The WAND® Information Infrastructure stores a complex web of inter-related product and services information.

#### WAND Taxonomy Structure and Cross Mappings



Spanish and French versions of NAICS are referred to as SCIAN. 6-digit codes are national and may be unique to each country.

At the hub of the system is WAND®'s proprietary thesaurus containing over 68,000 product and service types. This controlled vocabulary is built using a thesaurus construction software application compliant with the US national standard ANSI/NISO Z39.19-1993; "Guidelines for the Construction, Format, and Management of Monolingual Thesauri".

The thesaurus stores preferred names for products and services, plus relationships to many synonyms and variants forms. The terminology is also organized within poly-hierarchical structures, which provide drill-down menu-driven access.

Preferred product and service names are cross-referenced to many national and international commodity classification schema, such as the internationally used Harmonized System.

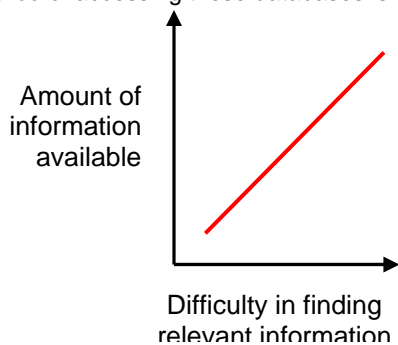
Preferred products and services are also linked to a massive database of attributes, which are organized into matrices of features and options. The matrix for any single product or service is known as a feature-option template and these templates provide a standardized method for capturing and retrieving highly specific product and service information.

## CONTROLLED VOCABULARY INDEXING

Individuals and organizations alike are experiencing one of the paradoxes of the information age: as we gain access to more and more data, it is becoming harder to find the particular information we need. Metaphorically speaking, society is drowning in data, yet thirsting for information. No where is this more evident than on the Worldwide Web.

### 3.1 Problems with Free Text Searching

Advances in information technology have enabled the proliferate growth of large databases. These information systems, on-line services and CD-ROM products; they are used by millions of people every day. The most common method of accessing these databases is a **textword** query, also known as a free text search.



A textword may be defined as any word that appears within the actual text of a database (this might be text confined to a specific data field in a structured database or the full text of an entire collection of documents in a text management and retrieval system). The inherent weakness of a textword query is that any particular word may mean several things and, conversely, any particular thing may be described by several different words. Many words used within a document are not key to, or indeed even part of the document's meaning.

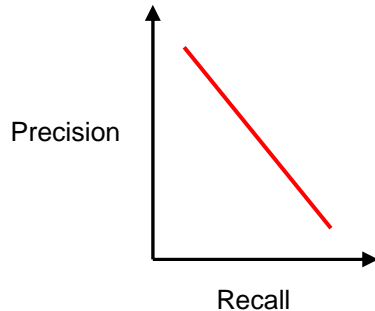
### 3.2 The Elasticity of Language

This phenomenon is what Dr. Peter Roget, author of the original Thesaurus of English Words and Phrases, eloquently dubbed "the elasticity of language." The elasticity of language has a significant impact on textword searches in large databases: It results in the user having to sift through masses of irrelevant data or, much more insidiously, in the user missing relevant information.

To illustrate Roget's language elasticity with a contemporary example, imagine that a medical researcher is searching for articles on AIDS. By submitting a query on the textword "aids," the researcher is likely to find many irrelevant articles, such as on hearing aids and Band-Aids. The researcher will probably then qualify the query to "aids and virus." This will improve the precision of the results; that is, of all the articles found, more of them are now likely to be about the subject the researcher had in mind. The downside is that the researcher has also diminished the recall of all possible articles on the subject, because the new query excludes any articles that happen not to have the word "virus" in the text. Acronyms and abbreviations also affect the query results: The user would miss relevant items if they did not also search on the full term "acquired immunodeficiency syndrome."

### 3.3 Precision and Recall

Study of the nature and limitations of textword searching has led information scientists to identify a natural trade-off between **precision** (the degree to which retrieved information accurately matches the intended subject of a query) and **recall** (the degree to which a query is able to retrieve all the relevant material).



If you want to retrieve every relevant item, then you will necessarily also have to sort through irrelevant items. If you only want to see relevant items, then you will also miss those articles that are relevant but which happened not to match all the search criteria.

### 3.4 Solving the Precision / Recall Dilemma

The precision / recall dilemma is an inherent limitation with textword searching, because it results from the innately elastic quality of natural language. Solving this dilemma led to the use of **controlled vocabularies**, also known as indexing languages, and the idea of concept-based retrieval rather than text-based retrieval.

The controlled vocabulary model involves a group of trained lexicographers and indexers resolving the elasticity of language within a database by developing and assigning standardized terminology within a thesaurus.

## ☰ TYPES OF SEARCH & BROWSE FUNCTIONALITY

WAND<sup>®</sup>'s data rich information infrastructure supports a variety of state-of-the-art search and browse functions.

**Hierarchical Browse** - drill down through an unfolding menu to hone in on the precise products or services

**Keyword Search** - search the entire thesaurus, be redirected from non-preferred synonyms to standardized, or preferred, products and services

**Attribute Search** - search for products or services that match highly specific criteria

**International Code Search** - search by SIC, NAICS or HS Codes to find cross-referenced products and services

## DETAILED EXAMPLE OF THE DEPTH AND BREADTH OF WAND®'s DATA SET

The **Agricultural Products Package** is just one of the industry indexes in the WAND® system. It contains products from such major groups as live animals, crops, and prepared food and drink.

- Over 2,500 products
- Over 3,400 synonyms
- Hierarchical menu with over 200 classes and subclasses
- Over 200,000 product attributes

### A keyword search on the word "lobster" yields 62 product names.

You can scan the alphabetically sorted list to the right. Latin genus-species names are often shown in parentheses after the common names because in the international fisheries trade, this is often the only way to verify the authenticity of the fish type.

Confused by the choices? Fortunately, WAND®'s drill-down menu system organizes products into hierarchies, which help users hone in on the specific product they desire.

#### Animal Products

Fishery Products  
Seafood  
Shellfish  
Crustaceans

#### Lobsters

Imitation Lobster  
Lobster Hepatopancreas  
Lobsters (Homarus spp.)  
Lobsters (Ibacus ciliatus)  
Lobsters (Jasus spp.)  
Lobsters (Metanephrops spp.)  
Lobsters (Nephrops norvegicus)  
Lobsters (Nephropsis aculeata)  
Lobsters (Panulirus spp.)  
Lobsters (Scyllarides spp.)  
Lobsters (Thenus orientalis)

American Lobsters (Homarus americanus)  
Blue Spiny Lobster (Panulirus inflatus)  
Brazilian Slipper Lobster (Scyllarides brasiliensis)  
California Spiny Lobster (Panulirus interruptus)  
Cape Spiny Lobster (Jasus Ialandii)  
Caribbean Lobsterette (Metanephrops binghami)  
Caribbean Spiny Lobster (Panulirus argus)  
Common Spiny Lobster (Palinurus elephas)  
Eastern Spiny Lobster (Jasus verreauxi)  
European Lobsters (Homarus gammarus)  
Flathead Locust Lobster (Thenus orientalis)  
Florida Lobsterette (Nephropsis aculeata)  
Gilchrist Spiny Lobster (Palinurus gilchristi)  
Green Spiny Lobster (Panulirus gracilis)  
Hawaiian Spiny Lobster (Panulirus marginatus)  
Imitation Lobster  
Japanese Fan Lobster (Ibacus ciliatus)  
Japanese Spiny Lobster (Panulirus japonicus)  
Lobster Appetizers  
Lobster Entrees  
Lobster Hepatopancreas  
Lobster Livers  
Lobster Meal  
Lobster Puffs  
Lobster Roe  
Lobster Soup  
Lobster Tomalley  
Lobsterettes (Metanephrops spp.)  
Lobsterettes (Nephropsis aculeata)  
Lobsters  
Lobsters (Homarus spp.)  
Lobsters (Ibacus ciliatus)  
Lobsters (Jasus spp.)  
Lobsters (Metanephrops spp.)  
Lobsters (Nephrops norvegicus)  
Lobsters (Nephropsis aculeata)  
Lobsters (Panulirus spp.)  
Lobsters (Scyllarides spp.)  
Lobsters (Thenus orientalis)  
Mud Spiny Lobster (Panulirus polyphagus)  
Natal Spiny Lobster (Palinurus delagoae)  
Northern Lobsters (Homarus americanus)  
Norway Lobster (Nephrops norvegicus)  
Ornate Spiny Lobster (Panulirus ornatus)  
Painted Spiny Lobster (Panulirus versicolor)  
Pink Spiny Lobster (Palinurus mauritanicus)  
Pronghorn Spiny Lobster (Panulirus penicillatus)  
Red Spiny Lobster (Jasus edwardsii)  
Ridged Slipper Lobster (Scyllarides nodifer)  
Rock Lobsters (Jasus spp.)  
Royal Spiny Lobster (Panulirus regius)  
Scalloped Spiny Lobster (Panulirus homarus)  
Slipper Lobsters (Ibacus ciliatus)  
Slipper Lobsters (Scyllarides spp.)  
Slipper Lobsters (Thenus orientalis)  
Smoothtail Spiny Lobster (Panulirus laevicauda)  
Southern Spiny Lobster (Jasus novaehollandiae)  
Spanish Slipper Lobster (Scyllarides aequinoctialis)  
Spiny Lobsters (Panulirus spp.)  
St. Paul Spiny Lobster (Jasus paulensis)  
Urugavian Lobster (Metanephrops rubellus)  
Western Red Lobster (Panulirus cygnus)

**One of the top-level product headings** in the Agricultural Products package is "Animal Products". By selecting it, one can drill down through a menu to reach the desired product heading, "**Lobsters**". This product heading groups together all the principle types of lobster and prepared lobster foods. For the purposes of this example, we are going to pick one of these types, the "**Panulirus**" species of lobster.

Before looking at the detailed product attributes for our selected lobster, let us first look at some of the many different names by which this particular species is identified.

The "Panulirus" species of lobster goes by some common names, such as the "Spiny Lobster" and the "Sea Crawfish". The species is also differentiated into sub-types, such as "Hawaiian" and "Japanese". These sub-types will also show up in the product attributes which follow this example.

To the right, you will see a list of the many names associated with this species. All of these variant names are contained in the WAND® database with hyperlinks back to the preferred product name.

Finally, here are the detailed product attributes for "Lobsters (Panulirus spp.)".

- Blue Spiny Lobster (Panulirus inflatus)
- California Spiny Lobster (Panulirus interruptus)
- Caribbean Spiny Lobster (Panulirus argus)
- Common Spiny Lobster (Panulirus elephas)
- Gilchrist Spiny Lobster (Panulirus gilchristi)
- Green Spiny Lobster (Panulirus gracilis)
- Hawaiian Spiny Lobster (Panulirus marginatus)
- Japanese Spiny Lobster (Panulirus japonicus)
- Mud Spiny Lobster (Panulirus polyphagus)
- Natal Spiny Lobster (Panulirus delagoae)
- Ornate Spiny Lobster (Panulirus ornatus)
- Painted Spiny Lobster (Panulirus versicolor)
- Panulirus spp.
- Pink Spiny Lobster (Panulirus mauritanicus)
- Pronghorn Spiny Lobster (Panulirus penicillatus)
- Royal Spiny Lobster (Panulirus regius)
- Scalloped Spiny Lobster (Panulirus homarus)
- Sea Crawfish (Panulirus spp.)
- Smoothtail Spiny Lobster (Panulirus laevicauda)
- Spiny Lobsters (Panulirus spp.)
- Western Red Lobster (Panulirus cygnus)

**01: Form**

- 01: Live
- 02: Fresh
- 03: Chilled
- 04: Frozen
- 05: Other

**02: Type**

- 01: Blue Spiny
- 02: California Spiny
- 03: Caribbean Spiny
- 04: Common Spiny
- 05: Gilchrist Spiny
- 06: Green Spiny
- 07: Hawaiian Spiny
- 08: Japanese Spiny
- 09: Mud Spiny
- 10: Natal Spiny
- 11: Ornate Spiny
- 12: Painted Spiny
- 13: Pink Spiny
- 14: Pronghorn Spiny
- 15: Royal Spiny
- 16: Scalloped Spiny
- 17: Smoothtail Spiny
- 18: Western Red
- 19: Other

**03: Preservation Method**

- 01: In Brine
- 02: Salted
- 03: Thermally Processed

**04: Other**

- 04: Freezing Method
- 01: Individually Quick Frozen
- 02: Other

**05: Grade**

- 01: Small Select
- 02: Premium
- 03: Medium Select
- 04: Large Select
- 05: Small Jumbo
- 06: Large Jumbo
- 07: Small Cull
- 08: Large Cull
- 09: Other

**06: Style**

- 01: Chix
- 02: Claw
- 03: Half
- 04: Meat
- 05: Minced
- 06: Pieces
- 07: Quarter
- 08: Tail
- 09: Whole
- 10: Other

**07: Preparation**

- 01: Uncooked
- 02: Partially Cooked
- 03: Fully Cooked

**08: Farm Raised/Wild**

- 01: Farm Raised
- 02: Wild

**09: Head On/Head Off**

- 01: Head On
- 02: Head Off

**10: Shell On/Shell Off**

- 01: Shell On
- 02: Shell Off

**11: Controlled Atmosphere**

- Packaging
- 01: Air (None)
- 02: Vacuum
- 03: Other

**12: Package Form**

- 01: Bag
- 02: Box
- 03: Can
- 04: Other

