TAIGER

iSEARCH

TECHNICAL SPECIFICATIONS

1. iSEARCH TECHNICAL SPECIFICATIONS

1.1. Features

FEATURE	ONTOLOGY BASED	TAXONOMY & THESAURUS-BASED	KEYWORD BASED
Spelling checker and “Did you mean” option	X	X	X
Predictive search	X	X	X
Instant search	X	X	X
Highlighting and snippets	X	X	X
Faceted search	X	X	X
Last searches produced	X	X	X
Most frequent and recently used terms queried	X	X	X
Rich media	X	X	X
Advanced search - Tags	X	X	X
Advanced search – Concepts & Instances	X	X	
Synonyms	X	X	X
Multilingualism	X	X	X
Natural Language	X	X	X
Category	X	X	X
Cross-Language	X	X	
Probabilistic tag cloud	X	X	
Taxonomy support	X	X	
Thesaurus support	X	X	
Ontology support	X		
Contextual	X	X	
Conceptual	X	X	
Browser	X		
Relationship viewer	X		
Triplets	X		

- **Spelling checker and "did you mean" option:** Suggests orthographically correct sentences when users mistype (i.e., "Don Quijotte" will trigger "Did you mean … Don Quijote?").
- **Predictive search:** Suggests relevant terms to complete a user query or sentence. For instance, as the user types "Don" the system generates a drop-down menu with suggestions: "Don Quijote", "Don Antonio", etc.
- **Instant search:** Builds on the category search by directly displaying the categorized results in a drop-down menu so the user can choose among them.
- **Highlighting and snippets**: Displays results using highlights and configurable snippets.
- **Faceted Search:** Makes it possible to search through structured documents, e.g., XML documents or databases, applying the elements and properties of documents to filter through combo boxes, sliders, drop-down lists, etc.
- **Last searches run:** Summary of the last searches run for tunning and configuration purposes.
- **Most frequent and recently used terms queried:**
- **Rich media:** Ability to search for and within multimedia content.
- **Advanced Search:** – Tags.
- **Advanced search – Concepts & Instances:** Defines complex Boolean relationships, directly applying the concepts and instances in the target repository conceptual models.
- **Synonyms:** Makes it possible to expand the search using common synonyms for the seed terms (i.e., "Don Quijote", "Alonso Quijano", "El ingenioso hidalgo Don Quijote de la Mancha").
- **Multilingualism:** Allows locating and retriving documents even if annotation and search
- languages differ. For example, "Don Quijote" will yield the same results as "Don Quixote"
- or "Don Quichotte."
- **Natural language:** Fully-fledged natural language (human) queries.
- **Category**: Complements previous search functionality by proposing categorized results. For example, as the user types "Don Quijote" iSearch will display a drop-down menu detailing DVDs, books or CDs from which the user can choose.
- **Cross-Language:** Allows users to access and retrieve assets regardless of the annotation language and the one used in the query; i.e., "Don Quijote" will trigger the same results as "Don Quixote" or "Don Quichotte".
- **Probabilistic tag cloud:** Expands the traditional concept of the tag cloud, enabling results filtering by including or excluding particular tags from the set of results.
- **Taxonomy support:** Support for taxonomies.
- **Thesaurus support:** Support for thesauri.
- **Ontology support:** Fully fledged ontological support.
- **Contextual:** Makes it possible to locate documents semantically related to the terms of the seed query, not only using the exact words but, more importantly, their meaning.
- **Conceptual:** Advanced search based on the concepts of the underlying information modelling
- formalism (i.e. taxonomies, thesauri, ontologies).
- **Browser:** Enables users to visually navigate the underlying ontology, while visually adding concepts to build queries or refine results.
- **Relationship viewer:** An intuitive, powerful, and visual way to filter the result set by navigating the closest concepts to the one included as part of the seed query.
- **Triplets:** Permits creation of highly complex queries that mimic natural language by editing variables in predefined sentences that represent some of the most common queries.

ADD-ONS

Content connectors	X	X	X
Cloud connectors	X	X	X
Social media connectors	X	X	X
Sentiment analysis	X	X	X
Summary generation	X	X	X
Speech-to-text	X	X	X
Keyword spotting	X	X	X
Interlinking & content enrichment	X		
Pre-loaded industry specific ontologies	X		

1.2. Add-ons

1.2.1. Content Connectors

The Taiger content connector layer sets up an easy-to-use yet powerful data access and federation mechanism. Able to seamlessly interface with relational databases, content management systems, collaboration platforms or messaging systems, it provides the necessary logic to navigate and ingest content into the Taiger enterprise search platform.

The Taiger content connector layer ensures seamless access to more than 40 different repositories. Should your installation require access to content available through other sources, we are able to develop it upon request.

1.2.2. Cloud Connectors

The Taiger cloud connector layer provides a single point of access to data stored in various popular cloud-based services. Relying on Taiger r's powerful semantic-enabled search and visual interface, users are able to ingest, index, and retrieve corporate or private information more precisely, saving time and avoiding the need to log onto and visit a multitude of services. Furthermore, the reliable workflow and fine granularity privacy mechanisms ensure that only the data owner and those assigned access are able to retrieve sensitive content, in accordance with the ongoing workflow policy.

Should your installation require access to content available through other sources, we are able to develop it upon request.

Features:

- Search across cloud-based services through a single access point.
- Support for the most widely used document and media types.
- Advanced semantic annotation and search facilities.
- Powerful visual interface to refine search results and find information faster.
- Web-based interface.
- Support for a wide array of languages and cross-language search.

1.2.3. Social Media Connectors

Taiger social media connectors set up an agile mechanism to search across social media platforms relying on a common interface. Through Taiger faceted search capabilities, users are able to filter tweets by tags, sentiment or topic. Furthermore, the browsing tools help extract information based on the user who published it or the brand it relates to. Just as with other Taiger connectors, strong privacy and workflow mechanisms are available to ensure that only users with the right access level are able to recover sensitive content.

Should your installation require access to content available through other sources, we are able to develop it upon request.

Features:

- Search across social media platforms through a single access point.
- Tweet indexing and filtering by tag.
- Search through published tweets, mentions, and private messages.
- Access to content published on your wall.
- Powerful filtering capabilities.

1.2.4. Sentiment Analysis

Sentiment determination is another step in the process of converting unstructured content to structured content, with the aim of spotting trends and patterns within the information.

The Taiger sentiment analysis add-on sets up the means to consistently and reliably measure emotion and, more importantly, a way to adequately ascertain just who that sentiment targets in the article. We do this by rating the positive or negative assertions that are associated with a document or entity.

Our technology combines the transparency of a dictionary approach—knowing exactly what words are applied—with sophisticated natural language techniques, to ensure that the right words are allocated to the right entities and themes. Currently we have over 250,000 scored sentiment-bearing phrases, part of several completely customizable dictionaries. This enables us to provide very fine-grained control, and our customers can have multiple sentiment dictionaries, each with subtle tweaks to cover different verticals. Furthermore, Taiger is able to identify “objective” vs. “subjective” sentences and use them to change sentiment weighting. For example, “I saw (movie name) last night, and it was great” should have more weight than “I heard (movie name) was great.”

1.2.5. Summary generation

Summarization facilitates shortening a document in order to present its meaning (in the best possible way) using a limited amount of words.

Taiger semantic analysis software accomplishes this at the sentence level; this means we are able to pick out the most important and representative sentences from the content and use them to form a meaningful summary.

1.2.6. Speech-to-Text

Speech-to-text technology is used to extract information from the audio track of videos and audios by processing and transcribing it to text. The result is a highly accurate time-stamped XML, which feeds directly into the annotation process for linguistic and semantic processing.

Taiger Speech-to-text technology stands out due to its accuracy and unparalleled performance. It supports the most widespread European and Asian languages, including various varieties of English and Spanish, and runs as a batch process able to support pre-recorded and live feeds. The app requires less than real time to complete the transcription process, reaching a maximum transcription speed of 1/8 of real time, and features accuracy of up to 90%, which is unprecedented on the market. It can process all major file types and the overall accuracy can be easily improved by feeding the system with a text document containing new words or expressions.

Features:

- Automated time-stamped video and audio transcription.
- Live input feed processing and speaker recognition.
- Processes 8 hours of audio in 1 hour.
- Accuracy improvement by feeding the system a text document with new words or expressions.

1.2.7. Keyword Spotting

Able to process audio signals reaching speeds of up to 15x real time, this adaptable state-of-the-art keyword spotting software appliance makes it possible to locate single words and groups of words. The system can process pre-recorded signals, as well as live input feeds such as news broadcasts or interviews, depositions, or symposia.

Taiger keyword spotting can be configured to work with the most widely-spoken European and international languages, thereby serving a wide range of needs and configurations. A convenient and easy-to-use GUI administration tool—fully integrated in Taiger products—allows preferred configuration set up and easy definition of the target word list.

The result is a time-stamped XML indicating the incidence of each word together with a confidence index to improve reliability and ensure quality control, used in a wide variety of scenarios ranging from search to alerts and notifications. In combination with the Speech-to-Text solution, the Taiger keyword spotting solution permits improvement of the transcription outcome by stressing the presence of controlled vocabularies.

Features:

- Automated keyword location.
- 15x real time processing.
- Configurable keyword source.
- Language-independent.
- Telephonic, microphonic, pre-recorded or live input feed.
- Software appliance.
- Time-stamped XML.
- Confidence index.
- Alerts and notifications.

1.2.8. Interlinking and Content Enrichment

While Taiger already provides mechanisms to search content available within the boundaries of the organization, current needs also demand access to data sources outside the corporate firewall. Thus, Taiger provides advanced technology able to enrich corporate content by interlinking it with other available external resources, such as an external website or the Wikipedia.

The interlinking and content enrichment mechanism is based on a two-step process:

- **Interlinking**: first, external content is aligned with the target information structure in terms of its entities and schemas. For example, assuming that there is a dataset with a list of different anti-aging creams marketed by a cosmetics company, this process will match their ingredients with those ingredients available in other external datasets. However, the mechanism is much more complex and requires the analysis of the graph similarities between the information found in external datasets and the information already existing within the organization.
- **Content enrichment**: second, in addition to the previous step, there is a content enrichment process responsible for providing mechanisms to navigate through the new dataset and extract information from it, showing information and relationships that were unknown to the organization. I.e., it might associate an anti-aging cream marketed by a cosmetics company provided as a query result with information about the ingredients contained; e.g., where an ingredient was used in the past or where it is used for other purposes.

1.2.9. Pre-loaded Sector Specific Ontologies

Taiger products have been designed to support ontologies, taxonomies, and thesauri. Yet they also contain all the features common to traditional keyword-based search engines, such as spelling suggestion and correction, relevancy scores, auto-suggestion, etc., and are able to work without the semantic enhancements provided by ontologies.

While thesauri and taxonomies provide little expressivity, ontologies set up a much richer mechanism to model the relevant concepts in a business application or domain. In a nutshell, they make it possible to characterize any type of relationship among formalized concept categories.

We know that organizations across verticals invest considerable resources in developing or buying formalized categories of concepts. Therefore, Taiger products have been designed in such a way that they are able to swiftly integrate and work with pre-existing models. The Taiger annotation engine will take care of re-indexing content based on the existing model as needed, while the search engine will employ it to retrieve results.

Sometimes existing models need to be updated or fine-tuned to meet a certain purpose. Taiger ontology and taxonomy modeling methodology has been honed over more than 12 years, during the deployment of our products in a wide range of business and industries. Our methodology aims at fulfilling targets, covering all the relevant stages of the ontology life cycle.

Ontology creation:

- Ontology population
- Ontology validation
- Ontology deployment
- Ontology evolution and maintenance

Additionally, Taiger offers a set of pre-loaded sector and application specific ontologies that can be readily used. This allows a significant reduction of the set-up and configuration time, while guaranteeing best practices.

1.3. Technology

1.3.1. Linguistic Resources

The fully-fledged linguistic resources available in all Taiger products provide complete language processing capabilities in a wide array of languages and linguistic varieties. Among other features, Taiger supports **tokenization, removal of stop words, language identification,** and **named entity recognition**. All these help improve search result precision and relevance, while keeping recall limited to avoid information overload.

- **Tokenization**: this process divides a stream of text—e.g., a paragraph, an e-mail, etc.—into smaller pieces called "tokens". These tokens can be single words, symbols, punctuation marks, etc. The output of this process is the centerpiece for the parsing process.
- **Stop words**: in most languages there are words that appear in profusion in every text—articles, prepositions, conjunctions, etc.—but convey little meaning. As search engines are based on the relevance of query words and their frequency in indexed documents, these words are removed in the indexation and querying processes to avoid recovering documents that are not relevant to the query terms.
- **Stemming**: the stemming process is responsible for reducing tokens to their root or stem form. Specifically, during the indexing process words such as "reduce", "reducing", "reduces" and "reduced" are identified and normalized to the stem "reduce". Later, in the query expansion phase, the same process is applied to the query terms, identifying any document that matches the stem form.
- **Lemmatization**: this is a broader process through which the system is able to identify the lemma and its part of speech, i.e. whether the lemma represents a noun, and adverb, a verb, etc. For example, "meeting" could be a verb or a noun, depending on the context. Lemmatization therefore helps the contextualization of information in a given document.
- **Phrasing**: while stemming and lemmatization are concerned with detecting single words, phrasing deals with identifying common expressions formed by more than one word, e.g., "free of charge" or "natural gas". This identification allows better disambiguation in the query and matching processes.
- **Synonym expansion**: In order to increase recall—i.e., the number of relevant documents for a given query—the terms in a query are expanded with their corresponding synonyms so that all relevant documents are taken into consideration. Synonyms can be taken from a gazetteer or using a SKOS plug-in.
- **Word decomposition**: Some languages such as German, Japanese, Russian or Dutch form words by concatenating simpler ones. For example "Wolkenkratzer"—"skyscraper" in English—is formed by the words "Wolken", meaning "clouds", and "Kratzer", meaning "scraper". The word decomposition feature provides the means to break complex words down into their basic components, in order to improve the query processing and document matching.
- **Language identification**: In a multilingual setting, the ability to automatically identify the content language is key. Taiger is able to recognize the text and encoding language in order to automatically apply the corresponding language setting, resources, and techniques.
- **Named entity recognition (NER)**: Taiger offers the means to identify well-known entities, such as people or location names, dates or figures, to name but a few. It allows the system to automatically add useful information to the indexed document base and to increase disambiguation strength. The system can be fed generic dictionaries or specific grammars, or be extended with private dictionaries, thus precisely meeting customer needs.
- **Spelling checker:** through powerful spelling checker technology Taiger is able to detect spelling mistakes in the user's query and suggest a corrected version, while returning useful answers based on the most correct version of the query. The system uses dictionaries and previously indexed documents to provide suggestions.

1.3.2. Semantics

Semantic technology and ontologies are at the heart of our products, supporting contextual annotation, searching, and navigation. Ontologies—i.e. knowledge maps—formalize categories, concepts, and relationships for a specific domain. The links connecting concepts have precise meaning, which models the structure and organizational know-how in a common and agreed way, ensuring consistent information presentation, understanding, and sharing.

Ontologies are the evolution of traditional database schemas, in the sense that they provide a structure for application data. Yet the levels of expressiveness and flexibility are much greater than those of databases, making ontology-based applications significantly more powerful.

Compared to thesauri and taxonomies, ontologies are the most advanced information modeling mechanism currently available. Taxonomies only allow representation of controlled vocabulary word listings—grouped according to their similarity in meaning—typically containing synonyms and sometimes antonyms. Thesauri expand the concept of taxonomies by providing a tree structure that unerringly organizes concepts through supertype-subtype relationships: for example, "car" is a subtype of "vehicle". Ontologies offer all the previous features, plus a graph-like structure to classify concepts and model the relationships among them, thus providing a superior, top-notch information modeling mechanism that only a handful of vendors in the market are able to fully exploit and offer with their products.

Semantic technology and ontologies also help contextualize traditional tags, enabling better understanding, description, and matching of an asset to a user's query. This is the case where tags do not provide the context of the information they are intended to convey.

1.3.3. Advanced Text Analytics

Ontologies and semantic technology are at the core of the advanced linguistic capabilities present in Taiger products. They help improve the performance of traditional phrasing, synonym expansion, and entity extraction by narrowing the context around a term or phrase to increase precision. The ontology normalization unequivocally matches entities to elements in an ontology or information graph. This novel functionality represents the first step towards semantically exploited content. As a result, we are able to make more of the information available, going from just identifying linguistic relationships to also pinpointing semantic ones.

All of these groundbreaking features, including the revelation of hidden assets and the exploitation of the semantic relationships among them, enable information access in ways that are unprecedented and beyond the reach of other market players. These advances clearly differentiate Taiger technology from the rest.

- **Ontology normalization**: this process makes it possible to identify not only generic entities—people, locations, companies, and dates—but also any other relevant entity in the target domain. Moreover, the system can associate it unequivocally with an entity in the corresponding information map. Thus, entities such as "San Francisco", "Frisco", or "SF" are normalized and matched against the entity http://myontology.com/…/#ontology/sanFrancisco. Consequently, any mention of any of those entities will be easily recognized as the city of San Francisco. This translates into an improved textual and semantic analysis.
- **Ontology-based phrasing**: By using ontologies, Taiger can improve its text analyses. In particular, applications can be customized to detect expressions that are made up of more than one word and are explicitly relevant to a specific domain. Thus, an application in the energy sector will understand "light crude" as something completely different from, but at the same time more relevant than, an application analyzing the same words in the food domain.

- **Ontology-based synonym expansion**: Just as with phrasing, the use of ontologies allows the definition of synonyms for entities that are part of specific domain knowledge. For example, a customer in the energy sector can define "crude" as a synonym for "petroleum", although outside this sector that definition may not make any sense.
- **Ontology-based entity extraction**: Using ontologies to improve and customize entity extraction provides increased precision in the entity detection and normalization process. This novel feature extends the generic functionality of Named Entity Recognition (NER) through the use of domain ontologies, providing a set of instances belonging to each concept. In this way, the system is able to detect not only generic entities such as people, companies, and locations, but also elements more specific to the client's interest, such as products in a particular sector, movies or securities, to name but a few.
- **Document summary**: When handling large volumes of information, the ability to rapidly browse through content and discard whatever is not relevant can result in a competitive advantage. Taiger summary generation technology produces accurate and reliable summaries in a fast and convenient way that helps save time and money.
- **Sentiment analysis**: A text usually contains a particular attitude regarding a topic or a contextual polarity. The ability to extract and measure this sentiment is called "sentiment analysis". Using a continuous scale ranging from +1 (most positive) to -1 (most negative), this advanced functionality provides the means to measure the sentiment in a particular document in a standardized manner.

1.3.4. Query Expansion

Having linguistic and semantic resources at the core of our technology and solutions enables Taiger to implement and improve a mechanism known as query expansion.

Users often do not formulate search queries using the best terms:

- Not precise enough: the seed query returns too many results; **low precision**.
- Not abstract enough: the seed query does not return any results at all; **low recall**.

To overcome these limitations, the query expansion techniques are applied to reformulate a seed query and improve information retrieval by increasing the number and quality of search results returned. Our products feature traditional and semantic query expansion.

- **Traditional query expansion**: the linguistic resources are used, for instance, to expand the seed query with synonyms, homonyms (words that share the same spelling but have different meanings) and other morphological forms. At the same time, the stop words are removed and spelling mistakes are corrected. As a result, overall recall is improved; e.g., the system also retrieves documents where "automobile" appears, when given a query with the term "car".
- **Semantic query expansion**: the advanced text analytics—including ontology-based normalization, phrasing, synonym expansion, or extraction—are put to work to enrich users' queries with related entities and relationships originating from the target ontology. For example, a query with the term "accommodation" would expand into a query with different types of accommodation available in the system (e.g. "hotel", "camping", etc.)

To compensate for the behavioral variability of queries with respect to the precision vs. recall trade-off, both traditional and semantic query expansions feature heuristics to fine-tune the critical parameters affecting the query expansion performance. They also contribute to deciding whether the expansion is useful for a particular seed query.

1.3.5. Search+

In classical keyword-based search engines, simple pattern-matching techniques are used to match the user's seed query with the assets in the repository. Nonetheless, when dealing with unstructured information, results are not always what users are seeking. More advanced search engines are able to take full advantage of the context of the user's seed query to increase the system's precision and yield significantly better search results.

Taiger search and browsing technology fully supports traditional keyword search and all its accompanying features. We complement it with groundbreaking contextual, multimedia, and cross-language search technology, in addition to faceted search. The result is a superior search and user experience.

- **Contextual**: by using contextual search, Taiger can locate documents semantically related to the terms of the seed query. It uses not only the exact words, but, more importantly, their meaning. For example, a query for "accommodation in UK" will return documents referencing different accommodation types (e.g., hotels, B&B, guesthouses, apartments, etc.) in UK locations (London, Manchester, Liverpool, etc.) Ontologies enable this advanced matching process by making it easier to extract information from the seed query and facilitating the corresponding document retrieval.
- **Conceptual**: through the use of ontologies, the conceptual search is able to understand and translate the seed query into a set of relevant matching concepts and relationships. For instance, the seed query "hotels in London with swimming pools for less than 90€" would be translated into: hotels located in London, hotels with a swimming pool, hotels with rates less than 90€. The corresponding asset set would include all those documents that precisely meet all these relationships.
- **Multimedia**: searching in audio and video files—just as we do currently with text documents—taking advantage of time-stamped metadata. Through contextualization technology, Taiger can unerringly search across and within rich media content, exploiting information context and seamlessly processing structured and unstructured metadata. The result: a superior technology that makes it possible to pinpoint a search result at the exact second a character appears or a word is pronounced.
- **Cross-language**: this feature enables asset access and retrieval regardless of the annotation language or the one used in the query.
- **Faceted search**: this permits a search through structured documents—e.g., XML documents, databases—using the same structures that shape documents. Users can filter the set of results by combo boxes, text fields, calendar objects, sliders for number ranges, drop-down lists, etc.

1.3.6. Visual Query Expansion

The Taiger GUI interface is one of the most differentiating and powerful features of our solutions and our company. The various configurable controls enable the iterant expansion of queries in order to visually and progressively refine the result set in unprecedented ways. The advanced, user-friendly tools belie the technological complexity, offering an exceptional user experience while providing swift, fresh, precise, and engaging target information, contextual exploration, and discovery. Now users can rapidly gain access to the information they need, saving time and other valuable resources while simultaneously increasing productivity.

We use six different tools, namely: probabilistic tag clouds, relationship viewers, browsing, advanced search, tags and concepts, and faceted search. All the controls work in a cooperative way, with the single goal of refining the result set in the desired way to facilitate information location.

- **Probabilistic tag cloud**: This control expands the traditional concept of the tag cloud in two ways. First, tags are calculated using only the results of each individual user query. This provides significantly more accurate tags than calculating them based on the contents of the whole repository. Second, query-specific tags can be used to filter the result set, by including or excluding particular tags from the set of results. For example, if ”Kendrick” appears in the probabilistic tag cloud, the user can indicate whether all the assets containing this tag in their description or caption should be excluded or maintained in the target result set.
- **Relationship viewer**: The relationship viewer provides a very intuitive, powerful, and visual way to filter the result set. Taking the main concept introduced in the user query as the starting point, the control presents all related concepts that are one degree away according to the current information map. The user can then navigate back and forth, exploring the different relationships among concepts and filtering the result set at will until the desired assets are found.
- **Browser**: This control offers an expanded graphical view of the whole underlying conceptual model. It enables the user to navigate in search of a concept or set of concepts to directly refine the result set by indicating whether concepts should be added or removed from the result set, just like in the case of the tag cloud. It also represents a useful tool when it comes to learning about the underlying ontology or discovering other relevant assets the user was unaware of. The browser supports multiple ontologies.
- **Ontology-based faceted search**: Expanding the concept of traditional faceted search through the use of ontologies, this control enables the creation of facets for unstructured content also, such as audios, videos, and text assets. Users are now able to perform advanced faceted search through structured and unstructured content alike, easily reducing a set of several thousand results to only a few results in just a couple of clicks.
- **Advanced search**: In addition to several common advanced search features, Taiger offers the ability to define complex Boolean relationships, directly applying the concepts and instances in the target repositories and ontologies. As with all the other controls, this one can be combined with the other visual query expansion tools, thereby providing users with precise filtering capabilities.
- **Tags and concepts**: An asset’s properties include tags and concepts that originate directly from the target repositories and ontologies, and which are used to annotate the asset. Taiger enables result set filtering through application of the same techniques as in the case of probabilistic tag clouds. The result set is therefore modified accordingly, enhancing the overall user experience and search precision.

1.3.7. Content Federation

The Taiger content federation layer provides a robust and extensible mechanism for accessing content residing in different repositories and platforms. The platform currently supports the most popular content, cloud, and social media connectors.

- **Content connectors** are able to seamlessly interface with more than 40 relational databases, content management systems, collaboration platforms or messaging systems; they provide the necessary logic to navigate and ingest content into Taiger middleware.
- **Cloud connectors** provide a single point of access to data residing in various popular cloud-based services. They enable information ingestion and indexing, as well as corporate and individual information retrieval, saving time and avoiding the need to visit and log onto a multitude of services.
- **Social media connectors** set up an agile mechanism to gain access to information and search across social media platforms, relying on a common interface and single point of access.

In all cases, strong privacy and workflow mechanisms are available to ensure that only users with the right access level are able to retrieve sensitive content, according to the ongoing workflow policy. When it comes

to administration, the Web-based administration interface is able to support large IT admin teams scattered in different locations.

1.4. Technical Data

1.4.1. Supported Repositories

- **RDBMS**
  - IBM DB2
  - JDBC
  - Microsoft SQL Server
  - MySQL
  - ODBC
  - Oracle RDBMS
  - Sybase

- **Content Management Systems**
  - Alfresco
  - EMC Document Content Server
  - IBM FileNET P8
  - IBM FileNET CS and IS
  - IBM Content Manager
  - IBM Content Manager On Demand
  - IBM Websphere Portal PDM
  - Interwoven Worksite NT
  - Open Text LiveLink
  - OpenText edocs (Hummingbird DM)
  - Oracle Stellent UCM
  - SAP KM
  - Xerox Docushare
  - IBM Enterprise Records

- **RM & Archiving**
  - IBM Enterprise Records
  - HP Trim Context (TowerSoft)
  - Symantec Enterprise Vault

- **Messaging Systems**
  - Lotus Notes
  - Symantec Enterprise Vault

- **Cloud**
  - Iron Mountain
  - Box
  - Huddle
  - Dropbox
  - Google Drive
  - Google Cloud Storage

    - Wuala
    - FilesAnywhere
    - Syncplicity
    - Microsoft Sky Drive

- **Web Content Management**
    - IBM Web Content Management
    - Interwoven Teamsite
    - Vignette

- **Product Lifecycle Management**
    - Enovia MatrixOne

- **Collaboration**
    - EMC CenterStage
    - EMC Documentum eRoom
    - Lotus Quickplace
    - Lotus QuickR
    - IBM Connections
    - MS SharePoint 2003, 2007, 2010

- **Generic Systems**
    - CMIS
    - JCR
    - File Servers
    - FTP
    - WebDAV Server

- **Other**
    - Twitter
    - GoogleDocs
    - SalesForce

1.4.2. Supported Cloud Facilities

- Dropbox
- Box.net
- Evernote
- Google Contacts
- Google Docs
- Google Calendar
- Salesforce
- SkyDrive
- Campfire
- MindMeister
- RSS

1.4.3. Supported Social Media Connectors

- Twitter
- Facebook

- YouTube
- Vimeo

1.4.4. Pre-Loaded Ontologies

- Energy
- Knowledge management
- Media and Entertainment
- Digital copyright management
- Football
- Financial services
- Banking

1.4.5. Supported Languages

	SPEECH-TO-TEXT	KEYWORD SPOTTING
English	X	X
Spanish *(Spain, Argentina, Mexico, Colombia, and Caribbean)*	X	X
French	X	X
Arabic	X	
Hebrew	X	
Swedish	X	
Mandarin Chinese	X	
Japanese	X	
Farsi	X	
Russian	X	
Danish	X	
Korean	X	
German	X	
Dutch	X	
Portuguese *(Portugal & Brazil)*	X	
Turkish	X	
Greek	X	
Italian	X	
Catalonian	X	X
Basque		X
Valencian		X
Galician		X

1.4.6. Supported Media Types

	SPEECH-TO-TEXT	KEYWORD SPOTTING
AVI	X	
FLV	X	
MP3	X	X
WMA	X	X
WMV	X	X
WAV	X	X
AU	X	
MPEG	X	
MPEG2	X	
VOX	X	
OGG	X	