DR.BABASAHEB AMBEDKAR MARATHWADA
UNIVERSITY,AURANGABAD

Sagar B.C.A. College, Jalna

A Project Report On

Information Retrieval on the Internet

Submitted To

UNDER THE GUIDANCE OF

AJITKUMAR

Submitted by

NARWADE AMOL BHASKAR

M.LiB SY

Year 2022-23

Department of Master Of Library and Information Science

(M.Lib. SY)

CERTIFICATE

This is to certify that, the following student

Narwade Amol Bhaskar

Has successfully completed the summer internship project

Information Retrieval on the Internet

In the partial fulfillment of the requirement of Master Of Library and Information Science course asexpected by **Dr. Babasaheb Ambedkar Marathwada University, Aurangabad** for Academic Year **2022-23**.

External Examiner

Internal Examiner

Head
Department of Humanities
Sagar BCA College, Jalna.

Principal

N. D. NAJARDHANE

PRINCIPAL
Sagar BCA College
Devmurti, Tq. Dist. Jalna

STUDENT DECLARATION

This is to declare that this Summer Training Project report on **"Information Retrieval on the Internet"** is a record of genuine work done by me under the guidance of **AJITKUMAR** in the partial fulfillment to the requirement for Master Of Library and Information Science I declare that this Summer Training project report is original and not submitted to anyother university before.

Signature of the Student:

Student's Name: Narwade Amol Bhaskar

PRN No. 2021015200582465

Exam Seat No. CMLD401235

ACKNOWLEDGEMENT

While conducting this report, I got support in many ways from many people .First I am deeply grateful to my project guide, **AJITKUMAR**. who helped me with full devotion and always supported me earnestly whenever it was needed. Without his guidance, mental &moral support and academic inputs this report was not possible.

This Training report could never have seen the light of the day without his co- operation of those Clients who participated in this. I am thankful to all of them for giving me their valuable time.

My friends have been biggest support for me at every juncture of life. They manifested their great interest in my research work also and always tried to make thingseasy for me.

A word of gratitude goes to my family members whose love; affection and understanding have enabled me to complete this end with ease.

At the end, I thank to Almighty for giving me courage and strength to conduct this project report.

(NARWADE AMOL)

Information Retrieval on the Internet

Outline

1. Introduction, 1-1 Search Engines, 1-2 Search Engine History, 1-3 Search Engine Features and Services, 1-4 Search Engine Architectures
2. Web Crawling, 2-1 Robots Exclusion Protocol, 2-2 Robots Meta Tag, 2-3 Multi-Threaded Spidering , 2-4 Focused Spidering, 2-5 Keeping Spidered Pages Up-to-Date
3. The Information Retrieval System, 3-1 Preprocessing the Document Collection, 3-2 Information Retrieval Models, 3-2-1 The Boolean Model, 3-2-2 The Vector Space Model, 3-2-3 Latent Semantic Indexing, 3-2-4 The Probabilistic Model, 3-4 Relevance Feedback
4. Evaluation of Information Retrieval Systems, 4-1 Precision and Recall, 4-2 F-measure and E-measure, 4-3 Mean Average Precision, 4-4 Novelty Ratio and Coverage Ratio
5. Practical Considerations, 5-1 Efficient Indexing, 5-2 Examples of Information Retrieval Systems, 5-2-1 SMART, 5-2-2 Lucene, 5-2-3 Lemur, 5-2-4 OKAPI, 5-3 Example of Large-Scale Search Engine Architecture: Google, 5-4 Page Ranking Algorithms, 5-4-1 Hubs and Authorities, 5-4-2 Google's PageRank, 5-5 Commercial Aspects of Search Engines: Sponsored Links
6. The Invisible Web, 6-1 Searchable Databases, 6-2 Excluded Pages
7. Other Types of Information Retrieval Systems, 7-1 Multimedia Information Retrieval, 7-2 Digital Libraries, 7-3 Distributed Information Retrieval Systems
8. Conclusion and Future Directions, 8-1 Natural Language Queries, 8-2 The Semantic Web and Use of Meta-Data, 8-3 Visualization and Categorization of Results
9. Glossary, Bibliography

Abstract

The main components of a search engine are the Web crawler which has the task of collecting webpages and the Information Retrieval system which has the task of retrieving text documents that answer a user query. In this chapter we present approached to Web crawling, Information Retrieval models, and methods used to evaluate the retrieval performance. Practical considerations include information about existing IR systems and a detailed example of a large-scale search engine (Google), including the idea of ranking webpages by their importance (the Hubs an Authorities algorithm, and Google's PageRank algorithm). Then we discuss the Invisible Web, the part of the Web that is not indexed by search engines. We briefly present other types of IR systems: digital libraries, multimedia retrieval systems (music, video, etc.), and distributed IR systems. We conclude with a discussion of the Semantic Web and future trends in visualizing search results and inputting queries in natural language.

INTRODUCTION

There is a huge quantity of text, audio, video, and other documents available on the Internet, on about any subject. Users need to be able to find relevant information to satisfy their particular information needs. There are two ways of searching for information: to use a search engines or to browse directories organized by categories (such as Yahoo Directories). There is still a large part of the Internet that is not accessible (for example private databases and intranets).

Information retrieval (IR) is the task of representing, storing, organizing, and offering access to information items. IR is different from data retrieval, which is about finding precise data in databases with a given structure. In IR systems, the information is not structured, it is contained in free form in text (webpages or other documents) or in multimedia content. The first IR systems implemented in 1970's were designed to work with small collections of text (for example legal documents). Some of these techniques are now used in search engines.

In this chapter we describe information retrieval techniques, focusing on the challenges faced by search engines. One particular challenge is the large scale, given by the huge number of webpages available on the Internet (for example, about 8 billion webpages were indexed by Google in 2005). Another challenge is inherent to any information retrieval system that deals with text: the ambiguity of the natural language (English or other languages) that makes it difficult to have perfect matches between documents and user queries.

The organization of this chapter is as follows. We briefly mention the search engines history, features, and services. We present the generic architecture of a search engine. We discuss its Web crawling component, which has the task to collect webpages to be indexed. Then we focus on the Information Retrieval component which has the task of retrieving documents (mainly text documents) that answer a user query. We present current methods used to evaluate the performance of the Information Retrieval component. Practical considerations include information about existing IR systems and a detailed example of a large-scale search engine (Google); we present methods for ranking webpages by their importance (the Hubs an Authorities algorithm and Google's PageRank algorithm). In another section, we discuss the Invisible Web, the part of the Web that is not indexed by search engines. We briefly present other types of IR systems: digital libraries, multimedia IR systems, and distributed IR systems. We conclude with a discussion of the Semantic Web and other future trends.

Search engines

There are many general-purpose search engines available on the Web. A resource containing up-to-date information on the most used search engines is: http://www.searchenginewatch.com.
Here are some popular search engines (in alphabetic order):

AllTheWeb http://www.alltheweb.com/
AltaVista http://www.altavista.com/
Excite http://www.excite.com/
Google http://www.google.com/
Hotbot http://www.hotbot.com/
Lycos http://www.lycos.com/
MSN Search http://search.msn.com/
Teoma http://teoma.com/
WiseNut http://www.wisenut.com/
Yahoo! http://search.yahoo.com/

Meta-search engines combine several existing search engines in order to provide documents relevant to a user query. Their task is reduced to ranking results from the different search engines and eliminating duplicates. Some examples are: http://www.metacrawler.com/, http://www.mamma.com/, and http://www.dogpile.com/.

Search Engine History

The very first tool used for searching on the Internet was called Archie (the name stands for "archive"). It was created in 1990 by Alan Emtage, a student at McGill University in Montreal. The program downloaded the directory listings of all the files located on public anonymous FTP sites, creating a searchable database of filenames. Gopher was created in 1991 by Mark McCahill at the University of Minnesota. While Archie indexed file names, Gopher indexed plain text documents. Two other programs, Veronica and Jughead, searched the files stored in Gopher index systems.

The first Web search engine used Wandex, a now-defunct index collected by the World Wide Web Wanderer, a web crawler developed by Matthew Gray at MIT in 1993. Another very early search engine, Aliweb, also appeared in 1993, and still runs today. The first "full text" crawler-based search engine was WebCrawler, 1994. Unlike its predecessors, it let users search for any word in any web page; this became the standard for all major search engines ever since. It was also the first one to be widely known to the public. Also in 1994, Lycos (which started at Carnegie Mellon University) came out, and became a major commercial endeavor.

Soon after, many search engines appeared and became popular. These included Excite, Infoseek, Inktomi, Northern Light, and AltaVista. In some ways, they competed with popular directories such as Yahoo!. Later, the directories integrated or added on search engine technology for greater functionality.

Search engines were also known for the Internet investing frenzy that occurred in the late 1990s. Several companies entered the market spectacularly, with record gains during their initial public offerings. Some have taken down their public search engine, and are marketing enterprise-only editions, such as Northern Light.

Around 2001, the Google search engine rose to prominence (Page and Brin, 1998). Its success was based in part on the concept of link popularity and PageRank, that uses the premise that good or desirable pages are pointed to by more pages than others. Google's minimalist user

interface was very popular with users, and has since spawned a number of imitators. Google is currently the most popular search engine. In 2005, it indexed approximately 8 billion pages, more than any other search engine. It also offers a growing range of Web services, such as Google Maps and online automatic translation tools.

In 2002, Yahoo! acquired Inktomi and in 2003, Yahoo! acquired Overture, which owned AlltheWeb and AltaVista. Despite owning its own search engine, Yahoo initially kept using Google to provide its users with search results. In 2004, Yahoo! launched its own search engine based on the combined technologies of its acquisitions and providing a service that gave pre-eminence to the Web search engine over its manually-maintained subject directory.

MSN Search is a search engine owned by Microsoft, which previously relied on others for its search engine listings. In early 2005 it started showing its own results, collected by its own crawler. Many other search engines tend to be *portals* that merely show the results from another company's search engine. For more details and search engine timelines see, for example, http://en.wikipedia.org/wiki/Search_engine.

Search Engine Features and Services

Search engines allow a user to input keywords that describe an information need. The also offer advanced search capabilities. Although they lead to more precise, they are less utilized by users. We briefly discuss some advanced search features. Boolean features (AND, OR, NOT) that allow retrieval of documents that contain all the keywords (AND), any of the keywords (OR), exclude some words (NOT), or combinations of these Boolean operators. The proximity feature searches for phrases or consecutive words (usually simple search can do this if the words are surrounded by double quotes). The search can be done only in particular fields, such as URLs or titles. Limits can be imposed on the type of retrieved pages: date, language, file types, etc.

Some search engines also offer services: news directories, image search, maps (such as Google Maps), language tools (such as automatic translation tools or interfaces in particular languages), newsgroup search, and other specialized searches.

Search Engine Architectures

The components of a search engine are: Web crawling (gathering webpages), indexing (representing and storing the information), retrieval (being able to retrieve documents relevant to user queries), and ranking the results in their order of relevance. Figure 1 presents a simplified view of the components of a search engine. More details about the main module, the IR system, will follow in the next sections.

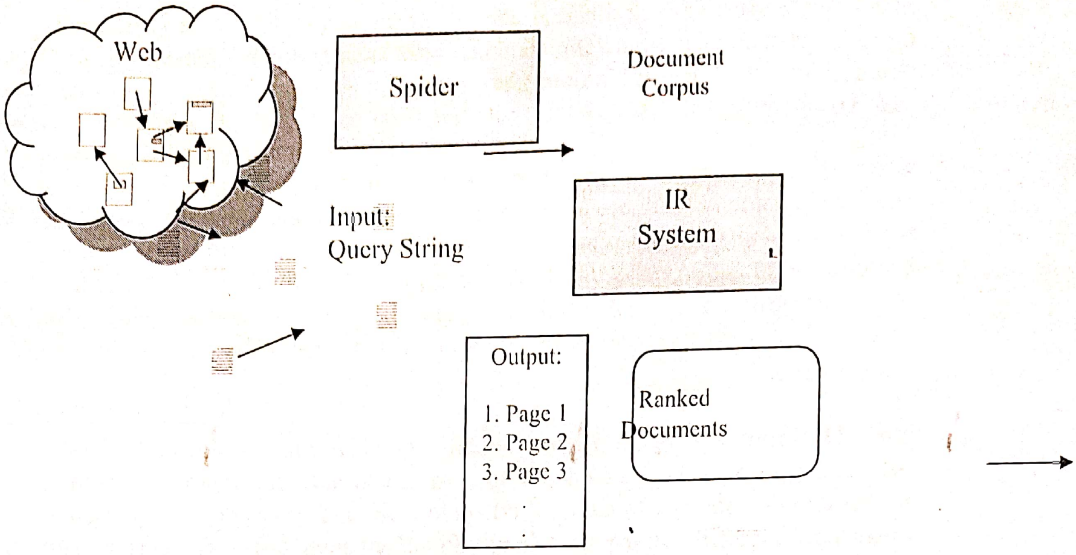


Figure 1. The simplified architecture of a search engine.

WEB CRAWLING

Web crawlers, also known as spiders or robots, have the task to collect webpages to build the text collection for the IR system. The text is extracted from the HTML code of the webpages. Some information related to the HTML format may be stored too. For example, text in headings or in bold font can be given higher weight than the rest of the text.

A crawler starts with one or more http addresses (a set of root URLs), and follows all the links on these pages recursively, to find additional pages. It can proceed by depth-first searching (follow the first link in a page and all the links in the new pages that it leads to, then come back to follow the rest of the links in the current page) or by breadth-first searching (follow all the links in the page for one step, then the links in the pages they point to, for one step, etc.). Breadth-first has the advantage that it explores uniformly outward from the root page but requires memory to store all the links waiting to be traversed on the previous level (exponential in the depth of the links structure). It is the standard spidering method. Depth-first requires less memory (linear in depth) but it might get "lost" pursuing a single thread. Both strategies can be easily implemented using a queue of links (URLs) to be visited next. How new links are added to the queue determines the search strategy. FIFO (first in first out, append to end of the queue) gives breadth-first search. LIFO (last in first out, add to front of queue) gives depth-first search. Heuristically ordering the queue gives a "focused crawler" that directs its search towards "interesting" pages. A spider needs to avoid visiting the same pages again when the links are circular; it needs to keep track of pages that were already visited.

To extract links from a webpage in order to collect candidate links to follow, HTML hyperlink fields are parsed. Here are two examples of hyperlinks:

<a href="http://www.site.uottawa.ca/~diana/csi4107">
<frame src="site-index.html">

If the URL is not specified, like in the last example, the link is relative to the current base URL. If a file name is not specified, a default name is used (such as index.html). The links are put into canonical form: the ending slash is removed, if there is one; internal references within the same page are removed, etc. Once the pages are collected, the text is extracted from the HTML documents, to be processed by the IR system.

Robot exclusion protocols are used to prevent certain sites or webpages from being indexed by Web crawlers. Web sites and pages can specify that robots should not crawl or index certain areas, by using the Robots Exclusion Protocol or the robots meta tag. The second one is newer and less well-adopted than the first one. These standards are conventions to be followed by "good robots". They cannot be enforced, but companies have been prosecuted for "disobeying" these conventions and "trespassing" on private cyberspace.

The Robots Exclusion Protocol

The Robots Exclusion Protocol is a site-wide specification of excluded directories. The site administrator has to put a "robots.txt" file at the root of the host's Web directory. See for example http://www.ebay.com/robots.txt. The file "robots.txt" is a list of excluded directories for a given robot (user-agent). This file contains blank lines to separate different user-agent disallowed directories. with one directory per "Disallow" line. No regular expression can be used as patterns for directories.

To exclude all robots from the entire site, the file would contain:

```
User-agent: *
Disallow: /
```

To exclude specific directories:

```
User-agent: *
Disallow: /tmp/
Disallow: /cgi-bin/
Disallow: /users/paranoid/
```

To exclude a specific robot:

```
User-agent: GoogleBot
Disallow: /
```

To allow a specific robot:

```
User-agent: GoogleBot
Disallow:
```

The Robots Meta Tag

An individual document tag can be used to exclude indexing or following links in a particular webpage. The HEAD section of a specific HTML document can include a robots meta tag, such as `<meta name="robots" content="none">`. The content value can be a pair of values for two aspects: `index` or `noindex` for allowing or disallowing the indexing of this page, and `follow` or `nofollow` for allowing or disallowing following the links in this page. There are two special values: `all = index,follow` and `none = noindex,nofollow`. Examples:

```
<meta name="robots" content="noindex,follow">
<meta name="robots" content="index,nofollow">
<meta name="robots" content="none">
```

Multi-Threaded Spidering

Network delays are frequent when downloading individual pages. It is best to have multiple threads running in parallel, each requesting a page from a different host. The URL's can be distributed to threads, to guarantee equitable distribution of requests across different hosts, in order to maximize through-put and avoid overloading any single server. For example, early Google spiders had multiple coordinated crawlers with about 300 threads each, together being able to download over 100 pages per second.

Focused Spidering

More "interesting" pages could be explored first. There are two styles of focus: topic-directed and link-directed. For the former, if the desired topic description or sample pages of interest are given, the spidering algorithm could sort the queue of links by the similarity (e.g. cosine metric) of their source pages and/or anchor text to this topic description. For the latter, the spider could keep track of in-degree and out-degree of each encountered page, and sort the queue to prefer popular pages with many in-coming links (*authorities*), or to prefer summary pages with many out-going links (*hubs*). See the section on page ranking algorithms for more details.

Keeping Spidered Pages Up-to-Date

The Web is very dynamic: there are many new pages, updated pages, deleted pages, etc. A search engine needs to periodically check spidered pages for updates and deletions. A spider could look in the HTML head information (e.g. meta tags on the last update) to determine if the page has changed, and only reload the entire the page if needed. It could track how often each page is updated and preferentially return to pages which are historically more dynamic. It could preferentially update pages that are accessed more often to optimize the freshness of popular pages.

THE INFORMATION RETRIEVAL SYSTEM

Figure 2 presents a more detailed view of the architecture of an IR system (Baeza-Yates and Berthier Ribeiro-Neto, 1999). Text Operations are used to preprocess the documents collections and to extract index words. The indexing module constructs an inverted index from words to document pointers. The searching module retrieves documents that contain given query words, using the inverted index. The ranking module scores all the retrieved documents according to a relevance metric. The user interface manages interaction with the user: the input of the query and the output of the ranked documents, including the visualization of results. The query operations can transform the query in order to improve retrieval (query expansion using synonyms from a thesaurus, query transformation using relevance feedback).

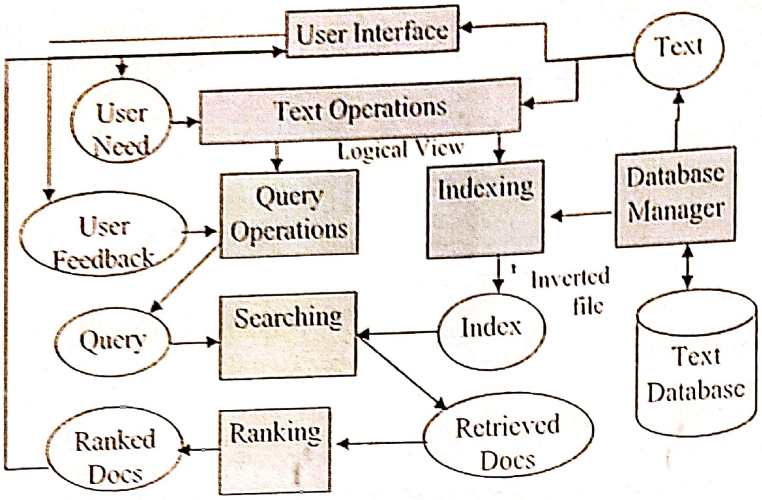


Figure 2. The architecture of an IR system: Text operations are applied of the text of the documents and on the description of the user information need in order to transform them in a simplified form needed for computation. The documents are indexed and the index is used to execute the search. After ranked documents are retrieved, the user can provide feedback which can be used to refine the query and restart the search for improved results.

Preprocessing the document collection

There are several preprocessing steps needed to prepare the document collection for the IR task. The first step is to filter out unwanted characters and markup (e.g. HTML tags, punctuation, numbers, etc.). Then the text needs to be broken into tokens (keywords) by using as delimiters white space and punctuation characters. This it not quite as straightforward as it seems, since words in texts are not always clearly delimited (for example, if the text is *You can't do this*, you can consider the apostrophe as a word separator to get two words. *can* and *t*, or ignore it as separator and consider one word *can't*, or expand the contacted form into two words *can* and *not* and use the white space as separator).

The keywords can be used as they are, or they can be transformed into a base form, for example nouns in the singular form, verbs in the infinitive form, etc. (e.g., *books* becomes *book*, *talked* becomes *talk*). A common approach is to stem the tokens to "stem" forms. For example, *computational* becomes *comput* and *computing* becomes *comput*. Stemming the terms before building the inverted index has the advantage that it reduces the size of the index, and allows for retrieval of webpages with various inflected forms of a word (for example, when searching for webpages with the word *computation*, the results will include webpages with *computations* and *computing*). Stemming is easier to do than computing base forms, because stemmers remove suffixes, without needing a full dictionary of words in a language. A popular and fast stemmer is Porter's stemmer.

Another useful preprocessing step is to remove very frequent words that appear in most of the documents and do not bear any meaningful content. They are called stopwords (e.g., *a*, *the*, *it*, *of*, *could*, etc.). An example of stopwords list can be found at:
http://www.lextek.com/manuals/onix/stopwords1.html.

Important phrases composed of two or more words could also be detected to be used as keywords (possibly using a domain specific dictionary, or using a statistical method for analyzing the text collection in order to detect sequences of words that appear together very often).

Now the text is ready for the next step, building the inverted index that stores for each keyword a list of documents that contain it, in order to allow for fast access during the retrieval step.

Information Retrieval Models

This section presents information retrieval models that can be applied on any text collection. Not all the IR models are easily scaled up to be able to deal with a very large collection, such as pages collected from the Web. The most important IR models are: the Boolean Model, the Vector Space Model, and the Probabilistic Model. Various extensions of these models are possible. We discuss one of them here, Latent Semantic Indexing, which is an extension of the Vector Space Model.

The Boolean Model

The Boolean model is the simplest to implement. A document is represented as a set of keywords. Queries are Boolean expressions of keywords, connected by AND, OR, and NOT, including the use of brackets to indicate the scope of these operators. For example, the query "all the hotels in Rio Brazil or Hilo Hawaii, but not Hilton" is typed by the user as:

[[Rio & Brazil] | [Hilo & Hawaii]] & hotel & !Hilton]

The output of the system is a list of documents that are relevant, but there will be no partial matches or ranking. The Boolean model is very rigid: AND means "all"; OR means "any". All matched documents will be returned, making it difficult to control the number of documents retrieved. All the matched documents satisfy the query to the same degree; that makes it difficult to rank the output. Another disadvantage of this model is that is it not easy for the users to express complex queries.

The Vector Space Model

The vector space model of information retrieval is a very successful statistical method proposed by Salton (1989). It generates weighted term vectors for each document in the collection, and for the user query. Then the retrieval is based on the *similarity* between the query vector and document vectors. The output documents are ranked according to this similarity. The similarity is based on the occurrence *frequencies* of the keywords in the query and in the documents.

Let's assume that t distinct terms remain after preprocessing; let's call them index terms or the vocabulary. These terms form a vector space with dimensionality t, the size of the vocabulary. Each term i, in a document j, is given a weight w_{ij}. Both the documents and the queries are expressed as t-dimensional vectors: $d_j = (w_{1j}, w_{2j}, \ldots, w_{tj})$.

A collection of N documents can be represented in the vector space model by a documents-by-terms matrix. An entry in the matrix corresponds to the "weight" of a term in the document; zero means the term has no significance in the document; it simply doesn't appear in the document. The matrix tends to contain lots of zeros.

	T_1	T_2	....	T_t
d_1	w_{11}	w_{21}	...	w_{t1}
d_2	w_{12}	w_{22}	...	w_{t2}
:	:	:		:
:	:	:		:
d_N	w_{1N}	w_{2N}	...	w_{tN}

The weights in the matrix can be 1 if the term occurs in the document and 0 if it does not (binary weights); but the more frequent terms in a document are more important, i.e., more indicative of the topic. Therefore it is good to use the frequencies of the terms as weights.

Let f_{ij} be the frequency of the term T_i in the document d_j

We can normalize the *term frequency* (*tf*) across the entire corpus: $tf_{ij} = f_{ij} / max\{f_{ij}\}$. Terms that appear in many *different* documents are *less* indicative of overall topic.

Let df_i be the document frequency of term T_i – the number of documents containing the term i, and let idf_i be the inverse document frequency of term T_i:

$$idf_i = log\ (N/df_i)$$

(where N is the total number of documents). The *idf* value is an indication of a term's *discrimination* power. The logarithm is used to dampen the effect relative to *tf*. A typical combined term importance indicator is *tf-idf* weighting:

$$w_{ij} = tf_{ij} \cdot idf_i = tf_{ij}\ log\ (N/df_i).$$

A term occurring frequently in the document but rarely in the rest of the collection is given high weight. Many other ways of determining term weights have been proposed. Experimentally, *tf-idf* has been found to work well.

The query is also transformed into a vector. It is typically treated as a document and also *tf-idf* weighted. Alternatively, the user could supply weights for the given query terms.

The similarity between vectors for the document d_j and the query q can be computed as the vector inner product:

$$sim(d_j, q) = \vec{d_j} \cdot \vec{q} = \sum_{i=1}^{t} w_{ij} \cdot w_{iq}$$

where w_{ij} is the weight of term i in document j and w_{iq} is the weight of term i in the query

For binary vectors, the inner product is the number of matched query terms in the document (the size of the intersection). For weighted term vectors, it is the sum of the products of the weights of the matched terms. There are several problems with the inner product: it does not have a bounded range of values; it favors long documents with a large number of unique terms; it measures how many terms are matched but not how many terms are *not* matched.

The cosine similarity measure tends to work better. The formula is the same as the inner product, but it is normalized by the length of the documents and the length of the query (the length of a vector is the square root of the sum of the squares of its components).

$$cosSim(\vec{d_j}, \vec{q}) = \frac{\vec{d_j} \cdot \vec{q}}{\left|\vec{d_j}\right| \cdot \left|\vec{q}\right|} = \frac{\sum_{i=1}^{t} (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^{t} w_{ij}^2 \cdot \sum_{i=1}^{t} w_{iq}^2}}$$

	T_1	T_2	....	T_t
D_1	w_{11}	w_{21}	...	w_{t1}
D_2	w_{12}	w_{22}	...	w_{t2}
:	:	:		:
:	:	:		:
D_N	w_{1N}	w_{2N}	...	w_{tN}

The weights in the matrix can be 1 if the term occurs in the document and 0 if it does not ary weights); but the more frequent terms in a document are more important, i.e., more cative of the topic. Therefore it is good to use the frequencies of the terms as weights.

f_{ij} be the frequency of the term T_i in the document d_j

can normalize the *term frequency* (*tf*) across the entire corpus: $tf_{ij} = f_{ij} / max\{f_{ij}\}$. Terms appear in many *different* documents are *less* indicative of overall topic.

df_i be the document frequency of term T_i – the number of documents containing the term i.

let idf_i be the inverse document frequency of term T_i:

$$idf_i = log\ (N/df_i)$$

here N is the total number of documents). The *idf* value is an indication of a term's *crimination* power. The logarithm is used to dampen the effect relative to *tf*. A typical nbined term importance indicator is *tf-idf* weighting:

$$w_{ij} = tf_{ij} \cdot idf_i = tf_{ij} \log\ (N/df_i).$$

A term occurring frequently in the document but rarely in the rest of the collection is en high weight. Many other ways of determining term weights have been proposed. perimentally, *tf-idf* has been found to work well.

The query is also transformed into a vector. It is typically treated as a document and also *df* weighted. Alternatively, the user could supply weights for the given query terms.

The similarity between vectors for the document d_j and the query q can be computed as vector inner product:

$$sim(d_j, q) = d_j \bullet q = \sum_{i=1}^{t} w_{ij} \cdot w_{iq}$$

ere w_{ij} is the weight of term i in document j and w_{iq} is the weight of term i in the query

For binary vectors, the inner product is the number of matched query terms in the cument (the size of the intersection). For weighted term vectors, it is the sum of the products the weights of the matched terms. There are several problems with the inner product: it does have a bounded range of values; it favors long documents with a large number of unique ms; it measures how many terms are matched but not how many terms are *not* matched.

The cosine similarity measure tends to work better. The formula is the same as the inner duct, but it is normalized by the length of the documents and the length of the query (the gth of a vector is the square root of the sum of the squares of its components).

$$cosSim(\vec{d_j}, \vec{q}) = \frac{\vec{d_j} \bullet \vec{q}}{\|\vec{d_j}\| \cdot \|\vec{q}\|} = \frac{\sum_{i=1}^{t} (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^{t} w_{ij}^2 \cdot \sum_{i=1}^{t} w_{iq}^2}}$$

The cosine measures the angles between the two vectors (the higher the cosine value – closer to 1, the smaller the angle between the vector of the document and the vector of the query, therefore a more relevant document). Because we consider only the angle, the length of the documents is not a problem anymore.

A naïve implementation of the vector space retrieval is straightforward but impractical: convert all the documents in collection C to *tf-idf* weighted vectors, for all the keywords in the vocabulary V; convert the query to a *tf-idf*-weighted vector q; then for each document d_i in C compute cosSim(d_i, q); sort the documents by decreasing score and present top-ranked documents to the user. The time complexity would be $O(|V|\cdot|C|)$. It would take very long for large V and C (for example, if $|V| = 10{,}000$ and $|C| = 100{,}000$ then $|V|\cdot|C| = 1{,}000{,}000{,}000$).

A practical implementation is based on the observation that documents containing none of the query words do not affect the final ranking. Identifying those documents that contain at least one query word is easily done by using an inverted index. The numerator in the cosine similarity formula will be calculated much faster because the multiplications where one of the terms is zero will not be executed.

The steps of a practical implementation are as follows. Step 1, pre-processing (tokenization, stopword removal, stemming), determines the keywords in the vocabulary to be used as index terms. Step 2 is the building of the inverted index, with an entry for each keyword in the vocabulary (see Figure 3). The index is a data structure that will allow fast access in the retrieval step (hash table, B-tree, sparse list, etc.) For each keyword, the index keeps a list of all the documents where it appears together with the corresponding term frequency (*tf*). It also keeps the total number of occurrences in all documents (for the *idf* score). So the *tf-idf* scores can be computed in one pass trough the collection. The cosine similarity also requires document lengths; a second pass to is needed to compute document vector lengths. The time complexity of indexing a document of n tokens is $O(n)$. So indexing m such documents takes $O(m\ n)$. Computing *idf* scores can be done during the same first pass. Therefore computing the vector lengths is also $O(m\ n)$. Completing the process takes $O(m\ n)$, which is also the complexity of just reading in the collection of documents.

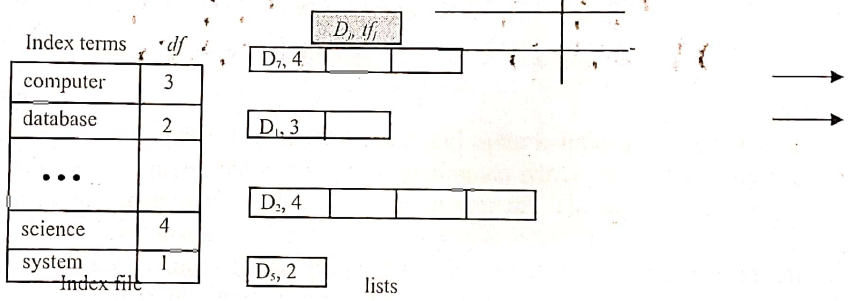


Figure 3 Example of inverted index: for each term, *df* is the number of documents in which it occurred; each list element records the document where the term occurred and how many times.

The last step is the retrieval process. The inverted index from Step 2 is used to find the limited set of documents that contain at least one of the query words. Then the cosine similarity

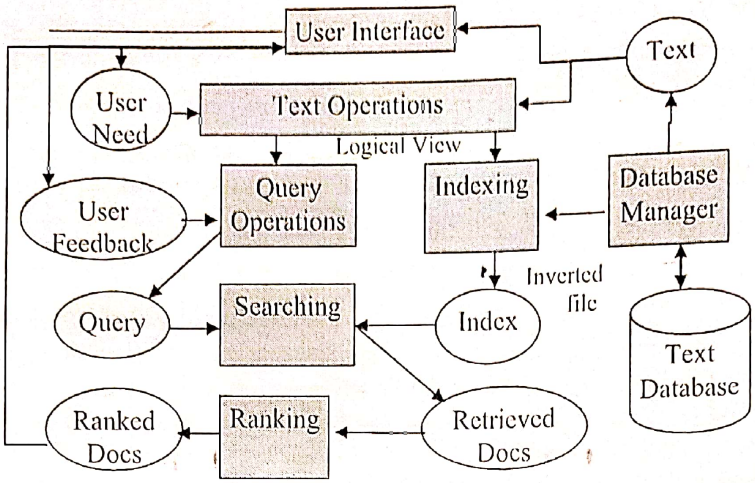


Figure 2. The architecture of an IR system: Text operations are applied of the text of the documents and on the description of the user information need in order to transform them in a simplified form needed for computation. The documents are indexed and the index is used to execute the search. After ranked documents are retrieved, the user can provide feedback which can be used to refine the query and restart the search for improved results.

Preprocessing the document collection

There are several preprocessing steps needed to prepare the document collection for the IR task. The first step is to filter out unwanted characters and markup (e.g. HTML tags, punctuation, numbers, etc.). Then the text needs to be broken into tokens (keywords) by using as delimiters white space and punctuation characters. This it not quite as straightforward as it seems, since words in texts are not always clearly delimited (for example, if the text is *You can't do this*, you can consider the apostrophe as a word separator to get two words *can* and *t*, or ignore it as separator and consider one word *can't*, or expand the contacted form into two words *can* and *not* and use the white space as separator).

The keywords can be used as they are, or they can be transformed into a base form, for example nouns in the singular form, verbs in the infinitive form, etc. (e.g., *books* becomes *book*, *talked* becomes *talk*). A common approach is to stem the tokens to "stem" forms. For example, *computational* becomes *comput* and *computing* becomes *comput*. Stemming the terms before building the inverted index has the advantage that it reduces the size of the index, and allows for retrieval of webpages with various inflected forms of a word (for example, when searching for webpages with the word *computation*, the results will include webpages with *computations* and *computing*). Stemming is easier to do than computing base forms, because stemmers remove suffixes, without needing a full dictionary of words in a language. A popular and fast stemmer is Porter's stemmer.

Another useful preprocessing step is to remove very frequent words that appear in most of the documents and do not bear any meaningful content. They are called stopwords (e.g., *a*, *the*, *it*, *of*, *could*, etc.). An example of stopwords list can be found at: http://www.lextek.com/manuals/onix/stopwords1.html.

Important phrases composed of two or more words could also be detected to be used as keywords (possibly using a domain specific dictionary, or using a statistical method for analyzing the text collection in order to detect sequences of words that appear together very often).

Now the text is ready for the next step, building the inverted index that stores for each keyword a list of documents that contain it, in order to allow for fast access during the retrieval step.

Information Retrieval Models

This section presents information retrieval models that can be applied on any text collection. Not all the IR models are easily scaled up to be able to deal with a very large collection, such as pages collected from the Web. The most important IR models are: the Boolean Model, the Vector Space Model, and the Probabilistic Model. Various extensions of these models are possible. We discuss one of them here, Latent Semantic Indexing, which is an extension of the Vector Space Model.

The Boolean Model

The Boolean model is the simplest to implement. A document is represented as a set of keywords. Queries are Boolean expressions of keywords, connected by AND, OR, and NOT, including the use of brackets to indicate the scope of these operators. For example, the query "all the hotels in Rio Brazil or Hilo Hawaii, but not Hilton" is typed by the user as:

[[Rio & Brazil] | [Hilo & Hawaii]] & hotel & !Hilton]

The output of the system is a list of documents that are relevant, but there will be no partial matches or ranking. The Boolean model is very rigid: AND means "all"; OR means "any". All matched documents will be returned, making it difficult to control the number of documents retrieved. All the matched documents satisfy the query to the same degree; that makes it difficult to rank the output. Another disadvantage of this model is that is it not easy for the users to express complex queries.

The Vector Space Model

The vector space model of information retrieval is a very successful statistical method proposed by Salton (1989). It generates weighted term vectors for each document in the collection, and for the user query. Then the retrieval is based on the *similarity* between the query vector and document vectors. The output documents are ranked according to this similarity. The similarity is based on the occurrence *frequencies* of the keywords in the query and in the documents.

Let's assume that t distinct terms remain after preprocessing; let's call them index terms or the vocabulary. These terms form a vector space with dimensionality t, the size of the vocabulary. Each term i, in a document j, is given a weight w_{ij}. Both the documents and the queries are expressed as t-dimensional vectors: $d_j = (w_{1j}, w_{2j}, \ldots, w_{tj})$.

A collection of N documents can be represented in the vector space model by a documents-by-terms matrix. An entry in the matrix corresponds to the "weight" of a term in the document; zero means the term has no significance in the document; it simply doesn't appear in the document. The matrix tends to contain lots of zeros.

	T_1	T_2	....	T_t
d_1	w_{11}	w_{21}	...	w_{t1}
d_2	w_{12}	w_{22}	...	w_{t2}
:	:	:		:
:	:	:		:
d_N	w_{1N}	w_{2N}	...	w_{tN}

The weights in the matrix can be 1 if the term occurs in the document and 0 if it does not (binary weights); but the more frequent terms in a document are more important, i.e., more indicative of the topic. Therefore it is good to use the frequencies of the terms as weights.
Let f_{ij} be the frequency of the term T_i in the document d_j
We can normalize the *term frequency* (*tf*) across the entire corpus: $tf_{ij} = f_{ij} / max\{f_{ij}\}$. Terms that appear in many *different* documents are *less* indicative of overall topic.
Let df_i be the document frequency of term T_i – the number of documents containing the term i, and let idf_i be the inverse document frequency of term T_i:

$$idf_i = log\ (N/df_i)$$

(where N is the total number of documents). The *idf* value is an indication of a term's *discrimination* power. The logarithm is used to dampen the effect relative to *tf*. A typical combined term importance indicator is *tf-idf* weighting:

$$w_{ij} = tf_{ij} \cdot idf_i = tf_{ij} log\ (N/df_i).$$

A term occurring frequently in the document but rarely in the rest of the collection is given high weight. Many other ways of determining term weights have been proposed. Experimentally, *tf-idf* has been found to work well.

The query is also transformed into a vector. It is typically treated as a document and also *tf-idf* weighted. Alternatively, the user could supply weights for the given query terms.

The similarity between vectors for the document d_j and the query q can be computed as the vector inner product:

$$sim(d_j, q) = d_j \bullet q = \sum_{i=1}^{t} w_{ij} \cdot w_{iq}$$

where w_{ij} is the weight of term i in document j and w_{iq} is the weight of term i in the query

For binary vectors, the inner product is the number of matched query terms in the document (the size of the intersection). For weighted term vectors, it is the sum of the products of the weights of the matched terms. There are several problems with the inner product: it does not have a bounded range of values; it favors long documents with a large number of unique terms; it measures how many terms are matched but not how many terms are *not* matched.

The cosine similarity measure tends to work better. The formula is the same as the inner product, but it is normalized by the length of the documents and the length of the query (the length of a vector is the square root of the sum of the squares of its components).

$$\cos Sim(d_j, q) = \frac{d_j \bullet q}{|d_j| \cdot |q|} = \frac{\sum_{i=1}^{t} (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^{t} w_{ij}^2 \cdot \sum_{i=1}^{t} w_{iq}^2}}$$

The cosine measures the angles between the two vectors (the higher the cosine value – closer to 1, the smaller the angle between the vector of the document and the vector of the query, therefore a more relevant document). Because we consider only the angle, the length of the documents is not a problem anymore.

A naïve implementation of the vector space retrieval is straightforward but impractical: convert all the documents in collection C to *tf-idf* weighted vectors, for all the keywords in the vocabulary V; convert the query to a *tf-idf*-weighted vector q; then for each document d_i in C compute cosSim(d_i, q); sort the documents by decreasing score and present top-ranked documents to the user. The time complexity would be $O(|V|\cdot|C|)$. It would take very long for large V and C (for example, if $|V| = 10{,}000$ and $|C| = 100{,}000$ then $|V|\cdot|C| = 1{,}000{,}000{,}000$).

A practical implementation is based on the observation that documents containing none of the query words do not affect the final ranking. Identifying those documents that contain at least one query word is easily done by using an inverted index. The numerator in the cosine similarity formula will be calculated much faster because the multiplications where one of the terms is zero will not be executed.

The steps of a practical implementation are as follows. Step 1, pre-processing (tokenization, stopword removal, stemming), determines the keywords in the vocabulary to be used as index terms. Step 2 is the building of the inverted index, with an entry for each keyword in the vocabulary (see Figure 3). The index is a data structure that will allow fast access in the retrieval step (hash table, B-tree, sparse list, etc.) For each keyword, the index keeps a list of all the documents where it appears together with the corresponding term frequency (*tf*). It also keeps the total number of occurrences in all documents (for the *idf* score). So the *tf-idf* scores can be computed in one pass trough the collection. The cosine similarity also requires document lengths; a second pass to is needed to compute document vector lengths. The time complexity of indexing a document of n tokens is $O(n)$. So indexing m such documents takes $O(m\ n)$. Computing *idf* scores can be done during the same first pass. Therefore computing the vector lengths is also $O(m\ n)$. Completing the process takes $O(m\ n)$, which is also the complexity of just reading in the collection of documents.

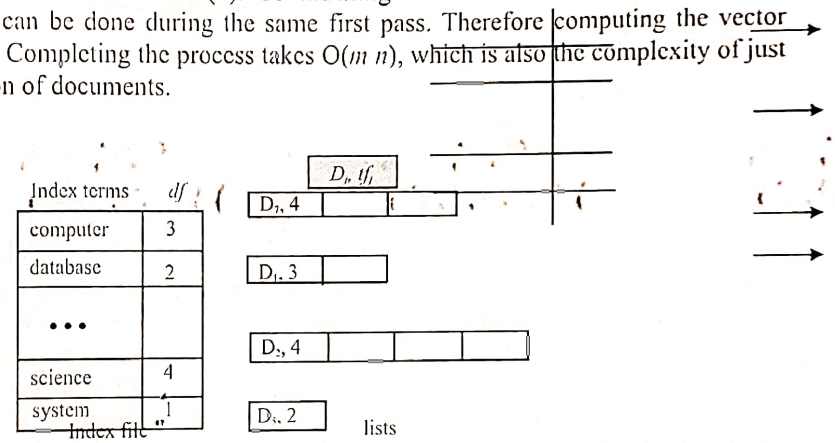


Figure 3 Example of inverted index: for each term, *df* is the number of documents in which it occurred; each list element records the document where the term occurred and how many times.

The last step is the retrieval process. The inverted index from Step 2 is used to find the limited set of documents that contain at least one of the query words. Then the cosine similarity

maximize the probability of relevance. Documents in the set R are predicted to be

he probabilistic ranking is computed as:

$$m(d_j,q) = P(R \mid d_j) / P(\neg R \mid d_j)$$

e ratio of the probability that the document d_j is relevant and the probability that it is not It reflects the odds of the document d_j being relevant, and minimizes the probability of ous judgment. Using Bayes rule (for two events A and B, the probability of A given B = P(B|A) P(A) / P(B)) we expand the formula:

$$sim(d_j,q) = \frac{P(d_j \mid R) \cdot P(R)}{P(d_j \mid \neg R) \cdot P(\neg R)} \cong \frac{P(d_j \mid R)}{P(d_j \mid \neg R)}$$

is the probability of randomly selecting the document d_j from the set R of relevant s. $P(R)$ stands for the probability that a document randomly selected from the collection is relevant. The meanings attached to $P(d_j \mid \neg R)$ and $P(\neg R)$ are analogous lementary. $P(R)$ and $P(\neg R)$ are the same for all the documents relative to the query. e replace the probability of each document by the product of the probabilities of the ntains. We assume the terms occur in a document independent of each other; this is a g assumption that works well in practice, even if in reality terms are not independent, ce of a term might trigger the presence of a closely related term. We obtain:

$$sim(d_j,q) \cong \frac{\prod_{k_i} P(k_i \mid R) \cdot \prod_{k_i} P(k_i \mid \neg R)}{\prod_{k_i} P(\neg k_i \mid R) \cdot \prod_{k_i} P(\neg k_i \mid \neg R)}$$

$_i \mid R)$ is probability that the index term k_i is present in a document randomly selected et R of relevant documents and $P(\neg k_i \mid R)$ is the probability that k_i is not present . The es for $\neg R$ have analogous meanings. Taking logarithms and ignoring factors that are or all the documents in the context of the same query we obtain:

$$sim(d_j,q) \cong \sum_i w_{i,j} \cdot w_{i,q} \cdot \left(\log \frac{P(k_i \mid R)}{P(\neg k_i \mid R)} + \log \frac{P(k_i \mid \neg R)}{P(\neg k_i \mid \neg R)}\right)$$

re binary weights, 1 if the index term is in the document or in the query, 0 if not.
$\neg k_i \mid R) = 1 - P(k_i \mid R)$ and $P(\neg k_i \mid \neg R) = 1 - P(k_i \mid \neg R)$.
bilities left to estimate are: $P(k_i \mid R)$ and $P(k_i \mid \neg R)$. They can have initial guesses:
0.5 and $P(k_i \mid \neg R) = df_i / N$, where df_i is the number of documents that contain k_i.
l guess is used to retrieve an initial set of document V, from which the subset V_i e index term k_i. The estimates are re-evaluated:
$_i \mid R) = V_i / V$ and $P(k_i \mid \neg R) = (df_i - V_i) / (N - V)$
ss can be repeated recursively. By doing so, the guess of the probabilities can be vithout the need of the user intervention (contrary to what we mentioned above).
last formulas pose problems for small values of V and V_i (such as V=1 and V_i= 0).
ent these problems, an adjustment factor is added, for example:
$_i \mid R) = (V_i + 0.5) / (V + 1)$ and $P(k_i \mid \neg R) = (df_i - V_i + 0.5) / (N - V + 1)$
opular variant of the probabilistic model is the Okapi formula (Robertson *et. al*, 2000).

Relevance Feedback

The users tend to ask short queries, even when the information need is complex. Irrelevant documents are retrieved as answers because on the ambiguity of the natural language (words have multiple senses). If we know that some of retrieved documents were relevant to the query, terms from those documents can be added to the query in order to be able to retrieve more relevant documents. This is called *relevance feedback*. Often, it is not possible to ask the user to judge the relevance of the retrieved documents. In this case *pseudo-relevance feedback* methods can be used. They assume the first few retrieved documents are relevant and use the most important terms from them to expand the query.

EVALUATION OF INFORMATION RETRIEVAL SYSTEMS

To compare the performance of information retrieval systems there is a need for standard test collections and benchmarks. The TREC forum (Text Retrieval Conference, http://trec.nist.gov/) provides test collections and organizes competition between IR systems every year, since 1992. In order to compute evaluation scores, we need to know the expected solution. Relevance judgments are produced by human judges and included in the standard test collections. CLEF (Cross-Language Evaluation Forum) is another evaluation forum that organizes competition between IR systems that allow queries or documents in multiple languages (http://www.clef-campaign.org/), since the year 2000.

In order to evaluate the performance of an IR system we need to measure how far down the ranked list of results will a user need to look to find some or all the relevant documents. The typical evaluation process starts with finding a collection of documents. A set of queries needs to be formulated. Then one or more human experts are needed to exhaustively label the relevant documents for each query. This assumes binary relevance judgments: a document is relevant or not to a query. This is simplification, because the relevancy is continuous: a document can be relevant to a certain degree. Even if relevancy is binary, it can be a difficult judgment to make. Relevancy, from a human standpoint, is subjective because it depends on a specific user's judgment; it is situational, it relates to the user's current needs; it depends on human perception and behavior; and it might be dynamic, it can change over time.

Once the test suite is assembled, we can compute numerical evaluation measures, for each query, and then average over all the queries in the test set.

Precision and Recall

Precision (P) measures the ability to retrieve top-ranked documents that are mostly relevant. Recall (R) measures the ability of the search to find all of the relevant items in the corpus.

$$P = \frac{\text{Number of relevant documents retrieved}}{\text{Total number of documents retrieved}}$$

$$R = \frac{\text{Number of relevant documents retrieved}}{\text{Total number of relevant documents}}$$

F-measure and E-measure

The F-measure combines precision and recall, taking their harmonic mean. The F-measure is high when both precision and recall are high.

$$F = \frac{2PR}{P+R} = \frac{2}{\frac{1}{R}+\frac{1}{P}}$$

A generalization of the F-measure is the E-measure, which allows emphasis on precision over recall or vice-versa. The value of the parameter β controls this trade-off: if β = 1 precision and recall are weighted equally (E=F), if β < 1 precision weights more, and if β > 1 recall weights more.

$$E = \frac{(1+\beta^2)PR}{\beta^2 P + R} = \frac{(1+\beta^2)}{\frac{\beta^2}{R}+\frac{1}{P}}$$

Figure 4 shows the distribution of the retrieved documents versus the relevant documents. In the upper part of the figure, the intersection of the two circles is the part that needs to be maximized by an IR system. In the lower part of the figure, the number of documents that need to be maximized is in the lower left corner and the upper right corner. The other two corners would contain zeros for an ideal IR system.

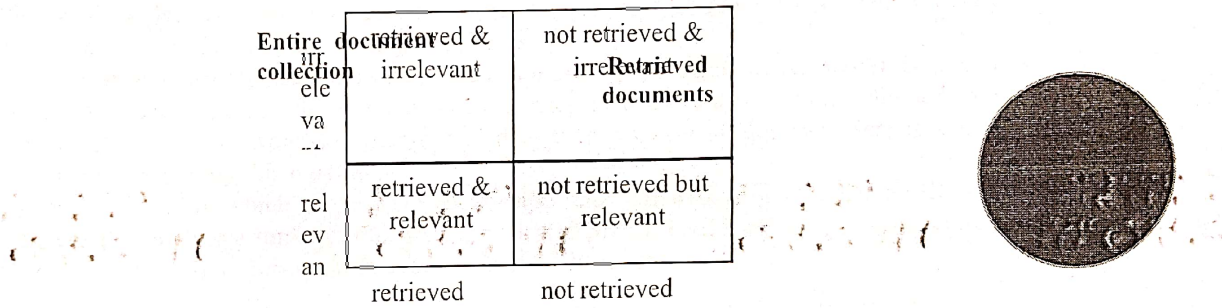


Figure 4. Retrieved versus relevant documents.

Sometimes, for very large text collections or the Web, it is difficult to find the total number of relevant items. In such cases, one solution is to sample across the collection and to perform relevance judgment on the sampled items. Another idea is to apply different retrieval algorithms or different IR systems to the same collection for the same query, then aggregate the relevant items from all of them and perform relevance judgments on this set.

Mean Average Precision

Usually precision is more important than recall in IR systems, if the user is looking for an answer to a query, not for all the possible answers. Recall can be important when a user needs to know all the relevant information on a topic. A system can increase precision by decreasing recall and vice-versa: there is a precision-recall tradeoff (for example, recall can be increased by simply retrieving more documents, but the precision will go down, since many retrieved documents will not be relevant). Precision-recall curves can be used to compare two IR systems for all values of precision and recall.

In fact precision and recall are not enough for evaluating IR systems. For example, if we have two systems that retrieve 10 documents each, 5 relevant and 5 not relevant, both have precision 0.5, but a system that has the first 5 retrieved documents relevant and the next 5 irrelevant is much better than a system that has the first 5 retrieved documents irrelevant and the next 5 relevant (because the user will be annoyed to have to check the irrelevant documents first). Modified measures that combine precision and recall and consider the order of the retrieved documents are needed.

Some good measures are: precision at 5 retrieved documents, precision at 10 retrieved documents or some other cut-off point; the R-Precision; the interpolated average precision; and the mean average precision. The trec_eval script can be used to compute many evaluation measures (http://trec.nist.gov/trec_eval/).

The *R-precision* is the precision at the R-th position in the ranking of the results for a query that has R known relevant documents.

The *interpolated average precision* computes precision at fixed recall intervals (11 points), to allow fair average over all the queries in the test set at the same recall levels. See Chapter 3 of (Baeza-Yates and Berthier Ribeiro-Neto, 1999) for more details. This measure is not much used use lately in evaluating IR systems.

The most widely-used measure is the *mean average precision* (MAP score). It computes precision at each point in the ranking where a relevant document was found, divided by the number of existing relevant documents (and then averages over all queries). Here is a simple example of computing this measure.

Given a query q, for which the relevant documents are d1, d6, d10, d15, d22, d26, an IR system retrieves the following ranking: d6, d2, d11, d3, d10, d1, d14, d15, d7, d23. We compute the precision and recall for this ranking at each retrieved document.

Rank	Document	Recall	Precision
1	d6	1/6 = 0.166	1/1 = 1.00
2	d2	1/6 = 0.166	1/2 = 0.50
3	d11	1/6 = 0.166	1/3 = 0.33
4	d3	1/6 = 0.166	1/4 = 0.25
5	d10	2/6 = 0.33	2/5 = 0.40
6	d1	3/6 = 0.50	3/6 = 0.50
7	d14	3/6 = 0.50	3/7 = 0.42
8	d15	4/6 = 0.66	4/8 = 0.50
9	d7	4/6 = 0.66	4/9 = 0.44
10	d23	4/6 = 0.66	4/10 = 0.40

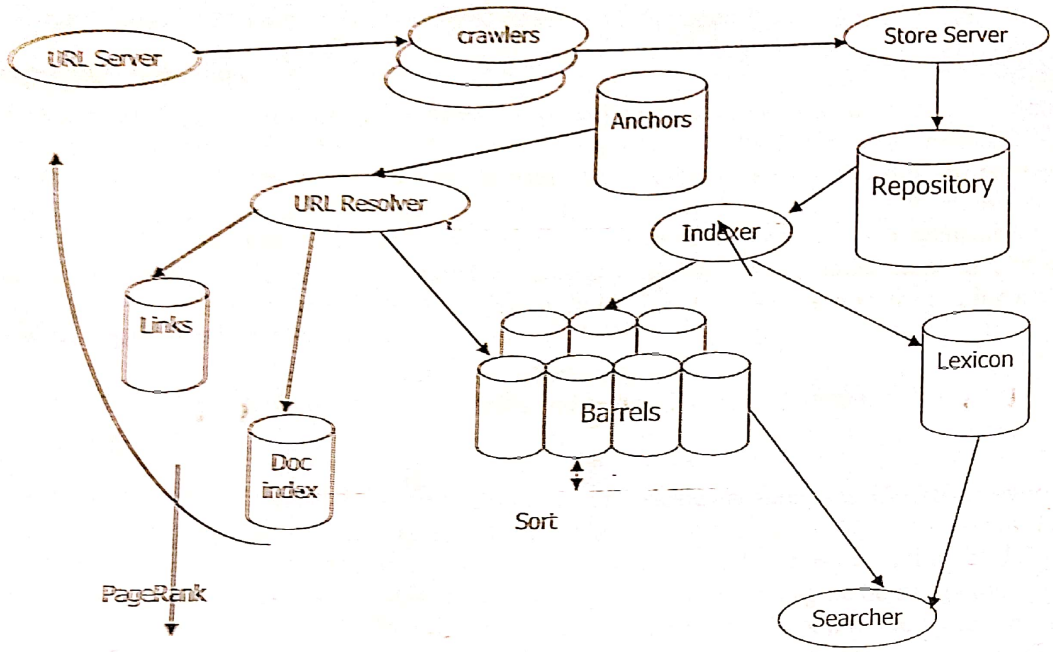


Figure 5. Google's architecture.

In order to deal with many small files in an efficient way (in both space requirements and access time), the system uses big virtual files addressable by 64 bit integers, and it supports compression. The "Repository" contains the full HTML code of every webpage (compressed), a document identifier, its length, and its URL. The "Document Index" keeps information about each document (the document identifier, the current document status, a pointer into the repository, a document checksum, and various statistics). The "Lexicon" is a repository of words, implemented as a list and a hash table of pointers. The list stores occurrences of a particular word in a particular document. It also records the types of hit: Plain (simple text), Fancy (in special HTML format such as bold or heading) and Anchor (text on a link).

There are two indexes: The Forward Index (for fast access to words using word identifiers, and to documents using document identifiers) and the Inverted Index (for the actual retrieval and similarity calculation).

Google associates the text of a link with the page of the link and the page where the link points to. The advantages of doing this are: the anchors often provide accurate descriptions; anchors may exist for documents which cannot be indexed (i.e., images, programs, and databases); propagating anchor text improves the quality of the results.

Page Ranking Algorithms

In addition to how relevant the retrieved webpages are to the user query, they can also be ranked by their importance. A webpage is important, for example, if there are many webpages that have

links to it. This section presents the Hubs and Authorities algorithm (Kleinberg, 1999) and Google's PageRank algorithm (Brin & Page, 1998), with examples.

Hubs and Authorities

Authorities are pages that are recognized as providing significant, trustworthy, and useful information on a topic. The *in-degree* of a page (the number of links that point to the page) is a simple measure of authority. *Hubs* are index pages that provide lots of useful links to relevant content pages (topic authorities). The algorithm developed by Kleinberg in 1998 attempts to computationally determine hubs and authorities on a particular topic through analysis of a relevant subgraph of the Web. It is based on mutually recursive facts: hubs point to lots of authorities and authorities are pointed to by lots of hubs. Together they tend to form a bipartite graph, as depicted in Figure 6.

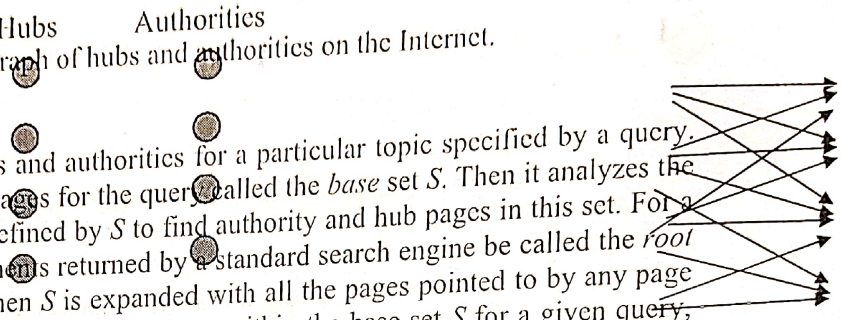


Figure 6. Bipartite graph of hubs and authorities on the Internet.

The algorithm computes hubs and authorities for a particular topic specified by a query. First, it determines a set of relevant pages for the query called the *base* set S. Then it analyzes the link structure of the Web subgraph defined by S to find authority and hub pages in this set. For a specific query Q, let the set of documents returned by a standard search engine be called the *root* set R. The set S is initialized to R. Then S is expanded with all the pages pointed to by any page in R and all the pages that point to any page in R. Even within the base set S for a given query, the nodes with highest in-degree are not necessarily authorities (they may be generally popular pages like Yahoo or Amazon).

The algorithm slowly converges on a mutually reinforcing set of hubs and authorities. For each page $p \in S$, an authority score a_p and a hub score h_p are maintained. All a_p and h_p are initialized with 1. The algorithm is repeated several times. At each iteration, the scores are maintained normalized, and the new scores use the values from the previous iteration.

$$\sum_{p \in S} (a_p)_2 = \sum_{p \in S} (h_p)_2 = 1$$

Authorities are pointed to by lots of good hubs (all pages q that point to p):

$$a_p = \sum_{q:q \to p} h_q$$

Hubs point to lots of good authorities (all pages q that p points at):

$$h_p = \sum_{q:p \to q} a_q$$

For example, if pages 1, 2, and 3 point to page 4, and page 4 points to pages 5, 6, and 7, the scores are computed as exemplified in Figure 7.

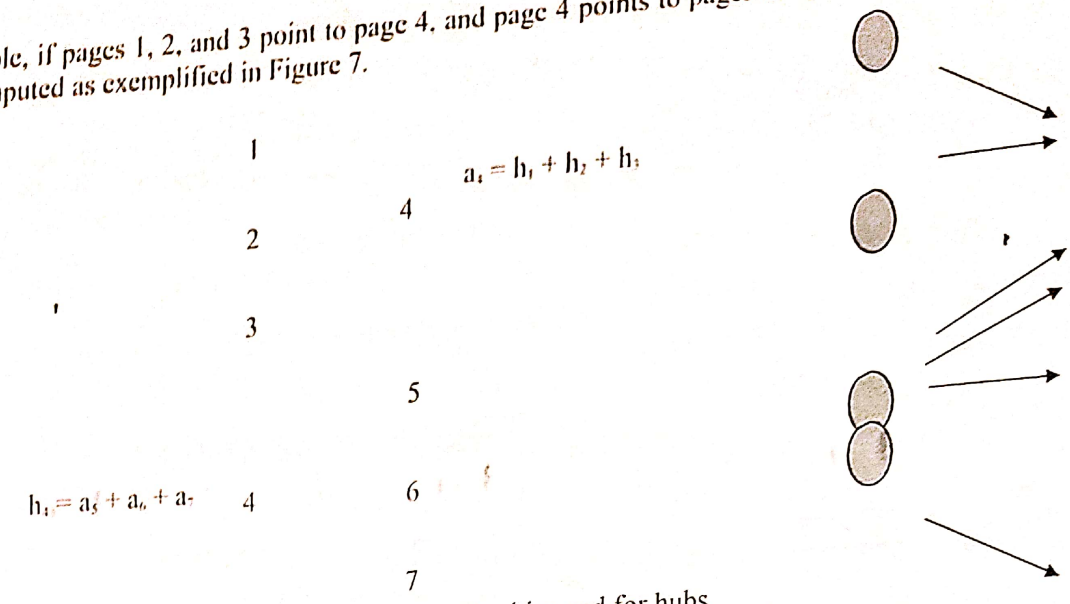


Figure 7. Example of computing scores for authorities and for hubs.

The hubs and authorities algorithm can be summarized as follows:

Initialize for all $p \in S$: $a_p = h_p = 1$

Repeat k times (where k is the number of iterations):

For all $p \in S$: $a_p = \sum_{q:q \to p} h_q$ (update authority scores)

For all $p \in S$: $h_p = \sum_{q:p \to q} a_q$ (update hub scores)

For all $p \in S$: $a_p = a_p / c$ where c is a constant such that: $\sum_{p \in S} (a_p / c)_2 = 1$

For all $p \in S$: $h_p = h_p / c$ where c is a constant such that: $\sum_{p \in S} (h_p / c)_2 = 1$

The algorithm converges to a *fix-point*, where the scores do not change at the next iteration. In practice, 20 iterations produce fairly stable results.

Google's PageRank

An alternative link-analysis method is PageRank, used by Google (the actual formula currently used by Google might be slightly different). PageRank does not attempt to capture the distinction between hubs and authorities, it ranks pages only by authority. It is applied to the entire Web rather than a local neighborhood of pages surrounding the results of a query.

If p is a given page and $q_1, \ldots, q_n$ are the pages that point to the page p, the page rank PR of p is given by the sum of the page ranks of all the pages $q_1, \ldots, q_n$ each of them divided by its number of outgoing links:

$$PR(p) = (1-d) + d \cdot (PR(q_1)/C(q_1) + \ldots + PR(q_n)/C(q_n))$$

where $C(q)$ is number of links going out of the page q, and d is damping factor which can be between 0 and 1 (usually d is set to 0.85).

Note that the sum of all ranks of all the webpages needs to add up to 1. In fact the links that go out of any page equally share its rank (due to the division of $PR(q)$ by $C(q)$). The page rank of a page p is the sum of the weights of all its incoming links. Figure 8 shows a simplified example where the PageRank values "flow" from pages to the pages they cite. After several iterations the PR values stabilize. Figure 9 shows an example of stable fix point.

Figure 8. Example of computing Page Rank scores.

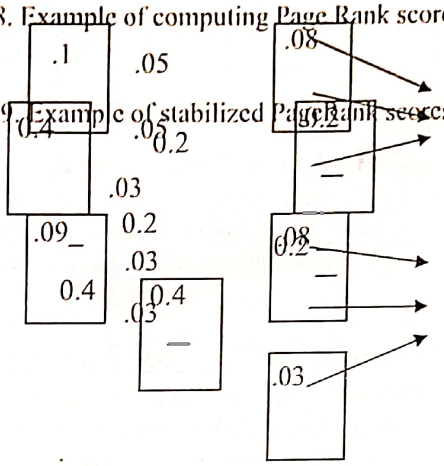
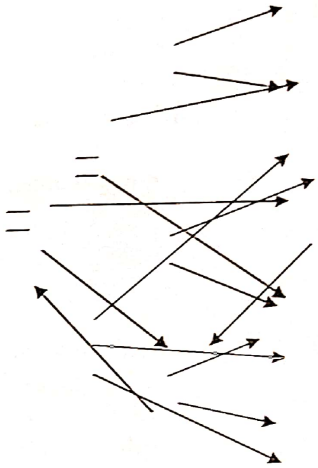


Figure 9. Example of stabilized PageRank scores.



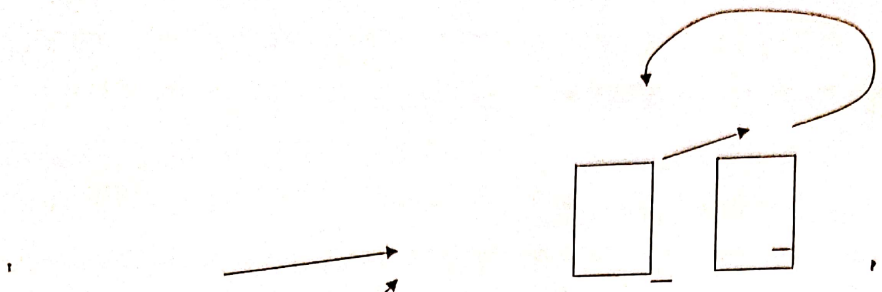


Figure 10. Example of "rank sink": when a group of pages only point to themselves but are pointed to by other pages.

There are complications when a group of pages only point to themselves but are pointed to by other pages; they act as a "rank sink" and absorb all the rank in the system (see Figure 10). That's why the additional factor *d* in the formula is needed. The PageRank algorithm models a "random surfer", that visits a page *p* with a probability given by its page rank *PR(p)*. The term *(1-d)* is the probability at each page that the "random surfer" will get bored and randomly jump somewhere else, allowing the surfer to get out of possible dead ends.

Commercial Aspects of Search Engines: Sponsored Links

Businesses pay to advertise on the major search engines. This would lead customers to their websites, if the customers following links returned by search engines as results of searching for specific terms. A business can bid for particular keywords. This is called *term leasing*. The links returned by search engines for commercial purposes are called *sponsored links*. They usually appear separate from other results, at the top, bottom or right side of the search engine results page. The *pay-per-click* advertising method allows search engines companies to charge a business proportional to the number of times users follow a sponsored link.

THE INVISIBLE WEB

The *Invisible Web* is the part of the Web that cannot be retrieved (seen) in the result pages from general Web search engines and in almost all subject directories.

Searchable Databases

Most of the Invisible Web is made up of the contents of thousands of specialized *searchable databases* that can be searched via the Web. The search results from many of these databases are delivered to the user in webpages that are generated just in answer to the search – *dynamic pages*. Such pages very often are not stored anywhere: it is easier and cheaper to dynamically generate the answer page for each query than to store all the possible pages containing all the possible answers to all the possible queries people could make to the database. Search engines cannot find these pages. If the only way to access webpages requires the user to *type something* or *select a combination of options*, spiders are unable to index the pages, because they cannot type or select options. Also, spiders crawl or navigate the Web by following the links in the webpages that are

already in pages collected by their parent search engine. If there is no link to a page, a spider cannot "see" it.

Some searchable databases require a fee, and the users log in using passwords. Many are free; here are a few examples of free searchable databases:

http://www.freepint.com/gary/direct.htm
http://opcit.eprints.org/explorearchives.shtml
http://www.deepwebresearch.info/

Of particular value in academic research are:

Librarians Index http://lii.org/
AcademicInfo http://www.academicinfo.net/
Infomine http://infomine.ucr.edu/

Excluded Pages

There are also some types of pages that search engine companies *exclude by policy*. There is no technical reason they could not include them if they wanted. It's a matter of selecting what to include in indexes that are already huge and expensive to operate.

Some search engines may choose not to include pages because the *format of the document* would be infrequently or unsuccessfully searched by the users of the search engine. Pages formatted in PDF or pages that have very little HTML text might be excluded (though lately Google and other search engines index PDF files by transforming them into plain text with minimal HTML markup). Search engines also have a hard time indexing the contents of documents in Flash, Shockwave, and other programs like Word, WordPerfect, etc. Pages consisting almost entirely of images are often excluded as well. *Script-based pages* are usually excluded. HTML links containing a ? lead to script-based pages. A script is a type of programming language that can be used to fetch and display webpages. They can be used to create all or part of a webpage, and to communicate with searchable databases. Most search engines are instructed not to crawl sites or include pages that use script technology, although it is often technically possible for them to do so. This is a policy decision. If spiders encounter a ? in a URL or link, they are programmed to back off, because they could encounter poorly written scripts or intentional "spider traps" designed to ensnare spiders, sometimes bogging them down in infinite loops that run up the cost and time it takes for spiders to do their work. This may result in the contents of an entire site using scripts being excluded from a search engine, or a search engine may crawl a safe part of a site and omit others.

OTHER TYPES OF INFORMATION RETRIEVAL SYSTEMS

Multimedia Information Retrieval

There is a lot of multimedia content on the Web. The information retrieval systems described above were adapted to work with collections of images, video, or music. A query can be expressed as text, or as a sample image, or by humming a melody. If the query is in text form, the IR system can use the text in the caption of the images, or the text description of the music (composer, singer, album, etc.) in order to find the information. In this case the traditional IR technology described above is used. If the query is an image or a piece of music, it can be treated as a digital signal. Techniques such as vector space model can be extended to compute the

similarity between two signals, where the features in the vectors will not be frequencies of occurrence in text, but features extracted by digital signal processing techniques.

A multimedia IR system differs from a traditional IR system in several ways. First, the structure of the multimedia objects is more complex than the structure of textual data: this requires integration of multimedia database management systems to adequately represent, manage, and store multimedia objects. Second, the similarity measure needs to be extended. The similarity measure is needed to match a query to a multimedia document, and to rank the retrieved multimedia documents. Third, query languages are more complex. Depending on the type of query, the search can be done only by content (image, music, etc.), only by text descriptions, or a combination of both. See chapters 11 and 12 of (Baeza-Yates and Berthier Ribeiro-Neto, 1999) for more details.

Digital Libraries

Traditional libraries are among the first institutions to use IR systems, to create catalogs of records for the material from the library. The catalogs can be search by users in the library or over the Web (online public access catalogs). These catalogs use database technology; the records are structured according to standards such as MARC (title, a few subject headings, and a classification number).

Modern libraries are being transformed to digital libraries as a result of the growth in electronic publishing, which makes information available in a digital form. Through the Web, a single interface provides access to local resources, as well as remote access to databases in the sciences, humanities, and business, including full-text journals, newspapers, and directories. Special collections, in multimedia not only in text format, become available through the same gateway. For more details about the technology of digital libraries see, for example, (Lesk, 1997).

Many libraries, particularly academic and large public libraries, have undertaken digital library project to achieve interoperability and ease of use and access. Two such projects are the Los Angeles Public Library's Virtual Electronic Library project (http://www.lapl.org), and University of Pennsylvania's Digital Library (http://www.library.upenn.edu). A digital library could have no connection to an actual library, for example the ACM Digital Library (http://portal.acm.org/dl.cfm) that contains journal and conference publications in Computer Science.

Digital libraries are more than complex IR systems. They are social systems centered around various communities of users. They also have component for building, cataloging, maintaining, and preserving repositories. There are many international or national digital library projects. One such project is the Digital Libraries Initiative (DLI) (phase one 1994-1998, phase two in progress) supported by the National Science Foundation (NSF), the Department of Defense Advanced Research Projects Agency (DARPA) and the National Aeronautics and Space Administration (NASA). The DLI phase one contained large research projects at six universities: University of Illinois Urbana-Champaign, Carnegie-Mellon University, Stanford University, University of California at Berkeley, University of California at Santa Barbara and University of Michigan. These projects are developing the next generation of tools for information discovery, management, retrieval and analysis.

Distributed Information Retrieval Systems

When the collection of documents is huge, it can be distributed over many computers. Parallel computing can be used to speed up processing. Document partitioning can be used to divide the search task into multiple, self-contained tasks that each involve extensive computation and data processing with little communication between them. Collections can be divided by topics, or for administrative purposes. When the collection is distributed, an index can be built for each partition, but a centralized index is still needed in order to direct the search for the terms in the user's query. To build a distributed IR system, algorithmic IR issues need to be considered together with engineering issues common to distributed systems in general. The main engineering issues involve: defining a search protocol for transmitting requests and results; designing servers that can efficiently accept requests, initiate subprocesses or threads to service requests, and exploit any locality inherent in the processing using appropriate caching techniques; designing a broker that can submit asynchronous search requests to multiple servers in parallel and combine the intermediate results into a final end user result. The algorithmic issues involve: how to distribute documents across the distributed search servers, how to decide which servers should receive a particular search request, and how to combine the results from the different servers.

A special type of distributed IR systems are Peer-to-Peer IR systems (P2P), when the information can be repeated on several computers, and there is no centralized access control. In a P2P system, the nodes (servers) are independent; each node can leave or enter the system any time. Examples of P2P systems are Gnutella and Napster. See chapter 9 of (Baeza-Yates and Berthier Ribeiro-Neto, 1999) for more details about distributed systems.

CONCLUSION AND FUTURE DIRECTIONS

This chapter presented an overview of the methods used in information retrieval and search engines. The technology of search engines is a very dynamic field, always looking for improvements and new ideas in order to satisfy user needs. Future trends in search engines include technology that is yet in the stage of research prototypes. Multimedia IR systems on the Web are becoming more important, as more video, music, and other types of data are available on the Web and fast Internet access becomes common.

Natural Language Queries

Text-based IR systems will also evolve. Users could express their queries in natural language, not just as keywords. This requires deeper syntactic and semantic analysis of the queries and the documents. Allowing the user to orally describe the information need into a microphone is a more natural way to interact with a search engine (Crestani, 2002). Spoken queries need to be translated into text queries using a Speech Recognition system (though current speech recognition technology would introduce recognition errors that might hurt retrieval performance). Cross-language Information Retrieval systems become available (Savoy, 2003). The queries can be a language in which the user feels comfortable, while the documents are in another language. This requires automatic translation of the queries before matching them to documents for retrieval.

Information Extraction techniques look at retrieving specific pieces of information from documents rather than showing to the user long lists of links to documents. These systems tend to work only in specific domains, such as biomedical text (Mooney and Bunescu, 2005) or newspaper text describing terrorist attacks (Rillof, 1999). Questions Answering techniques return a concise answer to a query expressed in natural language. They require deep semantic analysis in order to match queries to selected sentences in the documents (Harabagiu *et al.*, 2000). Simpler methods looked at exploring redundancy on the Web: extract answers from many webpages, and even if some answers are wrong, selecting the answer that has a majority of votes often leads to a correct response (Clarke *et al.*, 2001).

The Semantic Web and Use of Meta-Data

Most of the current forms of Web content are designed to be presented to humans: they are not understandable by computers. The Semantic Web aims at enhancing existing Web content with semantic structure in order to make it meaningful to computers as well as to humans. The Semantic Web project (http://www.semanticweb.org/) provides support for adding semantic annotations (meta-data that describes their content) to webpages or multimedia objects. To express the meta-data there is a need for standardized vocabularies and constructs explicitly and formally defined in domain *ontologies* (sets of domain concepts and relations between them).

The performance of current search engines and IR systems suffers because of the ambiguity of the natural language: words in documents and queries have multiple meanings and the retrieval results often include the wrong meanings in addition to the desired meanings. Better results will be achieved if webpages contain precise semantic annotations. This will allow search agents to navigate, collect, and utilize information on the Web in more reliable ways.

Visualization and Categorization of Results

Search engines tend to retrieve many documents in answer to a user query. Often users look only at the first 10 documents. When recall in important to the user, a long list is not a good way of displaying the results. The list does not show the distribution of the different categories of answers. Various ideas are tried in order to present the results in more manageable ways, for example 2-dimensional maps or 3-dimentional visualizations (Chen *et al.*, 1998). Automatic clustering techniques can be used to discover clusters of similar documents. Each cluster will then be an object in the visual representation.

GLOSSARY

Boolean Model: a classic model of document retrieval based on classic set theory. Uses Boolean operator such as AND, OR, NOT.

Digital Library: a complex system composed of: a repository of heterogeneous digital objects; descriptions of these objects (meta-data); a set of users; systems for capturing, indexing, cataloging, searching, browsing, delivery, archiving, and preserving the repository.

Distributed Information Retrieval: IR systems that distribute the data collection and the computation over multiple servers.

Index: a data structure built to speed up the search. For each keyword, it records the number of occurrences in documents and possibly other information.

Information Retrieval (IR): part of Computer Science that studies the retrieval of information (not data) from a collection of written documents. The retrieved documents aim at satisfying a user information need usually expressed in natural language.

Invisible Web: the part of the Web not indexed by search engines, mostly composed of searchable databases. These databases produce dynamic HTML pages as results to queries; therefore the pages cannot be indexed by search engines.

Latent Semantic Indexing: a model of information retrieval that extends the classic vector space model; it reduces the dimension of the vector space; the dimensions are no longer the index terms, they approximate concepts.

Mean Average Precision: an information retrieval performance measure that combines precision and recall and rewards relevant documents ranked higher in the list of retrieved documents. Computed as the average of the precision values for each relevant document in the ranked results.

Meta-data: description of the data (in XML or other description language)

Meta-search: a search technique where a single entry point is provided to multiple heterogeneous search engines. The user query is sent to these search engines and a unified list of results is presented to the user.

Multimedia Information Retrieval: IR systems that deal with images, video, audio, music or other multimedia objects.

Page Ranking: methods for ranking webpages by their popularity, for example based on the number of links that point to a page.

Peer-to-Peer Information Retrieval: Distributed IR systems where the nodes are independent computers that can leave or join the system any time.

Precision: an information retrieval performance measure that quantifies the fraction of the retrieved documents which are relevant.

Probabilistic Model: a model of information retrieval based on a probabilistic interpretation of document relevance to a user query.

Query: the expression of the user information need in the input language of the information system. Usually keywords and sometimes a few Boolean connectives (AND, OR, NOT).

Recall: an information retrieval performance measure that quantifies the fraction of known relevant documents that are among the retrieved documents.

Relevance Feedback: an interactive process of obtaining feedback from he user about the relevance or non-relevance of the retrieved documents.

Search Engine: An IR system designed to find information on the Web, to index webpages in order to be able to retrieve them as result of a user query.

Stemming: a technique for reducing a word to its root form.

Stopwords: words that occur frequently in texts, for example articles, prepositions, and conjunctions.

User Information Need: a natural language declaration of the informational need of a user.

Vector Space Model: a classic model of document retrieval based on representing documents and queries as vectors of index terms.

Web Crawler (Web Spider or Robot): a program that collects HTML pages from the Web by following links from the collected webpage.

BIBLIOGRAPHY

Ricardo Baeza-Yates and Berthier Ribeiro-Neto. 1999. *Modern Information Retrieval*, Addison-Wesley Publishing Company.

Sergey Brin and Lawrence Page. 1998. The anatomy of a large-scale hypertextual Web search engine. WWW7 / *Computer Networks* 30(1-7): 107-117.

Chris Buckley, Gerard Salton, and James Allan. 1993. Automatic retrieval with locality information using SMART. In *Proceedings of the First Text REtrieval Conference* (TREC-1), pages 59–72. NIST Special Publication 500-207.

Charles L. A. Clarke and Gordon V. Cormack and Thomas R. Lynam. 2001. Exploiting Redundancy in Question Answering. In *Research and Development in Information Retrieval*, pages 358-365.

Hsinchun Chen, Andrea L. Houston, Robin "R. Sewell, and Bruce L. Schatz. 1998, Internet browsing and searching: User evaluations of category maps and concept space techniques. *Journal of the American Society of Information Science*, 49(7):582-608.

Fabio Crestani. 2002. Spoken Query Processing for Interactive Information Retrieval. Data and Knowledge Engineering, 41(1):105-124.

Scott Deerwester, Susan T. Dumais, Thomas K. Landauer, George W. Furnas, and Richard Harshman. 1990. Indexing by latent semantic analysis. *Journal of the Society for Information Science*, 41(6):391-407.

Digital Library Initiative. 2006. http://www.dli2.nsf.gov/

Sanda Harabagiu, Marius Pasca and Steven Maiorano. 2000. Experiments with Open-Domain Textual Question Answering, in *Proceedings of COLING-2000*, August 2000, Saarbruken, Germany, pages 292-298.

John Kleinberg. 1999. Hubs, Authorities, and Communities. In *ACM Computing Surveys*, 31(4).

Michael Lesk. 1997. *Practical Digital Libraries; Books, Bytes, and Bucks.* Morgan Kaufmann.

Raymond Mooney and Razvan Bunescu. 2005. Mining Knowledge from Text Using Information Extraction. *SIGKDD Explorations, Special issue on Text Mining and Natural Language Processing*, 7(1):p.3-10.

Martin F. Porter. 1980. An algorithm for suffix stripping. Program, 14(3): 130-137. http://www.tartarus.org/martin/PorterStemmer/

Ellen Riloff. 1999. Information Extraction as a Stepping Stone toward Story Understanding. In *Computational Models of Reading and Understanding*, Ashwin Ram and Kenneth Moorman, eds., The MIT Press.

Stephen E. Robertson and Karen Sparck-Jones. 1976. Relevance weighting of search terms. Journal of the American Society for Information Sciences, 27(3):129-146.

Stephen E. Robertson, Steve Walker, and Micheline Hancock-Beaulieu. 2000. Experimentation as a way of life: Okapi at TREC. Information Processing and Management 36(1):95-108.

Gerard Salton and Chris Buckley. 1988. Term-weighting approaches in automatic retrieval. *Information Processing and Management*, 24(5):513-523.

Gerard Salton. 1989. Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer. Addison-Wesley Publishing Company.

Jacques Savoy. 2003. Cross-language information retrieval: experiments based on CLEF 2000 corpora. *Information Processing and Management* 39(1).

Semantic Web project. 2006. http://www.semanticweb.org/

DR.BABASAHEB AMBEDKAR MARATHWADA UNIVERSITY,AURANGABAD

Sagar B.C.A. College, Jalna

A Project Report On

Management of Intellectual Property Policy, Procedures and Forms

Submitted To

UNDER THE GUIDANCE OF

AJITKUMAR

Submitted by

WAKOKE MANISH SITARAM
M.LiB SY IV SEM

Year 2022-23

Department of Master Of Library and Information Science

(M.Lib. SY)

CERTIFICATE

This is to certify that, the following student

Wakode Manish Sitaram

Has successfully completed the summer internship project

Management of Intellectual Property Policy, Procedures and Forms

In the partial fulfillment of the requirement of Master Of Library and Information Science course asexpected by **Dr. Babasaheb Ambedkar Marathwada University, Aurangabad** for Academic Year **2022-23**

External Examiner

Internal Examiner

Head
Department of Humanities
Sagar BCA College, Jalna.

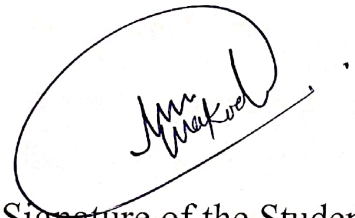
Principal

N. D. NAJARDHANE

PRINCIPAL
Sagar BCA College
Devmurti, Tq. Dist. Jalna

STUDENT DECLARATION

This is to declare that this Summer Training Project report on **"Management of Intellectual Property Policy, Procedures and Forms"** is a record of genuine work done by me under the guidance of **AJITKUMAR** in the partial fulfillment to the requirement for Master Of Library and Information Science I declare that this Summer Training project report is original and not submitted to anyother university before.



Signature of the Student:

Student's Name: Wakode Manish Sitaram

PRN No. 2021015200520832

Exam Seat No. CMLD401246

ACKNOWLEDGEMENT

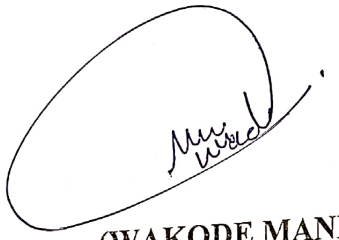
While conducting this report, I got support in many ways from many people .First I am deeply grateful to my project guide, **AJITKUMAR** who helped me with full devotion and always supported me earnestly whenever it was needed. Without his guidance, mental &moral support and academic inputs this report was not possible.

This Training report could never have seen the light of the day without his co- operation of those Clients who participated in this. I am thankful to all of them for giving me their valuable time.

My friends have been biggest support for me at every juncture of life. They manifested their great interest in my research work also and always tried to make thingseasy for me.

A word of gratitude goes to my family members whose love; affection and understanding have enabled me to complete this end with ease.

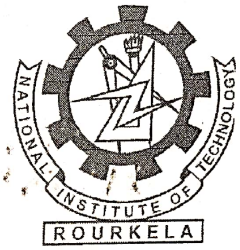
At the end, I thank to Almighty for giving me courage and strength to conduct this project report.



(WAKODE MANISH)

Management of Intellectual Property

Policy, Procedures and Forms



CONTENTS

TOPIC

Distance Education Materials
Implementation of Intellectual Property
Policy

6. Procedures for Management of Intellectual Property
Creating IP Awareness
Proposals for patent
application
Archiving theses containing Intellectual Property with commercial potential

FORMS

NITR/IP/1	Undertaking to be signed by all academic and technical staffjoining National Institute of Technology, Rourkela
NITR/IP/2	Undertaking by a person engaged by the Institute under "Workfor hire" terms
NITR/IP/3	Declaration by students (initiating work on Patentable Technologies)
NITR/IP/4	Handling and archiving of theses and dissertations submitted tothe National Institute of Technology, Rourkela
NITR/IP/5	Invention and Technology Disclosure Form
NITR/IP/5A	Invention and Technology Disclosure Form (Summary)
NITR/IP/6	Mutual Secrecy Agreement between NIT Rourkela and collaborating Institutions
NITR/IP/7	Letter of agreement between NIT Rourkela and collaboratinginstitutions for the Transfer of Proprietary Materials
NITR/IP/8	Letter of agreement for developing educational material maintained by a private firm

1. PREAMBLE

National Institute of Technology, Rourkela (hereinafter called the Institute) is dedicated to research, teaching, and extension of knowledge to the public. The Institute recognizes its responsibility to produce and disseminate knowledge. Inherent in this responsibility is the need to encourage the production of creative and scholarly works and the development of new and useful materials, devices, processes, and other intellectual property, some of which may have potential commercial value. These activities contribute to the professional development of the individuals involved, enhance the reputation of the Institute, provide additional educational opportunities to students, and promote public welfare. The Institute has a responsibility of bringing new knowledge into use by the general public. Such knowledge or technology often has commercial value and should be treated as a financial asset to be used, conserved and applied in such a way as to generate an appropriate financial return. Transfer of such information or technology through licensing satisfies both the above objectives, i.e., dissemination for use and realization of financial returns.

Technological and social developments in recent years have broadened the scope of information and technology that can have potential commercial value and, therefore, should be treated as assets subject to Institute ownership and control. In addition to new machines, compositions of matter, and written materials which traditionally have been the subject of patents and copyrights, computer software, video courses, etc. are now normal outcomes of Institute activities. Thus, a broad policy covering all aspects of intellectual property needs to be created to provide widespread protection to the originators of such property.

Over the past decade, appreciation of the commercial value of intellectual property has grown both within the academic community and in the society at large. Concerns related to confidentiality, publication, and ownership of intellectual property are now commonplace. The pace of modern science, resulting in new and useful inventions, initiated a need for a central policy in determining the course of the creation, protection, and commercialization of intellectual property in the Institute. This has resulted in establishment of the Intellectual Property Committee (IPC) and the Intellectual Property Policy (IPP) to encourage creation and protection of intellectual property in the Institute.

This IP Policy applies to all Institute employees and students, regular or contractual. Every member of the academic community, student, non-teaching and teaching staff alike, must be knowledgeable about intellectual property both to protect their own rights and to respect the rights of others. The Institute IP Policy is intended to encourage a healthy atmosphere conducive to research and development through a generous system of rewards and incentives for the creation of intellectual property while at the same time giving proper consideration to the economic rights and responsibilities of the Institute.

The strength of the Institute lies in its faculty, students, technical and administrative staff. This document is intended to introduce, regulate, and organize issues related to intellectual property within the Institute. It also reaffirms the Institute's commitment to scientific endeavors, academic excellence and the dissemination of knowledge. This policy is intended to spell out the responsibilities of the Institute and its employees and to establish a framework for ethical conduct.

2. THE INTELLECTUAL PROPERTY (IP) POLICY

The primary objective of the IP Policy of NIT Rourkela is to establish appropriate principles for creation, protection, ownership and management of intellectual property in the Institute. The endeavor is to provide an intellectual property environment that encourages the development of inventions and other intellectual creations for the best interest of the public, the creator, and the research sponsor, if any, and will permit the timely protection and disclosure of such intellectual property either by development and

commercialization after securing available protection, by publication, orboth.

This Policy is further intended to protect the respective interests of all participants by ensuring that the benefits of such property accrue to the public, to the inventor, to the Institute and to sponsors of specific research projects in varying degrees of protection, monetary return and recognition, as circumstances justify or require.

Objectives of the IP Policy :

The major objectives of the IP policy of NIT Rourkela are:

- To provide a superior environment to the employees and students of the Institute for creation, protection,and commercialization of intellectual property and to stimulate innovation.
- To encourage research, scholarship, and a spirit of inquiry, thereby generating new knowledge.

- To facilitate the transfer of knowledge and technology to intending users to promote utilization of such resources for benefit of the society.
- To provide an administrative system to determine the commercial significance of discoveries and developments and to assist in bringing these into public use.
- To provide for a equitable distribution of economic gains resulting from new intellectual property among the developer, author, or inventor (the originator), the Institute, and, where applicable, the sponsor.
- To provide incentives to originators in the form of personal development, professional recognition, and financial compensation.
- To safeguard, review and manage the intellectual property so that it may receive adequate and appropriate legal protection against unauthorized use.
- To encourage students at all levels to develop patentable technologies and to provide financial assistance from the Institute to the extent possible.
- To create awareness on IPR through conducting seminars, conferences, invited talks and lectures, and training programs among the academic community.
- To create respect for other people's intellectual property among members of the Institute community.

SCOPE OF THE IP POLICY

The IP Policy applies to potentially patentable inventions and discoveries, industrial designs, copyrightable materials such as books, publications, electronic courseware, computer programs, electronic circuits etc., protectable trademarks and trade secrets, which are developed using Institute equipment, supplies, facilities, employee time,or trade secret information, or which relate directly to the Institute's business, research or development. The Institute will encourage, recognize and protect all creative and scholarly works in form of patents, copyrights, industrial designs, trademarks and trade secret, as the case may be, developed by its employees/students as a result of their research or employment. The Institute will protect the rights of the originator regarding intellectual property created by him as per legal framework of the land. However, the Institute, as a corporate body, shall share this right with inventors – staff, student or guest, in a just and fair manner.

This Policy applies to all the employees of the Institute and visitors using Institute facilities under the supervision of Institute staff and to all the students of the Institute including doctoral and Postdoctoral fellows.

It covers all intellectual property conceived, first reduced to practice, written, or otherwise produced by all faculty, staff, and students of the Institute whether using Institute resources or not. It also covers intellectual property created by part time students, employees and visitors using Institute funds, facilities or other resources.

3. DEFINITIONS

Intellectual Property (IP): For the purpose of this policy, "Intellectual Policy" is defined as the tangible or intangible results of research, development, teaching, or other intellectual activity. Intellectual property may include the following products:

a. Patents on new and useful scientific or technical advancements by way of inventions,

discoveries, processes, computer hardware and software, unique materials, machines, devices, instruments, apparatuses, circuits, plant varieties etc.

b. Copyright in industrial and architectural design, models, engineering drawings, integrated circuit layout designs, computer software, animations and visualizations, information technology products and processes including hardware and software features, original innovative, creative or artistic works and their derivatives or adaptations, whether dramatic, musical, literary works, work of graphics or plastics art and cinematographic and animated films, teaching material for classroom and online courses such as courseware for distance education, original data and records of research, undisclosed and/or unpublished information etc.

c. Trademarks, service marks, logos, collective marks, certification marks, trade names etc.

possible commercialization. He has to ensure the use of licensed software and fair and just treatmentof others' intellectual property by the Institute and its members.

Intellectual Property Committee (IPC) : The Institute's Senate shall appoint an Intellectual Property Committee (IPC) to help administer intellectual property and to make suitable recommendations to Dean (SRICCE)/Director for implementation. Dean (SRICCE) will serve as the Chairman, and Professor-in-Charge (IP) shall serve as the Secretary of the IPC. In addition to Dean (SRICCE) and Professor-in-Charge (IP), the Senate will nominate two more members from among the faculty of the Institute. The tenure of Professor-in-Charge (IP) and the two members will be two years, preferably non-concurrent. A member may be appointed for a second term, but not more than4 (four) years in total.

The IPC will assist various departments and centres of the Institute in all matters relating to intellectual property. It will help various departments to secure protection for intellectual property where appropriate. It will maintain central databases and files of patent applications, issued patents, trademarks and copyrights, licenses and

agreements, coordinate with various departments in negotiating and preparing license and other agreements, review and approve all agreements relating to intellectual property. It will review causes of possible infringementson the Institute's intellectual property and take action as deemed necessary.

The legal interests of the Institute and its staff, faculty and students in any intellectual property, except traditional scholarly works, shall be determined by the IPC in accordance with the policy enumerated under items.

5. POLICY

General Policy

The intellectual property policy shall apply to all persons employed by the Institute – full-time and part- time faculty, visiting faculty, scientists employed by the Institute, as well as technical and administrative staff. It also applies to undergraduate, postgraduate and doctoral students as well as postdoctoral fellowsand visiting scientists.

This policy shall apply to all kinds of intellectual property (including, but not limited to, any invention, discovery, trademark, copyright, trade secret, technology, scientific or technological development, research data and computer software) regardless of whether the intellectual property is subject to protection under patent, trademark, copyright, or any other law. The institute will encourage and recognize the originator of intellectual property and protect the ownership for the creators.

The Institute will work towards protection through legal means of all creations of scholarly and educational materials, inventions, products, processes, art works, musical compositions and dramatic and non dramatic literary works related to the author's academic or professional field, regardless of the medium of expression. All such intellectual property shall be jointly owned by the originator/author and the Institute.

The Institute shall have sole ownership of all intellectual property created by an employee who was hired specifically to work on a target product or process (or other intellectual property) or was commissioned by the Institute or a component of the Institute for the specific objective leading to creation of the intellectual property. The Institute will assert its ownership of all intellectual property created by the outside agencies commissioned by the Institute for the specific purpose.

The intellectual property generated from research projects sponsored by government/ non-government agencies will be owned by the creator(s), the Principal Investigator or Chief Consultant, the Institute and the sponsoring agency. The sponsoring agency will bear 50% of the protection cost or forgo the rights to the intellectual property. In case the project was accepted by the Institute under terms different from thatstated herein, the terms agreed to shall prevail.

Intellectual Property Rights and Obligations

Intellectual property generated by a full-time employee or a full-time student of the Institute is the joint property of the originator and the Institute whether Institute resources are used or not. If an Institute employee or a full time student creates intellectual property while working in another organization, it will be jointly owned by the creator, NIT Rourkela and the host institution. In case of part-time employees or students or visiting professionals, intellectual property generated by use of Institute facilities and/or support only come under joint ownership of the Institute and the originator.

the geographical scope and duration of such protection shall be final. If the originator intends to seek protection internationally or through Patent Cooperation Treaty (PCT) application, the institute will permit the originator to proceed on his own or with the help of other individuals or agencies.

Royalty Income Sharing

5.3.1 In case the Institute succeeds in commercialization of intellectual property for the creator and licenses rights to third parties in consultation with the creator(s), the revenue generated through royalty paymentswill be equitably shared among the creators and the Institute.

Inventor(s)	Institute
60 %	40 %

In case of multiple originators of an IP, all the originators will decide among themselves how to share the proceeds of an intellectual property. If they fail to arrive at a consensus, the IPC will analyze all available information and make a recommendation to the Director. The decision of the Director shall be binding and final.

If there are other legitimate claimants to the IP, they will be grouped either under "inventors" or "Institute". While sponsoring or supporting organization will get their share of the proceeds from "Institute" share, individuals (including visiting professionals) who contributed to the invention will receive their share fromthat of the originators.

Ownership of Intellectual Property in Certain Circumstances

Where research has been sponsored by a private industry/ foundation or government agency and no prior agreement exists on sharing of intellectual property, licensing of patents shall be negotiated between thesponsor and the Institute.

The intellectual property policies and guidelines of the Institute are subject to, and thus amended and superseded by the specific terms pertaining to intellectual property rights included in Central or State grants and contracts, or grants and contracts with NGO's or private sponsors.

If the intellectual property has been generated as a work-for-hire, the employee or agency will retain the moral right to be identified as the creator of the intellectual property but right of commercialization restsonly with the Institute.

Ownership of Intellectual property Generated by students

It is a requirement in academics that a student must own the copyright of the thesis (since it is his or her original work) which he or she submits as partial fulfillment of the requirements for an academic degree. However, the student will grant a non-exclusive, non-transferable royalty-free license to the institute to use, in the course of non-commercial academic activity, the records and data generated in the course of his research. Furthermore, it is possible that the research that the student carries out as part of the program of study may result in the generation of intellectual property other than the text of the thesis. Supervisors should advise students during the course of their work that certain kind of research may lead to the generation of intellectual property which will require protection of its commercial value through confidentiality, for which the student will have to forgo publication during the period of sealing of a patent. Care should be taken at all stages to see that no conflict of interest arises between

If the intellectual property has been generated as a work-for-hire, the student will retain the moral right to be identified as the creator of the intellectual property, but right of commercialization rests with the Institute.

Disclosure and Confidentiality.

At an appropriate stage in the development of an invention, the originator shall make a written disclosure of the concepts to the IPC, providing all such particulars as are vital to judge its commercial prospects. The IPC shall promptly acknowledge, in writing, its receipt of the disclosure and the date of receipt. The originator shall send one copy of his proposed manuscript, prior to submission of thesis, to the IPC.

All the departments in the Institute will be bound by the non-disclosure and confidentiality terms to be clearly spelled in a separate document. Each department is under obligation to file their R&D manuscripts, ifany, on time to time basis, with the IPC. It is expressly understood by the departments that any information which relates to any Invention should be treated as Intellectual Property and therefore is not to be divulgedwithout the prior consent of the IPC.

The Originator who has communicated with the IPC under Clause 5.6.1 shall refrain from publishing, reading, dissipating, circulating or disclosing the conception in any form whatsoever, since non-disclosure is one of the most important qualifications for intellectual property protection. The originator may disclose such conception, upon a prior written permission from the IPC, once an application for a patent, trademark or copyright has been made on the conditions described herein and the commercial rights in the conception are secured to the Institute.

Commercialization.

For purposes of protection and commercialization of intellectual property on behalf of the Institute, patent, trademark or copyright coverage may be sought, or the property may be treated as proprietary information, technical know-how, or trade secret.

The IPC may determine whether the Institute has a legal interest in the commercialization of the property.However, the Institute is not legally bound to commercialization of each property and the originator maynot claim such right. It shall be in the sole discretion of the Director on advice of the IPC to determine commercialization of the property.

In seeking and developing commercialization of intellectual property, the Institute shall be guided by thefollowing principles:

(a) A primary objective and responsibility of the Institute shall be to assure that the products of itsintellectual activity are brought into the widest possible use for the general benefit of society.

(b) Intellectual property should be treated as an asset and an appropriate return should be sought.

(c) Active participation of the originator in all commercialization efforts shall be sought.

Electronic and Distance Education Materials

All original works submitted by the contributors for the purpose of electronic and/or distance education coursedevelopment shall remain the property of the concerned contributors and the Institute.

The Institute has the right to use the course (including all related materials) developed by the faculty member and/or other employees involved in the development of a distance learning and/or e-learning course for the Institute's own educational, research, and other purposes without any additional compensation to the faculty member or any other employee who is an author of the course. Academic departments determine which courses will be offered and who will teach these courses.

If the course and/or related materials are licensed, sold, or otherwise conveyed to a third party, the mechanism as per Clause 5.3.1 will govern the distribution of any proceeds. If the material is used in distance education activity generating revenue under a scheme where instructors or other Institute personnel receive specific compensation, the originator of the courseware is entitled to a fair share of the proceeds, the exact amount being determined by the Director.

member/technicalstaff signs the undertaking form no. NITR/IP/1 at the time of joining.

- At the time of registration for Autumn Semester, each student must sign and submit a declaration in formno. NITR/IP/3 to Prof-In-Charge (IP).
- All departments will provide financial support in a fruitful means to all student projects with possible commercisable outcome.
- The institute will provide additional funds with higher outlay for project works with a commercial potential,wherever necessary.
- The IPC will bring out brochures/newsletters and make necessary announcements in various media for creating awareness among academic community in regards to advancements on technology, patentable and commercial technologies, IP laws and amendments and legal aspects related to IP.

- The IPC will conduct seminars, workshops, talks by eminent scientists and practitioners on IP related issues,and shall organize student groups to create greater participation.
- At the time of submission of thesis, each student must sign and submit a declaration in form no. NITR/IP/4 to Prof-In-Charge (IP).

Proposals for patent application

Faculty members, technical staff and students interested to protect their intellectual creations under IP law of the land may apply to Professor-in-charge (IP) using the Invention and Technology Disclosure Form(Form No. NITR/IP/5).

Professor-in-charge(IP) will draw the attention of IPC members in a meeting for evaluating the IP substance for possible protection within two weeks. If the members agree to file for protection, the IPC will approach appropriate Government, private and legal entities to go forward with protection of the IP with due recommendation from Director.

The expenditures for protection of IP substance will be borne by the Institute from its non-plan "administrative expenses" head.

If the IP substance is not fully developed for possible protection, IPC will guide the originators whereto improve it. IPC may also give guidance on drafting the Patent forms etc. even with provisional specifications.

The Institute shall bear all the charges for patent search while filling up the patent form.

If the patent is granted, it becomes the joint property of the originator and the Institute.

The Institute has the prerogative of finding a suitable partner for commercialization of the patents for firsttwo years from the date of grant of the patent.

After two years, the originator may choose a suitable partner for commercialization of intellectual propertycreated by him/her. However, benefit sharing mechanism will be adhered to as per Clause no. 5.3.1. The originator, before going for technology transfer on his own, must seek the permission of IPC. The IPC should strive to dispose off the matter within two weeks.

Archiving theses containing Intellectual Property with commercial potential

Theses submitted by the students may have potential IP substance. It should be protected and commercialized for greater interest of humanity. Therefore, the students must come forward and the supervisors must motivate the students to work on patentable technologies. The student submitting a thesis must give a declaration in the Form no. NITR/IP/4.

The IPC will examine all the theses and find the suitability of protection of IP created by the students.

The Institute will also try for potential commercialization when a patent is granted. However, the Institutehas the prerogative of commercializing only for the first two years from the date of grant of patent. The revenue sharing mechanism will be adopted as per Clause 5.3.1.

If the Institute does not show any interest for patenting an invention contained in a thesis, the student may go for protection with formal information to the Institute. In such case, the thesis is protected and not disclosed to anyone for a period of one year after such a request is received

Form No. NITR/IP/1

National Institute of Technology
Rourkela

Undertaking to be signed by all academic and technical staff joiningNational Institute of Technology, Rourkela

1. This is to declare that I have read and understood the policy of the National Institute of Technology, Rourkela with respect to intellectual property and the rights therein, titled ['Policy'] and that I agree to be bound by it and to follow its provisions during the period of my employmentby the Institute.

2. I agree to report, disclosing full details, to the relevant authorities of the Institute any patentable or commercializable intellectual property that I may generate or participate in generating in accordance with the provisions of the Intellectual Property Policy.

3. (Strike out whichever is not applicable)

   (a) I certify that I am at present under no contractual obligation with any person ororganization, which are in conflict with the Policy.

   or

   (b) I am at present under the contractual obligations detailed below:

   _____

   _____

   _____

   (attach a separate sheet, if necessary)

4. I agree to share all intellectual property generated during the course of my work with the Institutein accordance with the Intellectual Property Policy of the Institute in vogue at the time of creationof the intellectual property.

5. I undertake to behave with dignity and broadness of mind while sharing intellectual property rights with my coworkers – students, faculty, technicians and other supporting staff as well as visitors.

Name :

Designation :

Department :

Employment Code :

(Signature)

Date

Form No. NITR/IP/3

National Institute of Technology
Rourkela

Initiating work on Patentable Technologies

Declaration by the Student(s) engaged in Project / Dissertation works

I/We, Mr./Ms./Dr. _____ Roll No.: _____ and Mr./Ms./Dr._____Roll No.:_____registered as research scholar or student of programs such as B.Tech./M.Tech./ M.Tech.(Res) / M.Sc. / Ph.D / D.Sc. in the Department of____, National Institute of Technology, Rourkela (hereinafter referred to as the "Institute") do hereby declare that I/we shall start working on a project entitled :_____

_____.

I/We declare that :

1. I/We shall respect the intellectual property of others. I/We shall not knowingly or unknowingly use any protected inventions / designs / materials / integrated circuits held by others for our purpose without paying the license fee.

2. I/We shall not use any unlicensed modeling / drafting / word processing /programming software for my/our purpose.

3. I/We shall not use any material protected under copyright law except beyond thescope of fair use for our purpose.

4. I/We shall request IPC for helping us for patent search for our work, the cost if any being borne by the Institute.

5. In the course of project work, if any IP is generated I/we shall proceed as the IP policy of the Institute for possible protection and subsequent commercialization.

Name of the Student

Signature of student
with date

Name of the Student

Signature of student
with date

Name of the Supervisor

Signature of the Supervisor with date

Signature of the Head of the Department with date

Form No. NITR/IP/4

National Institute of Technology Rourkela

Handling and Archiving of Theses and Dissertations submitted to theNational Institute of Technology, Rourkela

Declaration by the Author of the Thesis or Dissertation

I, Mr./Mrs./Miss/Dr._____Roll no._____ registered as a research Scholar or a student of programs such as B.Tech./ M.Tech./ M.Tech(Res)/ M.Sc./Ph.D/D.Sc in the Department of _____ National Institute of Technology, Rourkela (hereinafter referred to as the 'Institute') do hereby submit my thesis, entitled: _____ _____ (herein referred to as 'my thesis') in printed as well as in electronic forms for holding in the library of records of theInstitute.

I hereby declare that:

1. The electronic version of my thesis submitted herewith on CDROM is in PDF format.

2. My thesis is my original work of which the copyright vests in me and my thesis does not infringe or violate the rights of anyone else.

3. The contents of the electronic version of my thesis submitted herewith are the same as those submitted as final hard copy of my thesis after my viva voce and adjudication of my thesis on____(date).

4. I agree to abide by the terms and conditions of the Institute Policy on Intellectual Property (hereinafter Policy) currently in effect, as approved by the competent authority of the Institute.

5. I agree to allow the Institute to make available the abstract of my

thesis to any user in both hard copy (printed) and electronic forms.

6. For the Institute's own, non-commercial, academic use I grant to the Institute thenon-exclusive license to make limited copies of my thesis in whole or in part andto loan such copies at the Institute's discretion to academic persons and bodies approved from time to time by the Institute for non-commercial academic use. All usage under this clause will be governed by the relevant fair use provisions in the Policy and by the Indian Copyright Act in force at the time of submission of the thesis.

7. I agree to allow the Institute to place such copies of the electronic version of my thesis on the private intranet maintained by the Institute for its own academic community.

8. I agree to allow the Institute to publish such copies of the electronic version of my thesis on a public access website of the internet.

9. If in the opinion of the Institute my thesis contains patentable or copyrightable material and if the Institute decides to proceed with the process of securing copyrights and/or patents, I expressly authorize the Institute to do so. I also undertake not to disclose any of the patentable intellectual properties before being permitted by the Institute to do so, or for a period of one year from the date of final thesis examination, whichever is earlier.

10. In accordance with the Intellectual Property Policy of the Institute, I accept that any commercialisable intellectual property contained in my thesis is the joint property of myself, my coworkers, my supervisors and the Institute. I authorize the Institute to proceed with protection of the intellectual property rights in accordance with prevailing laws. I agree to abide by the provisions of the Institute Intellectual Property Right Policy to facilitate protection of the intellectual property contained in my thesis.

11. If I intend to file a patent based on my thesis when the Institute does not wish so, I shall notify my intention to the Institute. In such case, my thesis should be marked as patentable intellectual property and access to my thesis is restricted. No part of my thesis should be disclosed by the Institute to any person(s) without my written authorization for one year after my informing to the Institute to protect the IP on my own, within 2 years after the date of submission of the thesis or the period necessary for sealing the patent, whichever is earliest.

Name of student: **Name of supervisor(s):**

Signature of student: **Signature of supervisor(s):**

Signature of the Head of the Department

Form No. NITR/IP/5

National Institute of Technology Rourkela

Invention and Technology Disclosure Form

1. Title of the invention:

For Office use only

Proposal ID : NITR/IP/_____ - _____ / _____ Date of Receipt : _____

Acad yr Sl. No.

2. Inventors:

[*For visiting scientists, please give details of substantive employer.*]

SlNo.	Name	Employee Codeor Roll No	Position	Department	Email

3. **Brief description of the invention:** *(How this invention relates to new processes, systems, machines, compositions of matter etc.)*

4. **Detailed description of the invention**

State of prior art

(a) Prevailing **state of the art**?

(b) **Literature search** relating to this invention? *[Please include a copy of the resulting documentation, and reprints of publications.]*

(c) **Prior art/patent search** relating to this invention? *[Please include a copy of the resulting documentation, and reprints of patent documents: if a computer database search has been resorted to, please give the web site details and the Key Words used in the search.]*

Description : *(Describe the invention so that other Institute faculty who are knowledgeable in the field can evaluate its technical and commercial merits.)*

Novelty : *(Highlight the features described above that make the invention novel.)*

Inventiveness : *(Are the novel features inventive based on 4.1(a) above; and, if so how?)*

Advantages *(over comparable inventions or practices)*:

Testing : *(Has the invention been tested experimentally? If so details of experimental data to be supplied.)*

Form No. NITR/IP/5

National Institute of Technology Rourkela

Invention and Technology Disclosure Form

1. Title of the invention:

For Office use only

Proposal ID : NITR/IP/ ____ - ____ / ____ Date of Receipt : _____

Acad yr Sl. No.

2. Inventors:

[*For visiting scientists, please give details of substantive employer.*]

SlNo.	Name	Employee Codeor Roll No	Position	Department	Email

3. **Brief description of the invention:** *(How this invention relates to new processes, systems, machines, compositions of matter etc.)*

4. **Detailed description of the invention**

State of prior art

(a) Prevailing **state of the art**?

(b) **Literature search** relating to this invention? *[Please include a copy of the resulting documentation, and reprints of publications.]*

(c) **Prior art/patent search** relating to this invention? *[Please include a copy of the resulting documentation, and reprints of patent documents: if a computer database search has been resorted to, please give the web site details and the Key Words used in the search.]*

Description : *(Describe the invention so that other Institute faculty who are knowledgeable in the field can evaluate its technical and commercial merits.)*

Novelty : *(Highlight the features described above that make the invention novel.)*

Inventiveness : *(Are the novel features inventive based on 4.1(a) above; and, if so how?)*

Advantages *(over comparable inventions or practices)*:

Testing : *(Has the invention been tested experimentally? If so details of experimental data to be supplied.)*

5. Funding and support

I. Was there significant use of Institute equipment and facilities? **Yes/No**

II. Was the invention supported by research grants/contract from external sources? **Yes/No**

If **YES**, please give details :

(a) Sponsor : _____

(b) Grant/contract no. : _____

(c) Period of grant/contract : _____

(d) Principal investigator and co-investigator: _____
(even if they are not inventors within the purview of this document and will not share the credit androyalties)

(e) Has the sponsor been informed of the invention? *(state whether required under grant/contractaward conditions)*

(f) Was the work done under any other agreement? Give details.

6. Information for protection of IPR: conception and disclosure *(Accurate data is required as priordisclosure may affect possibility of obtaining patent rights.)*

	Date	References/comments
Date of conception of this invention.Has this date been documented? If so, where and how?		
Has this invention been presented at seminars/discussions other than those which form the requirement for the degree program of the student?		
Please provide the anticipated date of submission for publication or communication for presentation at seminar/conference etc. (Should not be earlier than one month from thisdate)		
Has the invention been reduced to practice?		

Potential marketability including commercial suggestions [viable size of industry, equipment, rawmaterial and manpower requirement under different skill levels, import component, exportpotential, other relevant economic information]

8. Prior disclosure and possible intent :

Has the invention been disclosed to industry representatives or their parties?

Has any commercial organization shown interest in this invention? Give details.

9. Development Stage :

What is the current stage of development of the invention as it relates to commercial utilization and marketability :

☐ Embryonic
Partially developed
Fully developed

10. Potential for international patent :

Does the invention have significant commercial potential in foreign countries?If so, where? Give details.

11. Declaration :

I/We declare that all statements made herein are true to the best of my/our knowledge. I/We hereby agree to hold the right of intellectual property of this invention jointly with National Institute of Technology, Rourkela. National Institute of Technology, Rourkela will share any royalty income derived from the invention with the inventor(s) according to the IP policy of the Institute in force. Intellectual Property of this invention will be protected by National Institute of Technology, Rourkela from time to time based on its merit and commercial viability.

SlNo.	Name	Signature	Date	Place

Form No. NITR/IP/5A

National Institute of Technology
Rourkela

Invention and Technology Disclosure Form
(Summary)

1. **Title of the invention:**

2. **Inventors:**

[*For visiting scientists, please give details of substantive employer.*]

Sl.No.	Name	Employee Code or Roll No	Position	Department	Email

3. **Brief description of the invention:** *(Not to exceed 100 words)*

4. **Prevailing state of the art :**

(a) Details of Patent search sites or other resources.

(b) Key words used for patent search.

(c) List of patents related to present invention.

5. **Novelty**

6. **Inventiveness**

7. **Advantages**

8. **Commercial Potential :** *(List of organization with possible interest in the invention)*

9. **Signature of Inventor(s) with date**

Form No. NITR/IP/6

National Institute of Technology
Rourkela

Mutual Secrecy Agreement
between
NIT Rourkela and collaborating Institutions

This agreement is between National Institute of Technology, Rourkela (the provider organization) and/or provider scientist(s),jointly called the first party, and the organization_____ _____(the recipient organization), the second party.

1. The first party is the owner of the invention called _____ (description) and/or owner of certain technical data/process technology/other information (proprietary information)developed through their own efforts.
   The organization_____(name), the second party, is the owner ofthe invention called_____(description) and /or certain technical data/process technology/other information (proprietary information) developed through its own business and R & D efforts.

2. The invention and/ or the proprietary information of the provider scientist(s), NIT Rourkela and _____ (organization) is proprietary and confidential and not public knowledge. This will be disclosed to one another under the terms of this agreement.

3. The parties to this agreement consider it desirable for each other to have access to above invention/proprietary information for discussing and evaluating possible collaborative research and development work and/or licensing activities relating thereto.

Therefore the parties agree as follows:

- All invention/proprietary information as used in this Agreement provided by one party is proprietary and confidential in connection with evaluation of invention and/or proprietary information for collaborative R & D and/or licensing work. These are disclosed to one another inwriting clearly marked confidential OR arise out of discussions during visits to laboratory/plants or any other facility of either party, and reduced to writing within thirty (30) days of such discussion.The date and time of the visit and personnel present during the visit should be recorded in writingby both parties.

- All parties agree to hold in confidence any or all invention/proprietary information disclosed andfurther agree not to disclose the same to third parties or use it for any other purpose other thandiscussion and internal evaluation provided in this document. However,

either party may disclose the invention/information/technical data/technology to its own employees assisting that party in making an evaluation, provided that all such employees shall have agreed to be bound by the secrecy terms of this agreement.

- The recipient of tangible products or materials consisting invention/technology from the other party agrees not to analyze or have a third party to analyze such tangible products or materials.
- All invention/proprietary information is and remains the property of the disclosing party and must be returned, in a form suitable to be returned, within ninety (90) days after the disclosing party makes a written request for its return or at the conclusion of evaluation or termination of the Agreement.
- The evaluation period during which information will be exchanged will be ordinarily one (1) year from the date of signing this Agreement unless extended by mutual consent of the parties in writing.
- The foregoing obligation with respect to invention/proprietary information received by any party who are signatories to this Agreement shall survive in the event of termination of this agreement.

This agreement is effective as of_____, and shall terminate on _____. The two parties can extend the agreement through mutual consent, in writing, andthe extension period shall be on a yearly basis. Either party may terminate this Agreement at its discretion immediately upon written notice to the other party.

This agreement is signed on_____ _____between:
NIT, Rourkela

Organization:

_____

(Name of signatory) **(Name of signatory)**

Designation : Dean (SRICCE) **Designation**

Address : National Institute of Technology **Address**
Rourkela - 769008

National Institute of Technology
Rourkela

Form No. NITR/IP/7

Letter of Agreement between NIT Rourkela and collaborating institutionsfor Transfer of Proprietary Materials

This agreement is between National Institute of Technology, Rourkela (the provider organization) and/or provider scientist(s) _____ _____ [name(s)], jointly called the first party, and the recipient scientist(s) _____ [name(s)] of organization_____ (the recipient organization), the second party.

The Material that is covered by the agreement includes _____ (description of the material) which is considered as proprietary material of the provider and NIT Rourkela. The provider scientist and NIT Rourkela shall be free, in their sole discretion, to distribute the Materials to others and to use it for their own purpose.

In response to the second party's request for Material, both the parties agree to the following before second party receives the Material :

(1) The Material shall be used by_____(recipient scientist) working at _____ (recipient organization) in research to study _____ (description of work). The material will be used for teaching and not-for-profit research purpose only and not for use in any product or process for profit-making commercial purpose. The material is provided at no cost or with a fee ofRs._____.

(2) Neither the Material nor this material treated by any means will be used in human subjects.

(3) The second party shall not distribute, release or disclose the Material to any person or entity other than laboratory personnel under recipient scientist's direct supervision, and the second party must undertake to ensure that no one will be allowed to take or send Material to any other location unless written permission is obtained from the first party.

(4) The second party is not allowed to have a third party analyze such tangible products or materials obtained from the first party without written and specific authorization from the first party.

(5) The second party will give a brief description of its research program and the nature of usage _____

of the material to the first party. The second party will acknowledge the first party before any publication or presentation based on research results with supplied material.

(6) The second party will return all unused material at the request of first party.

(7) The second party agrees to use the Material in compliance with all applicable statutes and regulations. The material may have hazardous properties. The providers make no representation and extend no warranties of any kind, either expressed or implied. The second party assumes all liability for claims for damages which may arise from the use, storage or disposal of the Material. However, the first party will be liable to the second party when the damage is caused by thegross negligence or willful misconduct of the first party.

This agreement is signed on_____**between:**

NIT, Rourkela	**Organization:**
(Name of signatory)	**(Name of signatory)**
Designation : Dean (SRICCE)	**Designation**
Address : National Institute of Technology Rourkela - 769008	**Address**

Form No. NITR/IP/8

National Institute of Technology Rourkela

Letter of Agreement for Developing Educational MaterialMaintained by a Private Firm

(Use PART 1 or PART 2 as the case may be)

This agreement is between National Institute of Technology, Rourkela (the provider organization) and/or provider subject matter expert(s) _____ _____ _____[name(s)], jointly called the first party, and the registrationorganisation_____ _____the second party.

PART 1 : Hiring and Licensing a firm

1. The Educational Material that is covered by the agreement includes _____ (description of the material) which is considered as proprietary material of the subject matter expert(s) provider andNIT Rourkela.

2. The provider subject matter expert and NIT Rourkela shall be free, in their sole discretion, to distribute the materials to others and to use it for their own purpose.

3. The subject matter experts undertake that utmost care has been taken to avoid any copyright infringement while developing the educational material except the amount permissible as governed by fair use.

4. The second party will be paid an amount of Rs._____for the period _____ months/years for (i) maintaining, (ii) distributing, (iii) popularizing and (iv) marketing (tick the correct choice) the educational material stated above. The second party agrees to work only in the mode of expression for the purpose assigned to it as mentioned above. Whatever may be themode of expression, the second party also agrees to ensure that it will restrict unauthorized copying of the materials.

5. If the materials are available in internet, the second party must display "Terms and Conditions" for use of the materials and "Disclaimer Policy" in regard to warranty in the home page of the website. The reliability, consistency, speed and security of the materials in internet is hereby ensured by the second party in accordance with discussion with subject matter experts (documented in the minutes of meeting no._____, dated__). However,the second party may mention its identity by describing as "maintained by _____", "marketed by_____" as the case may be. The second party further agrees to provide the feedback from the readers to the subjectmatter experts.

6. The second party is not permitted to modify any content in the supplied materials by themselves or through a third party during or after termination of this agreement. The subject matter expert or any other person authorized by NIT Rourkela has the right to modify the contents of the materialsas and when need arises.

7. The second party will return the materials to the first party after termination of this agreement.

8. The agreement can be cancelled by either party at any point of time.

PART 2 : Selling to a firm

1. The educational material that is covered by the agreement includes _____ (description of the material) which is considered as proprietary material of the subject matter expert(s) provider andNIT Rourkela.

6. The second party is not permitted to modify any content in the supplied materials by themselves or through a third party during or after termination of this agreement. The subject matter expert or any other person authorized by NIT Rourkela has the right to modify the contents of the materialsas and when need arises.

7. The second party will return the materials to the first party after termination of this agreement.

8. The agreement can be cancelled by either party at any point of time.

PART 2 : Selling to a firm

1. The educational material that is covered by the agreement includes _____ (description of the material) which is considered as proprietary material of the subject matter expert(s) provider andNIT Rourkela.

2. The second party will pay an amount of Rs._____to the first party for obtaining legal right of copyright of the material mentioned above for maintaining, distributing, popularizing and marketing in the form (i) print media, (ii) multimedia and (iii) internet. The second party agrees todeal in the mode mentioned above and not in any other mode.

3. The second party is permitted to modify the contents of the supplied materials either through thesubject matter expert or a third party on payment basis.

4. The subject matter expert and NIT Rourkela have the moral right to be recognized as the creator.

5. The subject matter experts undertake that utmost care has been taken to avoid any copyright infringement while developing the educational material except the amount permissible as governed by fair use.

This agreement is signed on_____**between:**

NIT, Rourkela

Organization:

_____

(Name of signatory)

(Name of signatory)

Designation : Dean(SRICCE)

Designation

Address : National Institute of Technology
Rourkela - 769008

Address

DR.BABASAHEB AMBEDKAR MARATHWADA UNIVERSITY,AURANGABAD

Sagar B.C.A. College, Jalna

A Project Report On

Management of Intellectual Property Policy, Procedures and Forms

Submitted To

UNDER THE GUIDANCE OF

AJITKUMAR

Submitted by

RATHOD KAILAS SANJAY
M.LiB SY IV SEM

Year 2022-23

Department of Master Of Library and Information Science

(M.Lib. SY)

CERTIFICATE

This is to certify that, the following student

Rathod Kailas Sanjay

Has successfully completed the summer internship project

Management of Intellectual Property Policy, Procedures and Forms

In the partial fulfillment of the requirement of Master Of Library and Information Science course asexpected by **Dr. Babasaheb Ambedkar Marathwada University, Aurangabad** for Academic Year **2022-23**

External Examiner

Internal Examiner

Head
Department of Humanities
Sagar BCA College, Jalna.

Principal

N. D. NAJARDHANE

PRINCIPAL
Sagar BCA College
Devmurti, Tq. Dist. Jalna

STUDENT DECLARATION

This is to declare that this Summer Training Project report on "**Management of Intellectual Property Policy, Procedures and Forms**" is a record of genuine work done by me under the guidance of AJITKUMAR in the partial fulfillment to the requirement for Master Of Library and Information Science I declare that this Summer Training project report is original and not submitted to anyother university before

Signature of the Student: K. Rathod

Student's Name: Rathod Kailas Sanjay

PRN No. 2021015200607256

Exam Seat No. CMLD401238

ACKNOWLEDGEMENT

While conducting this report, I got support in many ways from many people .First I am deeply grateful to my project guide, **AJITKUMAR** who helped me with full devotion and always supported me earnestly whenever it was needed. Without his guidance, mental &moral support and academic inputs this report was not possible.

This Training report could never have seen the light of the day without his co- operation of those Clients who participated in this. I am thankful to all of them for giving me their valuable time.

My friends have been biggest support for me at every juncture of life. They manifested their great interest in my research work also and always tried to make thingseasy for me.

A word of gratitude goes to my family members whose love; affection and understanding have enabled me to complete this end with ease.

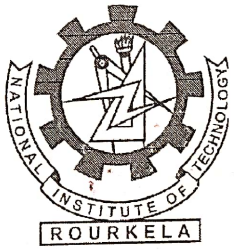
At the end, I thank to Almighty for giving me courage and strength to conduct this project report.

K. Rathod

(RATHOD KAILAS)

Management of Intellectual Property

Policy, Procedures and Forms



CONTENTS

TOPIC

Distance Education Materials
Implementation of Intellectual Property
Policy

6. Procedures for Management of Intellectual Property

Creating IP Awareness
Proposals for patent
application
Archiving theses containing Intellectual Property with commercial potential

FORMS

NITR/IP/1	Undertaking to be signed by all academic and technical staffjoining National Institute of Technology, Rourkela
NITR/IP/2	Undertaking by a person engaged by the Institute under "Workfor hire" terms
NITR/IP/3	Declaration by students (initiating work on Patentable Technologies)
NITR/IP/4	Handling and archiving of theses and dissertations submitted tothe National Institute of Technology, Rourkela
NITR/IP/5	Invention and Technology Disclosure Form
NITR/IP/5A	Invention and Technology Disclosure Form (Summary)
NITR/IP/6	Mutual Secrecy Agreement between NIT Rourkela and collaborating Institutions
NITR/IP/7	Letter of agreement between NIT Rourkela and collaboratinginstitutions for the Transfer of Proprietary Materials
NITR/IP/8	Letter of agreement for developing educational material maintained by a private firm

1. PREAMBLE

National Institute of Technology, Rourkela (hereinafter called the Institute) is dedicated to research, teaching, and extension of knowledge to the public. The Institute recognizes its responsibility to produce and disseminate knowledge. Inherent in this responsibility is the need to encourage the production of creative and scholarly works and the development of new and useful materials, devices, processes, and other intellectual property, some of which may have potential commercial value. These activities contribute to the professional development of the individuals involved, enhance the reputation of the Institute, provide additional educational opportunities to students, and promote public welfare. The Institute has a responsibility of bringing new knowledge into use by the general public. Such knowledge or technology often has commercial value and should be treated as a financial asset to be used, conserved and applied in such a way as to generate an appropriate financial return. Transfer of such information or technology through licensing satisfies both the above objectives, i.e., dissemination for use and realization of financial returns.

Technological and social developments in recent years have broadened the scope of information and technology that can have potential commercial value and, therefore, should be treated as assets subject to Institute ownership and control. In addition to new machines, compositions of matter, and written materials which traditionally have been the subject of patents and copyrights, computer software, video courses, etc. are now normal outcomes of Institute activities. Thus, a broad policy covering all aspects of intellectual property needs to be created to provide widespread protection to the originators of such property.

Over the past decade, appreciation of the commercial value of intellectual property has grown both within the academic community and in the society at large. Concerns related to confidentiality, publication, and ownership of intellectual property are now commonplace. The pace of modern science, resulting in new and useful inventions, initiated a need for a central policy in determining the course of the creation, protection, and commercialization of intellectual property in the Institute. This has resulted in establishment of the Intellectual Property Committee (IPC) and the Intellectual Property Policy (IPP) to encourage creation and protection of intellectual property in the Institute.

This IP Policy applies to all Institute employees and students, regular or contractual. Every member of the academic community, student, non-teaching and teaching staff alike, must be knowledgeable about intellectual property both to protect their own rights and to respect the rights of others. The Institute IP Policy is intended to encourage a healthy atmosphere conducive to research and development through a generous system of rewards and incentives for the creation of intellectual property while at the same time giving proper consideration to the economic rights and responsibilities of the Institute.

The strength of the Institute lies in its faculty, students, technical and administrative staff. This document is intended to introduce, regulate, and organize issues related to intellectual property within the Institute. It also reaffirms the Institute's commitment to scientific endeavors, academic excellence and the dissemination of knowledge. This policy is intended to spell out the responsibilities of the Institute and its employees and to establish a framework for ethical conduct.

2. THE INTELLECTUAL PROPERTY (IP) POLICY

The primary objective of the IP Policy of NIT Rourkela is to establish appropriate principles for creation, protection, ownership and management of intellectual property in the Institute. The endeavor is to provide an intellectual property environment that encourages the development of inventions and other intellectual creations for the best interest of the public, the creator, and the research sponsor, if any, and will permit the timely protection and disclosure of such intellectual property either by development and

commercialization after securing available protection, by publication, orboth.

This Policy is further intended to protect the respective interests of all participants by ensuring that the benefits of such property accrue to the public, to the inventor, to the Institute and to sponsors of specific research projects in varying degrees of protection, monetary return and recognition, as circumstances justify or require.

Objectives of the IP Policy :

The major objectives of the IP policy of NIT Rourkela are:

- To provide a superior environment to the employees and students of the Institute for creation, protection,and commercialization of intellectual property and to stimulate innovation.
- To encourage research, scholarship, and a spirit of inquiry, thereby generating new knowledge.

- To facilitate the transfer of knowledge and technology to intending users to promote utilization of suchresources for benefit of the society.
- To provide an administrative system to determine the commercial significance of discoveries anddevelopments and to assist in bringing these into public use.
- To provide for a equitable distribution of economic gains resulting from new intellectual property amongthe developer, author, or inventor (the originator), the Institute, and, where applicable, the sponsor.
- To provide incentives to originators in the form of personal development, professional recognition, andfinancial compensation.
- To safeguard, review and manage the intellectual property so that it may receive adequate and appropriatelegal protection against unauthorized use.
- To encourage students at all levels to develop patentable technologies and to provide financial assistancefrom the Institute to the extent possible.
- To create awareness on IPR through conducting seminars, conferences, invited talks and lectures, andtraining programs among the academic community.
- To create respect for other people's intellectual property among members of the Institute community.

SCOPE OF THE IP POLICY

The IP Policy applies to potentially patentable inventions and discoveries, industrial designs, copyrightable materials such as books, publications, electronic courseware, computer programs, electronic circuits etc., protectable trademarks and trade secrets, which are developed using Institute equipment, supplies, facilities, employee time, or trade secret information, or which relate directly to the Institute's business, research or development. The Institute will encourage, recognize and protect all creative and scholarly works in form of patents, copyrights, industrial designs, trademarks and trade secret, as the case may be, developed by its employees/students as a result of their research or employment. The Institute will protect the rights of the originator regarding intellectual property created by him as per legal framework of the land. However, the Institute, as a corporate body, shall share this right with inventors – staff, student or guest, in a just and fair manner.

This Policy applies to all the employees of the Institute and visitors using Institute facilities under the supervision of Institute staff and to all the students of the Institute including doctoral and Postdoctoral fellows.

It covers all intellectual property conceived, first reduced to practice, written, or otherwise produced by all faculty, staff, and students of the Institute whether using Institute resources or not. It also covers intellectual property created by part time students, employees and visitors using Institute funds, facilities or other resources.

3. DEFINITIONS

Intellectual Property (IP): For the purpose of this policy, "Intellectual Policy" is defined as the tangible or intangible results of research, development, teaching, or other intellectual activity. Intellectual property may include the following products:

a. Patents on new and useful scientific or technical advancements by way of inventions,

discoveries, processes, computer hardware and software, unique materials, machines, devices, instruments, apparatuses, circuits, plant varieties etc.

b. Copyright in industrial and architectural design, models, engineering drawings, integrated circuit layout designs, computer software, animations and visualizations, information technology products and processes including hardware and software features, original innovative, creative or artistic works and their derivatives or adaptations, whether dramatic, musical, literary works, work of graphics or plastics art and cinematographic and animated films, teaching material for classroom and online courses such as courseware for distance education, original data and records of research, undisclosed and/or unpublished information etc.

c. Trademarks, service marks, logos, collective marks, certification marks, trade names etc.

The three categories stated above are not mutually exclusive; a given article of intellectual property may include aspects of all three categories.

Fair use : The term "Fair Use" refers to the amount of copying or usage that may be permitted for a copyrighted material so that it does not obstruct the progress of human knowledge. Limited portions of a work can be copied without the right holder's permission for non-commercial and academic use, although the exact permissible percentage may have to be determined by the courts. In general, use of a small part of the work which does not hurt the present or potential market for that work is allowed under fair use, but there are many grey areas where the law has to be decided on a case-by-case basis. Fair use in the classroom during regular teaching is understood more liberally than that permissible in teaching for distance education through print or multimedia packages. This is because distance education packages are commercial products and hence permission has to be sought for the use of any intellectual property held by others. The possibility of fair use exists only in the case of copyright and does not apply to patents.

Employee : An "Employee" of the Institute is defined as any person receiving compensation for service, or any person volunteering services for the benefit of the Institute. The uncompensated activities of students in furtherance of their education shall not be considered service within the meaning of this policy, even if such activities benefit the Institute. A scholarship, fellowship, assistantship or any other payment received by a student during the course of his studies does not classify him as an employee.

Institute Personnel : Part-time and full-time members of the faculty, technical, administrative or the supporting staff and all other agents and employees, and undergraduate, postgraduate students, doctoral and postdoctoral fellows of the Institute.

Inventions : It is a general term which includes computer software, general instructional materials (including video tapes), novel machines, devices, compositions of matter (compounds, mixtures, genetically engineered cells, plants or animals), genetic forms, mask works, production processes, production methods, plant varieties, etc. Inventions will be considered as having been developed in the course of employment where conception and/ or development is in the individual's subject area of principal competence in scholarly activities for which the individual is employed.

Originator : Any person related to the Institute though a relationship which is in the form of a part-time or full- time member of the faculty, permanent or contractual staff, agent or employee, graduate or postgraduate student, doctoral or postdoctoral fellow of the Institute, who is involved, directly or indirectly, in the Invention as defined herein shall be deemed to be the Originator for the purpose of this document.

Work for Hire : Work for hire is defined for the purpose of this document as any work commissioned by the Institute. The ownership of the resulting intellectual property shall be assigned to the Institute through a written contract between the concerned parties.

4. IP MANAGEMENT PERSONNEL

Professor-in-charge (IP) : The Institute's Senate shall appoint a member of the faculty as professor-in-charge (IP) who will be responsible for day-to-day administration of IPR issues and shall work under the guidance of the Dean (SRICCE). He will serve as the member-secretary of the Intellectual Property Committee. The Professor-in- charge (IP) will be responsible for development and protection of intellectual properties of the Institute and find avenues for

possible commercialization. He has to ensure the use of licensed software and fair and just treatmentof others' intellectual property by the Institute and its members.

Intellectual Property Committee (IPC) : The Institute's Senate shall appoint an Intellectual Property Committee (IPC) to help administer intellectual property and to make suitable recommendations to Dean (SRICCE)/Director for implementation. Dean (SRICCE) will serve as the Chairman, and Professor-in-Charge (IP) shall serve as the Secretary of the IPC. In addition to Dean (SRICCE) and Professor-in-Charge (IP), the Senate will nominate two more members from among the faculty of the Institute. The tenure of Professor-in-Charge (IP) and the two members will be two years, preferably non-concurrent. A member may be appointed for a second term, but not more than4 (four) years in total.

The IPC will assist various departments and centres of the Institute in all matters relating to intellectual property. It will help various departments to secure protection for intellectual property where appropriate. It will maintain central databases and files of patent applications, issued patents, trademarks and copyrights, licenses and

agreements, coordinate with various departments in negotiating and preparing license and other agreements, review and approve all agreements relating to intellectual property. It will review causes of possible infringementson the Institute's intellectual property and take action as deemed necessary.

The legal interests of the Institute and its staff, faculty and students in any intellectual property, except traditional scholarly works, shall be determined by the IPC in accordance with the policy enumerated under items.

5. POLICY

General Policy

The intellectual property policy shall apply to all persons employed by the Institute – full-time and part- time faculty, visiting faculty, scientists employed by the Institute, as well as technical and administrative staff. It also applies to undergraduate, postgraduate and doctoral students as well as postdoctoral fellowsand visiting scientists.

This policy shall apply to all kinds of intellectual property (including, but not limited to, any invention, discovery, trademark, copyright, trade secret, technology, scientific or technological development, research data and computer software) regardless of whether the intellectual property is subject to protection under patent, trademark, copyright, or any other law. The institute will encourage and recognize the originator of intellectual property and protect the ownership for the creators.

The Institute will work towards protection through legal means of all creations of scholarly and educational materials, inventions, products, processes, art works, musical compositions and dramatic and non dramatic literary works related to the author's academic or professional field, regardless of the medium of expression. All such intellectual property shall be jointly owned by the originator/author and the Institute.

The Institute shall have sole ownership of all intellectual property created by an employee who was hired specifically to work on a target product or process (or other intellectual property) or was commissioned by the Institute or a component of the Institute for the specific objective leading to creation of the intellectual property. The Institute will assert its ownership of all intellectual property created by the outside agencies commissioned by the Institute for the specific purpose.

The intellectual property generated from research projects sponsored by government/ non-government agencies will be owned by the creator(s), the Principal Investigator or Chief Consultant, the Institute and the sponsoring agency. The sponsoring agency will bear 50% of the protection cost or forgo the rights to the intellectual property. In case the project was accepted by the Institute under terms different from thatstated herein, the terms agreed to shall prevail.

Intellectual Property Rights and Obligations

Intellectual property generated by a full-time employee or a full-time student of the Institute is the joint property of the originator and the Institute whether Institute resources are used or not. If an Institute employee or a full time student creates intellectual property while working in another organization, it will be jointly owned by the creator, NIT Rourkela and the host institution. In case of part-time employees or students or visiting professionals, intellectual property generated by use of Institute facilities and/or support only come under joint ownership of the Institute and the originator.

Before any intellectual property is disclosed to any outside agency, other individual, commercial or academic organization, press or public is published by the originator himself the creator shall submit a reasonably complete and detailed disclosure of such intellectual property to the IPC for determining if anykind of protection is possible under appropriate laws.

When a request from the originator is received, the IPC shall decide how, when, and where the intellectual property is to be protected. It will proceed either through its own efforts or through those of an appropriate Government or private firm or attorney to obtain protection and manage the intellectual property. Outside counsel services may be contracted with the consent of the Director. If the IPC does not find the property appropriate for protection by the Institute, the originator becomes the sole owner of the property and is at liberty to apply for protection under national laws. The Institute will issue a letter foregoing its claim to the property in future.

A decision by the Institute to seek patent or other available protection for intellectual property shall not obligate the Institute to pursue such protection internationally. The Institute's decision relating to

the geographical scope and duration of such protection shall be final. If the originator intends to seek protection internationally or through Patent Cooperation Treaty (PCT) application, the institute will permit the originator to proceed on his own or with the help of other individuals or agencies.

Royalty Income Sharing

5.3.1 In case the Institute succeeds in commercialization of intellectual property for the creator and licenses rights to third parties in consultation with the creator(s), the revenue generated through royalty payments will be equitably shared among the creators and the Institute.

Inventor(s)	Institute
60 %	40 %

In case of multiple originators of an IP, all the originators will decide among themselves how to share the proceeds of an intellectual property. If they fail to arrive at a consensus, the IPC will analyze all available information and make a recommendation to the Director. The decision of the Director shall be binding and final.

If there are other legitimate claimants to the IP, they will be grouped either under "inventors" or "Institute". While sponsoring or supporting organization will get their share of the proceeds from "Institute" share, individuals (including visiting professionals) who contributed to the invention will receive their share from that of the originators.

Ownership of Intellectual Property in Certain Circumstances

Where research has been sponsored by a private industry/ foundation or government agency and no prior agreement exists on sharing of intellectual property, licensing of patents shall be negotiated between the sponsor and the Institute.

The intellectual property policies and guidelines of the Institute are subject to, and thus amended and superseded by the specific terms pertaining to intellectual property rights included in Central or State grants and contracts, or grants and contracts with NGO's or private sponsors.

If the intellectual property has been generated as a work-for-hire, the employee or agency will retain the moral right to be identified as the creator of the intellectual property but right of commercialization rests only with the Institute.

Ownership of Intellectual property Generated by students

It is a requirement in academics that a student must own the copyright of the thesis (since it is his or her original work) which he or she submits as partial fulfillment of the requirements for an academic degree. However, the student will grant a non-exclusive, non-transferable royalty-free license to the institute to use, in the course of non-commercial academic activity, the records and data generated in the course of his research. Furthermore, it is possible that the research that the student carries out as part of the program of study may result in the generation of intellectual property other than the text of the thesis. Supervisors should advise students during the course of their work that certain kind of research may lead to the generation of intellectual property which will require protection of its commercial value through confidentiality, for which the student will have to forgo publication during the period of sealing of a patent. Care should be taken at all stages to see that no conflict of interest arises between

the student's academic activities and his or her generation of intellectual property. The copyright of the thesis in whichthis intellectual property is described or outlined will remain with the student while the institute will restrict access to the thesis for a limited period depending on commercial value as decided by the IPC. The institute will try to obtain a patent for the invention on behalf of the student and benefit-sharing mechanism will be abided by as proposed in section 5.3.1.

If a student is employed to assist in execution of a sponsored project or programme, the intellectual property rights originating from his contribution to the project will be governed by the terms of the contract between the institute and the sponsoring agency.

If the intellectual property has been generated as a work-for-hire, the student will retain the moral right to be identified as the creator of the intellectual property, but right of commercialization rests with the Institute.

Disclosure and Confidentiality.

At an appropriate stage in the development of an invention, the originator shall make a written disclosure of the concepts to the IPC, providing all such particulars as are vital to judge its commercial prospects. The IPC shall promptly acknowledge, in writing, its receipt of the disclosure and the date of receipt. The originator shall send one copy of his proposed manuscript, prior to submission of thesis, to the IPC.

All the departments in the Institute will be bound by the non-disclosure and confidentiality terms to be clearly spelled in a separate document. Each department is under obligation to file their R&D manuscripts, ifany, on time to time basis, with the IPC. It is expressly understood by the departments that any information which relates to any Invention should be treated as Intellectual Property and therefore is not to be divulgedwithout the prior consent of the IPC.

The Originator who has communicated with the IPC under Clause 5.6.1 shall refrain from publishing, reading, dissipating, circulating or disclosing the conception in any form whatsoever, since non-disclosure is one of the most important qualifications for intellectual property protection. The originator may disclose such conception, upon a prior written permission from the IPC, once an application for a patent, trademark or copyright has been made on the conditions described herein and the commercial rights in the conception are secured to the Institute.

Commercialization.

For purposes of protection and commercialization of intellectual property on behalf of the Institute, patent, trademark or copyright coverage may be sought, or the property may be treated as proprietary information, technical know-how, or trade secret.

The IPC may determine whether the Institute has a legal interest in the commercialization of the property.However, the Institute is not legally bound to commercialization of each property and the originator maynot claim such right. It shall be in the sole discretion of the Director on advice of the IPC to determine commercialization of the property.

In seeking and developing commercialization of intellectual property, the Institute shall be guided by thefollowing principles:

(a) A primary objective and responsibility of the Institute shall be to assure that the products of itsintellectual activity are brought into the widest possible use for the general benefit of society.

(b) Intellectual property should be treated as an asset and an appropriate return should be sought.

(c) Active participation of the originator in all commercialization efforts shall be sought.

Electronic and Distance Education Materials

All original works submitted by the contributors for the purpose of electronic and/or distance education coursedevelopment shall remain the property of the concerned contributors and the Institute.

The Institute has the right to use the course (including all related materials) developed by the faculty member and/or other employees involved in the development of a distance learning and/or e-learning course for the Institute's own educational, research, and other purposes without any additional compensation to the faculty member or any other employee who is an author of the course. Academic departments determine which courses will be offered and who will teach these courses.

If the course and/or related materials are licensed, sold, or otherwise conveyed to a third party, the mechanism as per Clause 5.3.1 will govern the distribution of any proceeds. If the material is used in distance education activity generating revenue under a scheme where instructors or other Institute personnel receive specific compensation, the originator of the courseware is entitled to a fair share of the proceeds, the exact amount being determined by the Director.

As to revisions and updates of a distance learning course, the faculty member(s) or other employee(s) who are the authors of the course will have the primary responsibility for revising and updating the course and related materials as long as he/she/they are employed by the Institute. If a faculty member (or other employee) who is an author is no longer employed by the Institute or is unable to revise and update the course as needed (as determined by the department), the material may be revised and updated by other Institute faculty members and other employees.

While developing the instruction materials in educational multimedia, digital imaging, and distance learning mode, faculty members are advised to avoid copyright infringement.

Implementation of Intellectual Property Policies.

The IPC shall prepare and distribute to the various departments copies of this document and other recommendations as may be considered appropriate for the implementation of the provisions of intellectual property policies and guidelines adopted by the Institute.

The policies set forth herein constitute an understanding which is binding on Institute faculty, staff, and students as a condition of their participation in Institute research, teaching, and service programmes.

6. Procedures for Management of Intellectual Property

Creating IP Awareness

The IPC will take a proactive approach for generation and protection of intellectual property in the Institute. The committee shall identify all intellectual property developed in the Institute through review of UG, PG and doctoral projects, inviting ideas from students and faculty, and by all other means of collecting information. If it judges that there is a reasonable chance for successful commercialization of an invention, it shall (1) advise the originator on the steps to be taken for protecting the ownership rights to the property, and (2) refer the matter to the Director with clear recommendations for appropriate course of action on the part of the Institute. On persuation of the IPC, or on his own initiative, the originator(s) will place a formal proposal before the IPC with details of the invention or other IP product. The IPC will examine the proposal taking help of internal or external experts, if necessary. The originator(s) may be required to make a demonstration of the product or process, or give a presentation before the Committee.

In some instances the IPC may find that an invention has not been developed to an extent where a decision can be made on patentability or commercialization. In such cases, it will request the originator to provide additional information or data that might help in making a decision, or advise the originator to report back to the IPC after the discovery is brought to a more advanced stage.

In close consultation and collaboration with the originator, the IPC shall determine the appropriate method of protection of the property and, where appropriate, obtain such protection. The process for licensing, selling, or otherwise conveying intellectual property will not involve the use of sealed bids. All costs associated with these actions shall be borne by the Institute, except that such costs shall be offset against future income. When a technology or other invention or intellectual property is commercialized, the net income from such commercialization will be distributed among the originator(s), the Institute and other stake holders. The IPC will guide the Institute on the distribution formula.

- IPC in coordination with Deputy/Assistant Registrar(Admin) will ensure that every faculty

member/technicalstaff signs the undertaking form no. NITR/IP/1 at the time of joining.

- At the time of registration for Autumn Semester, each student must sign and submit a declaration in formno. NITR/IP/3 to Prof-In-Charge (IP).
- All departments will provide financial support in a fruitful means to all student projects with possible commercisable outcome.
- The institute will provide additional funds with higher outlay for project works with a commercial potential, wherever necessary.
- The IPC will bring out brochures/newsletters and make necessary announcements in various media for creating awareness among academic community in regards to advancements on technology, patentable and commercial technologies, IP laws and amendments and legal aspects related to IP.

- The IPC will conduct seminars, workshops, talks by eminent scientists and practitioners on IP related issues,and shall organize student groups to create greater participation.
- At the time of submission of thesis, each student must sign and submit a declaration in form no. NITR/IP/4 to Prof-In-Charge (IP).

Proposals for patent application

Faculty members, technical staff and students interested to protect their intellectual creations under IP law of the land may apply to Professor-in-charge (IP) using the Invention and Technology Disclosure Form(Form No. NITR/IP/5).

Professor-in-charge(IP) will draw the attention of IPC members in a meeting for evaluating the IP substance for possible protection within two weeks. If the members agree to file for protection, the IPC will approach appropriate Government, private and legal entities to go forward with protection of the IP with due recommendation from Director.

The expenditures for protection of IP substance will be borne by the Institute from its non-plan "administrative expenses" head.

If the IP substance is not fully developed for possible protection, IPC will guide the originators whereto improve it. IPC may also give guidance on drafting the Patent forms etc. even with provisional specifications.

The Institute shall bear all the charges for patent search while filling up the patent form.

If the patent is granted, it becomes the joint property of the originator and the Institute.

The Institute has the prerogative of finding a suitable partner for commercialization of the patents for firsttwo years from the date of grant of the patent.

After two years, the originator may choose a suitable partner for commercialization of intellectual propertycreated by him/her. However, benefit sharing mechanism will be adhered to as per Clause no. 5.3.1. The originator, before going for technology transfer on his own, must seek the permission of IPC. The IPC should strive to dispose off the matter within two weeks.

Archiving theses containing Intellectual Property with commercial potential

Theses submitted by the students may have potential IP substance. It should be protected and commercialized for greater interest of humanity. Therefore, the students must come forward and the supervisors must motivate the students to work on patentable technologies. The student submitting a thesis must give a declaration in the Form no. NITR/IP/4.

The IPC will examine all the theses and find the suitability of protection of IP created by the students.

The Institute will also try for potential commercialization when a patent is granted. However, the Institutehas the prerogative of commercializing only for the first two years from the date of grant of patent. The revenue sharing mechanism will be adopted as per Clause 5.3.1.

If the Institute does not show any interest for patenting an invention contained in a thesis, the student may go for protection with formal information to the Institute. In such case, the thesis is protected and not disclosed to anyone for a period of one year after such a request is received

Form No. NITR/IP/1

National Institute of Technology
Rourkela

Undertaking to be signed by all academic and technical staff joiningNational Institute of Technology, Rourkela

1. This is to declare that I have read and understood the policy of the National Institute of Technology, Rourkela with respect to intellectual property and the rights therein, titled [`Policy`] and that I agree to be bound by it and to follow its provisions during the period of my employmentby the Institute.

2. I agree to report, disclosing full details, to the relevant authorities of the Institute any patentable or commercializable intellectual property that I may generate or participate in generating in accordance with the provisions of the Intellectual Property Policy.

3. (Strike out whichever is not applicable)

   (a) I certify that I am at present under no contractual obligation with any person ororganization, which are in conflict with the Policy.

   or

   (b) I am at present under the contractual obligations detailed below:

   _____

   _____

   _____

   (attach a separate sheet, if necessary)

4. I agree to share all intellectual property generated during the course of my work with the Institutein accordance with the Intellectual Property Policy of the Institute in vogue at the time of creationof the intellectual property.

5. I undertake to behave with dignity and broadness of mind while sharing intellectual property rights with my coworkers – students, faculty, technicians and other supporting staff as well as visitors.

Name :

Designation :

Department :

Employment Code :

(Signature)

Date

Form No. NITR/IP/2

National Institute of Technology
Rourkela

Undertaking by a person engaged by the Institute under "Work for hire" terms

I, _____ hereby certify that the work: _____ (the "Work"; attach additional sheet if necessary to accurately describe the work) is specially commissioned by National Institute of Technology, Rourkela and is to be considered a "work for hire".

I undertake that during the course of my work, I shall not use any material protected under copyright lawsbeyond the scope of fair use except those for which explicit permission of the owner has been obtained. Further, I am not entitled to reproduce elsewhere any potential copyrightable material generated during the period of this agreement or after its termination except that which falls under fair use. I shall retain only moral rights to this material. Furthermore, no patentable invention/technology/innovation/trademarks developed by myself, and others I shall be working with, will be disclosed by me to any other party upon termination of this agreement. I understand that any prior disclosure by myself, directly or indirectly, either during the period of this work-for-hire agreement or after its termination, shall render me prosecutable as per laws that may be in force at the time.

I hereby assign and/or transfer to Institute, absolutely and forever, all rights related to intellectual property generated during commissioning of the work or after its termination.

Signed this_____day of_____(month),_____(year)
Name:

Address:

Work-for-hire agreement tenable at Department:

In the project:

Signature

Date

Form No. NITR/IP/3

National Institute of Technology
Rourkela

Initiating work on Patentable Technologies

Declaration by the Student(s) engaged in Project / Dissertation works

I/We, Mr./Ms./Dr. _____ Roll No.: _____ and Mr./Ms./Dr._____Roll No.:_____registered as research scholar or student of programs such as B.Tech./M.Tech./ M.Tech.(Res) / M.Sc. / Ph.D / D.Sc. in the Department of____, National Institute of Technology, Rourkela (hereinafter referred to as the "Institute") do hereby declare that I/we shall start working on a project entitled :_____

_____.

I/We declare that :

1. I/We shall respect the intellectual property of others. I/We shall not knowingly or unknowingly use any protected inventions / designs / materials / integrated circuits held by others for our purpose without paying the license fee.

2. I/We shall not use any unlicensed modeling / drafting / word processing /programming software for my/our purpose.

3. I/We shall not use any material protected under copyright law except beyond thescope of fair use for our purpose.

4. I/We shall request IPC for helping us for patent search for our work, the cost if any being borne by the Institute.

5. In the course of project work, if any IP is generated I/we shall proceed as the IP policy of the Institute for possible protection and subsequent commercialization.

Name of the Student

Signature of student
with date

Name of the Student

Signature of student
with date

Name of the Supervisor

Signature of the Supervisor with date

Signature of the Head of the Department with date

Form No. NITR/IP/4

National Institute of Technology Rourkela

Handling and Archiving of Theses and Dissertations submitted to theNational Institute of Technology, Rourkela

Declaration by the Author of the Thesis or Dissertation

I, Mr./Mrs./Miss/Dr._____Roll no._____ registered as a research Scholar or a student of programs such as B.Tech./ M.Tech./ M.Tech(Res)/ M.Sc./Ph.D/D.Sc in the Department of _____ National Institute of Technology, Rourkela (hereinafter referred to as the 'Institute') do hereby submit my thesis, entitled: _____ _____ (herein referred to as 'my thesis') in printed as well as in electronic forms for holding in the library of records of theInstitute.

I hereby declare that:

1. The electronic version of my thesis submitted herewith on CDROM is in PDF format.
2. My thesis is my original work of which the copyright vests in me and my thesis does not infringe or violate the rights of anyone else.
3. The contents of the electronic version of my thesis submitted herewith are the same as those submitted as final hard copy of my thesis after my viva voce and adjudication of my thesis on____(date).
4. I agree to abide by the terms and conditions of the Institute Policy on Intellectual Property (hereinafter Policy) currently in effect, as approved by the competent authority of the Institute.
5. I agree to allow the Institute to make available the abstract of my _____.

thesis to any user in both hard copy (printed) and electronic forms.

6. For the Institute's own, non-commercial, academic use I grant to the Institute thenon-exclusive license to make limited copies of my thesis in whole or in part andto loan such copies at the Institute's discretion to academic persons and bodies approved from time to time by the Institute for non-commercial academic use. All usage under this clause will be governed by the relevant fair use provisions in the Policy and by the Indian Copyright Act in force at the time of submission of the thesis.

7. I agree to allow the Institute to place such copies of the electronic version of my thesis on the private intranet maintained by the Institute for its own academic community.

8. I agree to allow the Institute to publish such copies of the electronic version of my thesis on a public access website of the internet.

9. If in the opinion of the Institute my thesis contains patentable or copyrightable material and if the Institute decides to proceed with the process of securing copyrights and/or patents, I expressly authorize the Institute to do so. I also undertake not to disclose any of the patentable intellectual properties before being permitted by the Institute to do so, or for a period of one year from the date of final thesis examination, whichever is earlier.

10. In accordance with the Intellectual Property Policy of the Institute, I accept that any commercialisable intellectual property contained in my thesis is the joint property of myself, my coworkers, my supervisors and the Institute. I authorize the Institute to proceed with protection of the intellectual property rights in accordance with prevailing laws. I agree to abide by the provisions of the Institute Intellectual Property Right Policy to facilitate protection of the intellectual property contained in my thesis.

11. If I intend to file a patent based on my thesis when the Institute does not wish so, I shall notify my intention to the Institute. In such case, my thesis should be marked as patentable intellectual property, and access to my thesis is restricted. No part of my thesis should be disclosed by the Institute to any person(s) without my written authorization for one year after my informing to the Institute to protect the IP on my own, within 2 years after the date of submission of the thesis or the period necessary for sealing the patent, whichever is earliest.

Name of student: **Name of supervisor(s):**

Signature of student: **Signature of supervisor(s):**

Signature of the Head of the Department

Form No. NITR/IP/5

National Institute of Technology Rourkela

Invention and Technology Disclosure Form

1. Title of the invention:

For Office use only

Proposal ID : _____

NITR/IP/____ - ____ / _____ Date of Receipt : _____

Acad yr Sl. No.

2. Inventors:

[For visiting scientists, please give details of substantive employer.]

SlNo.	Name	Employee Codeor Roll No	Position	Department	Email

3. **Brief description of the invention:** *(How this invention relates to new processes, systems, machines, compositions of matter etc.)*

4. **Detailed description of the invention**

State of prior art

(a) Prevailing **state of the art?**

(b) **Literature search** relating to this invention? *[Please include a copy of the resulting documentation, and reprints of publications.]*

(c) **Prior art/patent search** relating to this invention? *[Please include a copy of the resulting documentation, and reprints of patent documents: if a computer database search has been resorted to, please give the web site details and the Key Words used in the search.]*

Description : *(Describe the invention so that other Institute faculty who are knowledgeable in the field can evaluate its technical and commercial merits.)*

5. Funding and support

I. Was there significant use of Institute equipment and facilities? **Yes/No**

II. Was the invention supported by research grants/contract from external sources? **Yes/No**

If **YES,** please give details :

(a) Sponsor : _____

(b) Grant/contract no. : _____

(c) Period of grant/contract : _____

(d) Principal investigator and co-investigator: _____
(even if they are not inventors within the purview of this document and will not share the credit androyalties)

(e) Has the sponsor been informed of the invention? *(state whether required under grant/contractaward conditions)*

(f) Was the work done under any other agreement? Give details.

6. Information for protection of IPR: conception and disclosure *(Accurate data is requïred as priordisclosure may affect possibility of obtaining patent rights.)*

	Date	References/comments
Date of conception of this invention.Has this date been documented? If so, where and how?		
Has this invention been presented at seminars/discussions other than those which form the requirement for the degree program of the student?		
Please provide the anticipated date of submission for publication or communication for presentation at seminar/conference etc. (Should not be earlier than one month from thisdate)		
Has the invention been reduced to practice?		

7. Commercial potential

Possible uses or application areas or products that may embody some aspects of the technology :

List of probable users of the technology (class of industries/organizations or target companies) :

List of probable organizations who may be interested in technology transfer (target industries orcompanies or other oraganisation) :

Potential marketability including commercial suggestions [viable size of industry, equipment, rawmaterial and manpower requirement under different skill levels, import component, exportpotential, other relevant economic information]

8. Prior disclosure and possible intent :

Has the invention been disclosed to industry representatives or their parties?

Has any commercial organization shown interest in this invention? Give details.

9. Development Stage :

What is the current stage of development of the invention as it relates to commercial utilization and marketability :

- ☐ Embryonic
- ☐ Partially developedFully developed

10. Potential for international patent :

Does the invention have significant commercial potential in foreign countries?If so, where? Give details.

11. Declaration :

I/We declare that all statements made herein are true to the best of my/our knowledge. I/We hereby agree to hold the right of intellectual property of this invention jointly with National Institute of Technology, Rourkela. National Institute of Technology, Rourkela will share any royalty income derived from the invention with the inventor(s) according to the IP policy of the Institute in force. Intellectual Property of this invention will be protected by National Institute of Technology, Rourkela from time to time based on its merit and commercial viability.

SlNo.	Name	Signature	Date	Place

1.				
2.				
3.				
4.				
5.				

Note:

(1) A patent confers the right upon an inventor to commercially exploit an invention for a limited period of time. Patent can be lost by disclosure of the details of an invention to the public before the filling of a patent. Unlike copyright, patent is not an automatic right. To obtain a patent, the proposed invention should be novel (not published elsewhere), inventive (not obvious to persons familiar with the state of art) and industrially applicable (should have utility). Once the patent is sealed, the patentee can sue for damages anyone who attempts to exploit the patented invention without the consent of the patentee.

(2) This document should be prepared with due care. The formal patent application will be prepared only from the information provided herein.

(3) The completed disclosure form with annexures should be submitted to :

Professor-in-Charge (IP)
National Institute of Technology, Rourkela

Form No. NITR/IP/5A

National Institute of Technology
Rourkela

Invention and Technology Disclosure Form
(Summary)

1. **Title of the invention:**

2. **Inventors:**

[For visiting scientists, please give details of substantive employer.]

Sl No.	Name	Employee Code or Roll No	Position	Department	Email

3. **Brief description of the invention:** *(Not to exceed 100 words)*

4. **Prevailing state of the art :**
   (a) Details of Patent search sites or other resources.

   (b) Key words used for patent search.

   (c) List of patents related to present invention.

5. **Novelty**

6. **Inventiveness**

7. **Advantages**

8. **Commercial Potential :** *(List of organization with possible interest in the invention)*

9. **Signature of Inventor(s) with date**

Form No. NITR/IP/6

National Institute of Technology Rourkela

Mutual Secrecy Agreement between NIT Rourkela and collaborating Institutions

This agreement is between National Institute of Technology, Rourkela (the provider organization) and/or provider scientist(s),jointly called the first party, and the organization_____ _____(the recipient organization), the second party.

1. The first party is the owner of the invention called _____ (description) and/or owner of certain technical data/process technology/other information (proprietary information)developed through their own efforts.
   The organization_____(name), the second party, is the owner ofthe invention called_____(description) and /or certain technical data/process technology/other information (proprietary information) developed through its own business and R & D efforts.

2. The invention and/ or the proprietary information of the provider scientist(s), NIT Rourkela and _____ (organization) is proprietary and confidential and not public knowledge. This will be disclosed to one another under the terms of this agreement.

3. The parties to this agreement consider it desirable for each other to have access to above invention/proprietary information for discussing and evaluating possible collaborative research and development work and/or licensing activities relating thereto.

Therefore the parties agree as follows:

- All invention/proprietary information as used in this Agreement provided by one party is proprietary and confidential in connection with evaluation of invention and/or proprietary information for collaborative R & D and/or licensing work. These are disclosed to one another inwriting clearly marked confidential OR arise out of discussions during visits to laboratory/plants or any other facility of either party, and reduced to writing within thirty (30) days of such discussion.The date and time of the visit and personnel present during the visit should be recorded in writingby both parties.

- All parties agree to hold in confidence any or all invention/proprietary information disclosed andfurther agree not to disclose the same to third parties or use it for any other purpose other thandiscussion and internal evaluation provided in this document. However,

either party may disclose the invention/information/technical data/technology to its own employees assisting that party inmaking an evaluation, provided that all such employees shall have agreed to be bound by the secrecy terms of this agreement.

- The recipient of tangible products or materials consisting invention/technology from the other party agrees not to analyze or have a third party to analyze such tangible products or materials.
- All invention/proprietary information is and remains the property of the disclosing party and must be returned, in a form suitable to be returned, within ninety (90) days after the disclosing party makes a written request for its return or at the conclusion of evaluation or termination of the Agreement.
- The evaluation period during which information will be exchanged will be ordinarily one (1) yearfrom the date of signing this Agreement unless extended by mutual consent of the parties in writing.
- The foregoing obligation with respect to invention/proprietary information received by any partywho are signatories to this Agreement shall survive in the event of termination of this agreement.

This agreement is effective as of_____, and shall terminate on _____. The two parties can extend the agreement through mutual consent, in writing, andthe extension period shall be on a yearly basis. Either party may terminate this Agreement at its discretion immediately upon written notice to the other party.

This agreement is signed on_____between:
NIT, Rourkela

Organization:

_____

(Name of signatory)

Designation : Dean (SRICCE)

Address : National Institute of Technology
Rourkela - 769008

(Name of signatory)

Designation

Address

Form No. NITR/IP/7

National Institute of Technology Rourkela

Letter of Agreement between NIT Rourkela and collaborating institutionsfor Transfer of Proprietary Materials

This agreement is between National Institute of Technology, Rourkela (the provider organization) and/or provider scientist(s) _____ _____ [name(s)], jointly called the first party, and the recipient scientist(s) _____ [name(s)] of organization_____ (the recipient organization), the second party.

The Material that is covered by the agreement includes _____ (description of the material) which is considered as proprietary material of the provider and NIT Rourkela. The provider scientist and NIT Rourkela shall be free, in their sole discretion, to distribute the Materials to others and to use it for their own purpose.

In response to the second party's request for Material, both the parties agree to the following before second party receives the Material :

(1) The Material shall be used by_____(recipient scientist) working at _____ (recipient organization) in research to study _____ (description of work). The material will be used for teaching and not-for-profit research purpose only and not for use in any product or process for profit-making commercial purpose. The material is provided at no cost or with a fee ofRs._____.

(2) Neither the Material nor this material treated by any means will be used in human subjects.

(3) The second party shall not distribute, release or disclose the Material to any person or entity other than laboratory personnel under recipient scientist's direct supervision, and the second party must undertake to ensure that no one will be allowed to take or send Material to any other location unless written permission is obtained from the first party.

(4) The second party is not allowed to have a third party analyze such tangible products or materials obtained from the first party without written and specific authorization from the first party.

(5) The second party will give a brief description of its research program and the nature of usage

of the material to the first party. The second party will acknowledge the first party before any publication or presentation based on research results with supplied material.

(6) The second party will return all unused material at the request of first party.

(7) The second party agrees to use the Material in compliance with all applicable statutes and regulations. The material may have hazardous properties. The providers make no representation and extend no warranties of any kind, either expressed or implied. The second party assumes all liability for claims for damages which may arise from the use. storage or disposal of the Material. However, the first party will be liable to the second party when the damage is caused by thegross negligence or willful misconduct of the first party.

This agreement is signed on _____ between:

NIT, Rourkela	**Organization:**
(Name of signatory)	**(Name of signatory)**
Designation : Dean (SRICCE)	**Designation**
Address : National Institute of Technology Rourkela - 769008	**Address**

Form No. NITR/IP/8

National Institute of Technology Rourkela

Letter of Agreement for Developing Educational Material Maintained by a Private Firm

(Use PART 1 or PART 2 as the case may be)

This agreement is between National Institute of Technology, Rourkela (the provider organization) and/or provider subject matter expert(s) _____ _____ _____[name(s)], jointly called the first party, and the reorganisation_____ _____the second party.

PART 1 : Hiring and Licensing a firm

1. The Educational Material that is covered by the agreement includes _____ (description of the material) which is considered as proprietary material of the subject matter expert(s) provider andNIT Rourkela.

2. The provider subject matter expert and NIT Rourkela shall be free, in their sole discretion, to distribute the materials to others and to use it for their own purpose.

3. The subject matter experts undertake that utmost care has been taken to avoid any copyright infringement while developing the educational material except the amount permissible as governed by fair use.

4. The second party will be paid an amount of Rs._____for the period _____ months/years for (i) maintaining, (ii) distributing, (iii) popularizing and (iv) marketing (tick the correct choice) the educational material stated above. The second party agrees to work only in the mode of expression for the purpose assigned to it as mentioned above. Whatever may be themode of expression, the second party also agrees to ensure that it will restrict unauthorized copying of the materials.

5. If the materials are available in internet, the second party must display "Terms and Conditions" for use of the materials and "Disclaimer Policy" in regard to warranty in the home page of the website. The reliability, consistency, speed and security of the materials in internet is hereby ensured by the second party in accordance with discussion with subject matter experts (documented in the minutes of meeting no._____, dated__). However,the second party may mention its identity by describing as "maintained by _____", "marketed by_____" as the case may be. The second party further agrees to provide the feedback from the readers to the subjectmatter experts.

6. The second party is not permitted to modify any content in the supplied materials by themselves or through a third party during or after termination of this agreement. The subject matter expert or any other person authorized by NIT Rourkela has the right to modify the contents of the materialsas and when need arises.

7. The second party will return the materials to the first party after termination of this agreement.

8. The agreement can be cancelled by either party at any point of time.

PART 2 : Selling to a firm

1. The educational material that is covered by the agreement includes _____ (description of the material) which is considered as proprietary material of the subject matter expert(s) provider andNIT Rourkela.

2. The second party will pay an amount of Rs._____to the first party for obtaining legal right of copyright of the material mentioned above for maintaining, distributing, popularizing and marketing in the form (i) print media, (ii) multimedia and (iii) internet. The second party agrees todeal in the mode mentioned above and not in any other mode.

3. The second party is permitted to modify the contents of the supplied materials either through thesubject matter expert or a third party on payment basis.

4. The subject matter expert and NIT Rourkela have the moral right to be recognized as the creator.

5. The subject matter experts undertake that utmost care has been taken to avoid any copyright infringement while developing the educational material except the amount permissible as governed by fair use.

This agreement is signed on_____**between:**

NIT, Rourkela

(Name of signatory)

Designation : Dean(SRICCE)

Address : National Institute of Technology
Rourkela - 769008

Organization:

(Name of signatory)

Designation

Address