Improving Results for the INEX 2009 and 2010 Relevant in Context Tasks

A thesis

submitted to the faculty of the graduate school

of the University of Minnesota

by

Reena Rachel Narendravarapu

In partial fulfillment of the requirements

for the degree of

Master of Science

Dr. Donald B. Crouch

August, 2011

Acknowledgements

I would like to take this opportunity to express my gratitude to all those who have helped and supported me during the course of my thesis.

First and foremost, I would like to thank God Almighty for his abundant grace and guidance in every step of my life.

I would like to thank Dr. Carolyn Crouch and Dr. Donald Crouch for their constant support and guidance throughout these two years.

I am thankful to the Computer Science Department faculty at University of Minnesota Duluth including Dr. Ted Pedersen, Dr. Pete Willamsen, Dr. Hudson Turner and Dr. Chris Prince who imparted wonderful Computer Science knowledge during my MS. A special note of thanks to my Math professor, Dr. Joe Gallian who is one of the best professors I have ever worked with. I would also like to thank Lori Lucia, Clare Ford and Jim Luttinen for their continuous help.

I thank all my friends, seniors and my colleagues, Natasha Acquilla, Bhagyashri Mahule and Radhika Banhatti for their commitment and support throughout the work. I also thank Supraja Nagalla and Sai Subramanyam Chittilla for processing the document collection.

I am very thankful to my family for their unconditional love and support in all my endeavors without which this would not have been possible.

Abstract

Information Retrieval Systems focus on retrieving information relevant to the user's query. Many strategies have been developed to retrieve documents. Due to the increase in the data across the web, it is very important to retrieve relevant elements at the appropriate level of granularity. Our element retrieval system called Flexible retrieval (Flex) works with semi-structured documents to retrieve elements at run time.

The goal of this thesis is to improve the results of INEX Ad Hoc 2009 and 2010 Relevant in Context (RiC) tasks. The RiC task returns a set of focused elements that are ordered by document. Snippets are incorporated as a part of the 2010 INEX RiC task. Snippets give an overview of the document and hence should be short with few irrelevant characters so as not to lose user interest. Appropriate retrieval techniques are developed to accommodate snippets. The Restricted Relevant in Context (RRiC) task, in which a snippet can have a maximum length of 500 characters, is also described.

Table of Contents

List of Figures

List of Tables

1. Introduction

Information retrieval (IR) is a field of computing which deals primarily with searching large collection of documents and retrieving documents based on their similarity to the query. Much current research in this field is due to the vast amount of data available on the World Wide Web (WWW). IR focuses on techniques used to search and retrieve information from documents and provide the user with the most relevant information with respect to their query. Extensible markup language (XML) provides a way to represent documents on the web so that storage and retrieval can be effectively performed. Most textual data on the web is represented in XML.

This research describes and evaluates algorithms to facilitate such retrieval. It is performed as a part of the INEX (Initiative for the Evaluation of XML Retrieval) Ad Hoc track. The goal of INEX is to enhance the state of XML retrieval. INEX provides XML document and query collections for this purpose and acts as a uniform platform for the evaluation of results submitted by all the participants across the globe. The current INEX collection is semi-structured, which means that the documents do not strictly follow a Document Type Definition (DTD). Our research group at University of Minnesota Duluth participates in the INEX Ad Hoc track and focuses on the strategies to retrieve elements from semi-structured documents. Descriptions of the document collection, query collection, evaluation metrics and the tasks for 2009 and 2010 INEX Ad Hoc track are found in Chapter 2.

The focus of XML retrieval is to reduce the granularity of search results from the entire document to the element level. The Smart retrieval system [11] is used for basic retrieval functions. Smart is based on the Vector Space Model [12]. The Vector Space Model represents documents and queries as vectors. The vectors are maintained in term frequency form. The similarity between vectors is calculated using an appropriate similarity function producing a ranked output. More detail is provided in Chapter 2.

Our system for flexible retrieval, which we call Flex, is used to retrieve rank-ordered lists of elements from a pre-selected set of documents. These results are identical to those produced by retrieval against an all-element index of the collection. Dynamic or flexible retrieval enables us to retrieve elements at run time, i.e., it is dynamic in nature. It is efficient with respect to space and cost. Flex considers the paragraph as the basic unit of retrieval. A more detailed description of Flex is provided in [5].

The objective of this research is to improve the results of the Relevant in Context (RiC) task as described in the INEX 2009 and 2010 Ad Hoc tracks [9] [1]. The RiC task is basically designed to produce a rank ordered list of non-overlapping elements that are grouped by document. In 2010 Ad Hoc track, it focuses on *snippet* retrieval, which provides a short summary of the search results (to accommodate the resource restricted conditions such as small screen mobile device). The strategies used for this task are discussed in Chapter 3. Chapter 4 describes the experiments performed for the Relevant in Context tasks and analyzes the results of these experiments. Conclusions and future work are presented in Chapter 5.

2. Overview

This chapter gives an overview of INEX and its tracks, with emphasis on 2009 and 2010 Ad-Hoc Tracks. The Vector Space Model [12], Smart retrieval system [11] and the core of the retrieval system, are explained briefly. The relevance assessments and the evaluation measures used to assess the performance of the system are also explained.

2.1 Retrieval Engine

The retrieval engine is the building block of the retrieval process. The XML documents are retrieved based on Vector Space Model (VSM) [12]. Smart is used for basic retrieval functions. The Vector Space Model and the Smart information retrieval system are described below.

2.1.1 Vector Space Model (VSM)

The basic retrieval model used in this research is the Vector Space Model [12]. It is an algebraic model which represents the text document as a vector with the words within the text as its components. In this model, each document in the document collection and each query from the query set are represented as n-dimensional vectors whose components are the word types (unique terms) occurring in them. These terms are weighted based on their frequency in the document. Similarity between the document and query is measured by calculating the inner product between the vectors. This provides a rank-ordered set of vectors.

2.1.2 Smart Retrieval Engine

Smart [11] is a retrieval system based on the Vector Space Model [12]. It has an indexing component which is responsible for constructing vectors for each document and query. The *nnn* or *term frequency* vectors used during indexing do not take the length of the document or query into account. Hence, these vectors are reweighted using Singhal's *Lnu-ltu* weighting scheme [13] which ensures that the length disparities between the

vectors don't bias results unfairly in favor of longer documents. The document vectors are converted to *Lnu*-weighted vectors and the query vectors are converted to *ltu* vectors. To reweight the document and query vectors, collection-specific parameters, namely, slope and pivot, are calculated as described in [3]. The pivot is calculated as the average number of unique terms in a document collection. The slope value is estimated experimentally. Slope value is selected by using the value which produces the best MAiP when evaluated using the Thorough metric [3]. Smart retrieval produces a ranked list of documents that closely correlate with the query. We use Smart for article retrieval.

2.2 INEX

The ***IN****itiative for the* ***E****valuation of* ***X****ML-Retrieval* (INEX) [6] was founded in 2002. It provides a platform for development and evaluation of information retrieval algorithms for XML documents. It is a competition in the field of Information Retrieval, and various corporations and universities participate in it. All the participants develop their own retrieval strategies and INEX evaluates all these strategies using various evaluation techniques.

INEX has different tracks such ad Ad-Hoc, Book, Efficiency, Snippet Retrieval, Data-Centric, XML Mining, Link-the-Wiki, etc. The research group at University of Minnesota, Duluth participates in the Ad-Hoc track of INEX. The Ad-Hoc track is centered on the Focused and RiC tasks.

INEX provides a large document collection, queries and uniform evaluation metrics. All the participants use this collection and query set in testing the design and implementation of their algorithms. The evaluation measures provided by INEX serve as a common measure to compare the performance of participants.

2.3 Document Collection

Over the past years, the INEX document collection changed drastically. Prior to 2006, INEX provided the IEEE document collection which followed the Document Type

Definition (DTD) format. Since 2008, a large collection from Wikipedia has been used. These XML documents contain approximately 30,000 different tags, along with untagged text. This semi-structured collection is approximately 5.6GB in size. In 2009, INEX provided a XML Wikipedia collection which is nearly 10 times larger than any previous collection. It is semi-structured with a size of 50.9GB. It has 2,666,190 articles with over 30,000 unique tags. This same collection has been used for the past three years (2009, 2010 and 2011). A sample document from the raw document collection is shown in Figure 1.

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- generated by CLiX/Wiki2XML [MPI-Inf, MMCI@UdS] $LastChangedRevision: 92 $
on 16.04.2009 16:46:54[mciao0827] -->
<!DOCTYPE article SYSTEM "../article.dtd">
<article xmlns:xlink="http://www.w3.org/1999/xlink">
<header>
<title>Paging</title>
<id>311193</id>
<revision>
<id>244253059</id>
<timestamp>2008-10-09T23:42:12Z</timestamp>
<contributor>
<username>Saga City</username>
<id>138511</id>
</contributor>
</revision>
<categories>
<category>All pages needing cleanup</category>
<category>All articles that may contain original research</category>
<category>Virtual memory</category>
...
</categories>
</header>
<bdy>
This article is about computer virtual memory.  For the wireless communication devices,
see <link xlink:type="simple" xlink:href="../045/229045.xml"> pager</link> and
<link xlink:type="simple" xlink:href="../045/229045.xml"> radio paging</link>.
...
<sec>
<st>Overview</st>
<p>
The main functions of paging are performed when a program tries to access pages that are not
currently mapped to physical memory (<link xlink:type="simple" xlink:href="../847/25847.xml"> RAM
</link>). This situation is known as a <link xlink:type="simple" xlink:href="../143/1157143.xml">
page fault</link>. The operating system must then take control and handle the page fault, in a manner
 invisible to the program. Therefore, the operating system must:
...
</p>
...
<ss1>
<st>
 Anticipatory paging </st>
<p>
This technique preloads a process's non-resident pages that are likely to be referenced in the near
future (taking advantage of <link xlink:type="simple" xlink:href="../028/64028.xml"> locality of
reference</link>). Such strategies attempt to reduce the number of page faults a process experiences.</p>
</ss1>
...
</sec>
...
</bdy>
</article>
```

Figure 1: Sample Document (311193.xml) from 2009 Document Collection

2.4 Query Set

Each participating group creates a set of candidate topics and submits them to INEX. A number of factors are taken into consideration while selecting a topic. They need to reflect real needs of operational systems, be diverse, differ in their convergence etc. The topics are submitted in a Content only and Structure (CO+S) format. The CO+S topic consists of five fields which are explained in Table 1. The INEX website gives a detailed description of the procedure for topic development [7].

Table 1: Fields in a CO+S Topic

Field	Description
<title>	Content Only (CO) queries are given. It serves as a summary of the content of the information needed by the user.
<castitle>	Content And Structure (CAS) queries are given. It specifies structural requirements of the information needed.
<phrasetitle>	Phrase queries are given which is optional.
<description>	Natural language definition of the information given in one or two sentences.
<narrative>	Detailed definition of what makes elements relevant and irrelevant is given.

After all the participating groups submit their topics, INEX considers all the queries and forms the query set for that year. The 2009 collection has 115 queries. In 2010, there are 107 queries. A sample query is shown in Figure 2.

```
</topic>
<topic id="2010057" ct_no="380">
  <title>Einstein Relativity theory</title>
  <castitle>//article[about(., Einstein)]//sec[about(.,relativity theory)]</castitle>
  <phrasetitle>Einstein "Relativity Theory"</phrasetitle>
  <description>Find documents on Einstein in general containing a section on General
  or Special RelativityTheory</description>
  <narrative>Documents about Einstein in general with a section on General or Special Theory
  of Relativity is sought here. Documents about other academic achievements of Einstein is
  also somewhat relevant. Documents about books and films on Einstein are not relevant.
  Documents about religious and political views of Einstein are irrelevant.</narrative>
</topic>
```

Figure 2: Sample Query (Query ID: 2010011) from the 2010 Query-set

2.5 Relevance Assessments

Relevance Assessments are conducted by each participating group using the GPXrai tool [9] [1]. The topics which are judged by a participating group are usually those that were submitted by that group originally. A subset of the document collection is assessed manually for each topic. During relevance assessment, the relevant passages are highlighted using the GPXrai tool. The articles are pooled to form a set of elements that are judged relevant for that query, thus forming the basis for evaluation of the results. The highlighted text is converted into file-offset and length (FOL) format and later used for evaluation. In FOL format, the file-offset refers to the position of the first character of the relevant element and length refers to the number of characters after the offset. (FOL format is explained in detail in Chapter 3.)

2.6 2009 Ad Hoc Tasks

This section presents a description of the tasks of the 2009 INEX Ad Hoc track.

2.6.1 Thorough Task

The Thorough Task seeks to return all the elements that are relevant in a document, hence permitting overlap of elements retrieved [9]. For a given query, all the elements are retrieved and arranged according to their rank. It is evaluated using mean average interpolated precision (MAiP) metric. See [3] for details.

2.6.2 Focused Task

The Focused retrieval task [9] seeks the retrieval of focused elements for each query. For this task, a ranked list of focused (i.e., non-overlapping) elements is returned. The XML document maintains a tree-like structure. If a child is found "relevant," then the parent is also relevant to some degree. In this case, both the parent and child are relevant. Certain overlap removal techniques are used to return only one of these overlapping elements. (The overlap techniques are described in detail in Chapter 3.) The metric used to evaluate Focused retrieval is interpolated precision at 1% recall i.e., IP[0.01]. See [5] for details.

2.6.3 Relevant in Context (RiC) Task

This task is used to return, for each query, all its focused elements, grouped by document [9]. For this purpose, the focused elements are considered and are grouped according to a rank-ordered list of documents. It is evaluated using mean average generalized precision (MAgP) metric.

Evaluation Metric

Each of Ad Hoc tasks uses different evaluation metrics [9]. To understand them better, we need to understand precision and recall. In an experimental document collection, prior assessment has identified relevant elements, and during the retrieval both relevant and irrelevant elements are retrieved. *Recall* is defined as the ratio of the number of relevant elements retrieved to the total number of relevant elements in the document collection. *Precision* is defined as the ratio of the number of relevant elements retrieved to the total number of elements retrieved (both relevant and irrelevant).

In the 2009 evaluation metrics, recall (R) is the fraction of the highlighted text that is retrieved, and precision (P) is the fraction of retrieved text that is highlighted. This thesis focuses on the Ad Hoc RiC task. Its evaluation measure is defined below.

The results of the RiC task are evaluated using Mean Average generalized Precision (MAgP) [9]. This metric gives an overall performance estimate. The formulas to calculate generalized precision (gP) and average generalized precision (AgP) are given the Figures 3 and 4, respectively. A generalized precision is calculated as the sum of document scores up to an article-rank divided by the article-rank. Mean average generalized precision (MAgP) is calculated basically as the mean of the AgP values of individual topics.

2.7 2010 Ad Hoc Tasks

The INEX 2010 Ad Hoc track addresses the impact of length/reading effort and hence focuses on snippet retrieval [1]. In resource-restricted conditions such as a mobile phones or document summary in a hit-list, it is essential that the best relevant elements or

passages are found and retrieved instead of displaying all the relevant elements. The best elements retrieved allow searchers to jump directly to relevant parts of the document; these elements are termed *snippets*. This section describes the tasks involved in 2010 Ad Hoc track.

$$gP[r] = \sum_{i=1}^{r} S(d_i)$$

where

$gP[r]$ is the generalized precision at rank r

d_r is the document with rank r

$S(d_i)$ is the document score indicating the amount of relevant text retrieved. It varies between 0 and 1, where 0 indicates irrelevant document or none of the retrieved text is relevant and 1 indicates all relevant text without any irrelevant text has been retrieved

Figure 3: Formula to Calculate generalized Precision (gP)

$$AgP = \frac{\sum_{r=1}^{|L|}(IsRel(d_r) * gP[r])}{N_{rel}}$$

where

d_r is the document with rank r

$|L|$ is the length of ranked list of documents

$IsRel(d_r)$ is 1 if document d_r contains highlighted relevant text

$IsRel(d_r)$ is 0 if document d_r doesn't contain any highlighted relevant text

$gP[r]$ is the generalized precision at rank r calculated using formula in Figure 3

N_{rel} is the total number of relevant documents

Figure 4: Formula to Calculate Average generalized Precision (AgP)

2.7.1 Relevant in Context (RiC) Task

This task is similar to the 2009 RiC task [9], but it can be viewed as a form of snippet retrieval. It involves two phases, namely, fetching and browsing [1]. In the fetching phase, the potentially relevant articles are identified; in the browsing phase, relevant elements within those articles are to be identified. The retrieved elements should not overlap and should have fewer than 300 intervening irrelevant characters between them.

Evaluation Metric

The 2010 RiC task [1] evaluation metrics differ slightly from those of 2009 [9]. This task uses the *tolerance to irrelevance* T2I(300) measure [1] proposed by the University of Tampere [2]. It takes the reading length and reading order within each document into consideration. Specifically, the per-document score is the character precision at a tolerance to irrelevance (T2I) point. The non-relevant elements retrieved during the retrieval process are penalized strongly. The T2I(300) score per document is used in the measure based on generalized precision and recall in which the number of irrelevant characters between two elements must not exceed 300. The overall performance of this task is measured using MAgP. See Section 2.5.3 for details.

2.7.2 Restricted Relevant in Context (RRiC) Task

This task is a variant of the RiC task (2.7.1) where only 500 characters per article are retrieved, thus simulating resource-restricted conditions directly [1]. A ranked list of articles is required and a maximum of 500 characters per article are retrieved without any overlaps.

Evaluation Metric

The measure proposed for the RiC task (see Section 2.7.1, above) is used for this task also, assuming that it agrees with the MAgP metric used in 2009 Ad Hoc track [1] [9]. Hence, Mean Average generalized Precision (MAgP) can also be used to evaluate the performance of both RiC and RRiC tasks. See Section 2.5.3 for details.

2.7.3 Restricted Focused Task

This task is similar to the 2009 Ad Hoc Focused task [1]. It gives a quick overview of the information relevant in the document collection for the query set. The results are restricted to a maximum of 1,000 characters per topic without any overlaps in the elements retrieved. A ranked list of non-overlapping elements or passages for each topic is returned. A set-based precision over the retrieved characters (character precision) is used as an evaluation metric for the Restricted Focused task [1] [2]. See [5] for details.

2.7.4 Efficiency Task

This task focuses on efficiency rather than effectiveness and exposes the trade-off between them [1]. It is similar to the Thorough Task of the 2009 Ad Hoc track but the results are restricted to a maximum of 15, 150 or 15,000 elements per query. Hence this task returns a ranked list of elements which may have overlapping elements. The results of the efficiency task are evaluated using interpolated Precision (iP) and MAiP [1]. See [3] for details.

2.8 Flexible Retrieval (Flex)

For element retrieval, our method of dynamic element or flexible retrieval called Flex [4], is used. Flex retrieves the elements at run time by generating the child nodes and their correlations first. In the next step, it uses the content of each child to build the parent node. As element retrieval is performed dynamically, flexible retrieval is time and memory efficient.

3. Relevant in Context Tasks

This chapter gives details of the Relevant in Context task and overlap removal techniques. The concept of snippets and the snippet retrieval methodology are also explained. It gives an overview of the fetching and browsing phases that are used to retrieve articles and the elements within those elements.

3.1 2009 Relevant in Context (RiC) Task

The scenario underlying the RiC task is to return non-overlapping elements, grouped by document from a ranked list of articles. Hence this task has two different phases, namely, fetching and browsing. The fetching phase involves identifying the articles that are similar to the query. In the browsing phase, a set of non-overlapping elements within those articles is identified. Later, the results obtained are evaluated. Details are provided below.

3.1.1 Fetching Phase

For the given set of queries, the articles that are similar to the query are identified. For a given query, the *n* top-ranked documents are retrieved.

The following steps are performed before browsing; they build the ground-work necessary for the flexible retrieval of elements. Figure 5 shows the steps involved in this process.

Producing Extended Doctrees

The original document is taken and a preorder traversal of the document tree is done to produce the extended doctree. This is a representation of the original document. It has the expanded path of each element in the document. Figure 6 shows a sample expanded path.

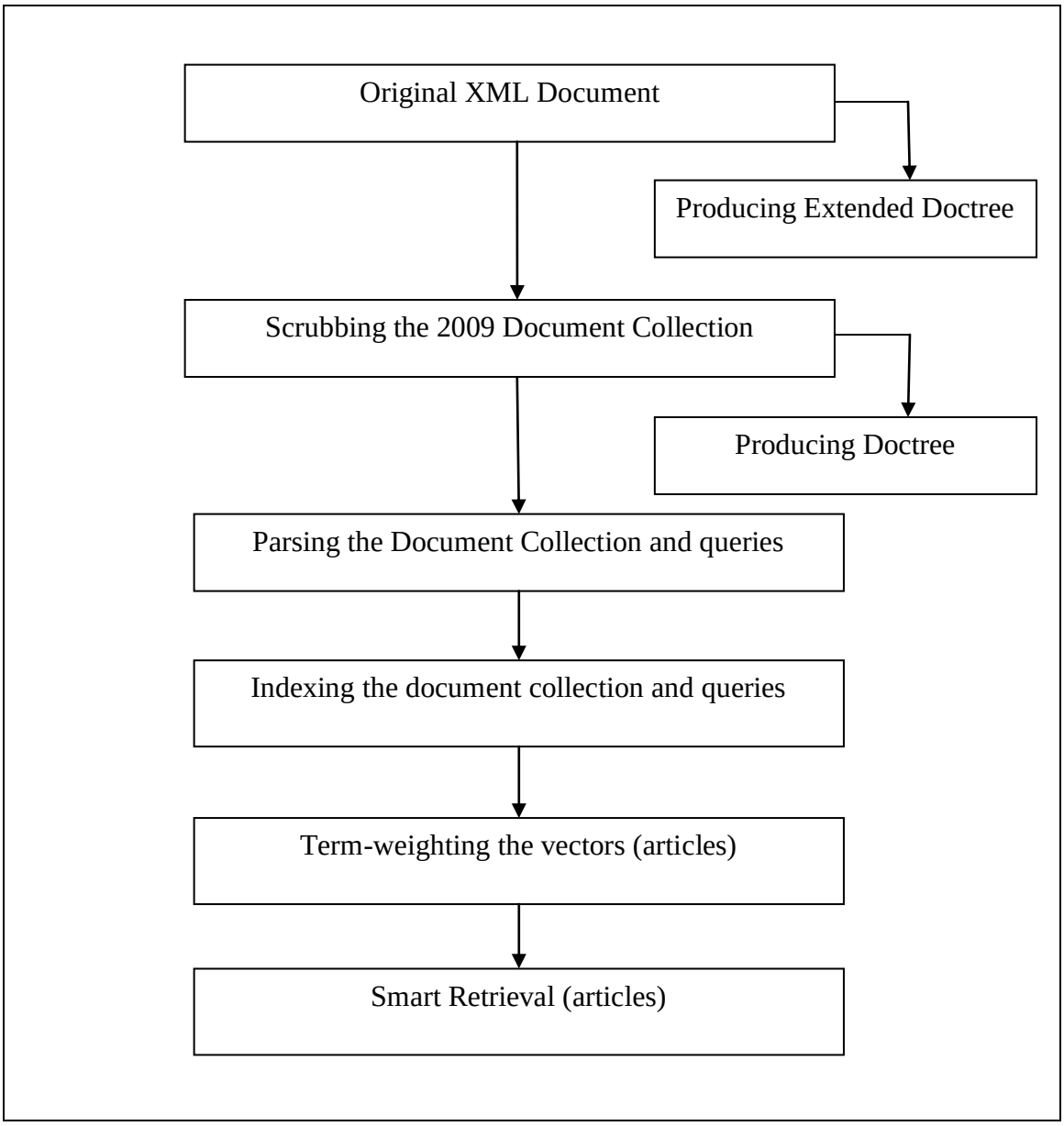

Figure 5: Steps in Fetching Phase

Scrubbing the Document Collection

In this phase, all the unwanted tags from the 2009 document collection (which has 30,000+ unique tags) are removed, but the text between these tags is retained. The untagged text within the semi-structured collection is tagged using *mt* tags. After scrubbing, the document collection is ready for parsing. See [3] for details.

```
/article[1]/
/article[1]/artifact[1]/way[1]/road[1]/header[1]/
/article[1]/artifact[1]/way[1]/road[1]/header[1]/title[1]/
/article[1]/artifact[1]/way[1]/road[1]/header[1]/categories[1]/
/article[1]/artifact[1]/way[1]/road[1]/bdy[1]/
/article[1]/artifact[1]/way[1]/road[1]/bdy[1]/table[1]/
/article[1]/artifact[1]/way[1]/road[1]/bdy[1]/p[1]/
/article[1]/artifact[1]/way[1]/road[1]/bdy[1]/sec[1]/
/article[1]/artifact[1]/way[1]/road[1]/bdy[1]/sec[1]/st[1]/
/article[1]/artifact[1]/way[1]/road[1]/bdy[1]/sec[1]/p[1]/
/article[1]/artifact[1]/way[1]/road[1]/bdy[1]/sec[1]/p[3]/
```

Figure 6: Sample Expanded Doctree for Document 4049328.xml

Producing Doctrees

At this stage, the doctree for each scrubbed document is built. Building the doctrees involves the preorder traversal of the document tree. The doctree contains the Xpath of every recognizable element. These doctrees are later used in flexible retrieval. See Figure 7 for a typical doctree. (Figure 6 displays the corresponding expanded doctree.)

The first column in the doctree is the Xpath, the second column represents the number of children and the third column represents the number of siblings to the right.

```
/article[1]/                        2   0
/article[1]/header[1]/              2   1
/article[1]/header[1]/title[1]/     0   1
/article[1]/header[1]/categories[1]/ 0  0
/article[1]/bdy[1]/                 3   0
/article[1]/bdy[1]/table[1]/        0   1
/article[1]/bdy[1]/p[1]/            0   1
/article[1]/bdy[1]/sec[1]/          3   0
/article[1]/bdy[1]/sec[1]/st[1]/    0   1
/article[1]/bdy[1]/sec[1]/p[1]/     0   1
/article[1]/bdy[1]/sec[1]/p[3]/     0   0
```

Figure 7: Sample Doctree for Document 4049328.xml

Parsing the Document Collection and Queries

The document collection is parsed after it is scrubbed. Parsing is the process of recognizing text within a matching set of XML tags. In this stage, we produce 10 different parses, namely, levels 0-7 the *paras*, and the *para+mt* parses. Each parse corresponds to one level in the document. (See [3] for details.) The queries are also parsed for use during the retrieval process.

Indexing the Document Collection and Queries

The parses produced in the previous step are used as input for indexing. Indexing is performed for level-0 (*articles*), *paras*, *para+mt* and *all-elements*. The level-0 parse is used as input to article index. *Para* parses and *para+mt* parses produce the *para* and *para+mt* indexes respectively. The parses of levels 0-7 and *para* parse are combined to produce all-element index. The article index is used for article retrieval, the *para+mt* index is used by Flex to seed the doctrees, the all-element index is used for slope and pivot experiments, and the para index is used to generate n_stats. The result of indexing is a set of *nnn*-weighted vectors (term frequency vectors), *dict.words* (dictionary of all unique words in the collection), *inv.words* (inverted file) and *textloc.txt* (information about the physical location of the elements being indexed). The queries are also indexed using Smart[11]. See [3] for details.

The textloc.txt file is given as input to the *generate_docid_docpath_mapping.pl* script, which produces the *docid-docpath mapping* file that maps unique Smart identifiers the elements of the document. See Figure 8 for a typical docid-docpath mapping file. The first column in the docid-docpath mapping file is the Smart identifier and the second column represents the Wiki document identifier and the Xpath.

Term-weighting the vectors

The nnn-weighted document and query vectors, (*doc.nnn*) and (*query.nnn*) are converted into *Lnu*-weighted (*doc.Lnu*) and *ltu*-weighted (*query.ltu*) vectors using the *Lnu-ltu* weighting scheme [13]. This scheme attempts to ensure that the length disparities

between the vectors don't bias results unfairly in favor of longer vectors. (The article index is reweighted separately and requires its own slope and pivot values. Calculation of slope and pivot values for this weighing scheme is described in [3].)

```
12518253 4049328/article[1]/header[1]/title[1]/
12518254 4049328/article[1]/header[1]/categories[1]/
12518255 4049328/article[1]/bdy[1]/table[1]/
12518256 4049328/article[1]/bdy[1]/p[1]/
12518257 4049328/article[1]/bdy[1]/sec[1]/st[1]/
12518258 4049328/article[1]/bdy[1]/sec[1]/p[1]/
12518259 4049328/article[1]/bdy[1]/sec[1]/p[3]/
```

Figure 8: Sample Docid-Docpath Mapping File of *para+mt* Parse

Smart Retrieval

The similarity score between the *Lnu*-weighted document vector and *ltu*-weighted query vector is calculated by taking the inner product between them, producing a rank-ordered list of articles for each query (*tr.Lnu.txt*) [11]. This file contains the Smart identifier of the document and its rank. Using the *docid docpath mapping* of articles, this *tr.Lnu.txt* file is converted into a rank-ordered article list with an actual Wiki document identifier. Figure 9 shows a sample article list. The first column corresponds to the query number and the second column represents the rank-ordered articles.

```
query_1 5178308
query_1 3773462
query_1 2597206
query_1 16589150
query_1 4049328
query_1 5507790
query_1 8884612
query_1 348097
query_2 8642120
query_2 1545004
query_2 6733337
query_2 411441
…..
```

Figure 9: Sample Article List

3.1.2 Browsing Phase

A relevant article may contain relevant information spread across different elements. The purpose of this phase is to identify the set of elements corresponding to the relevant information in each relevant article. The elements retrieved in this phase must not overlap.

The following steps are performed to retrieve the elements from the *n* top-ranked documents produced by article retrieval. Figure 10 shows the step-by-step procedure for retrieving the non-overlapping or focused elements.

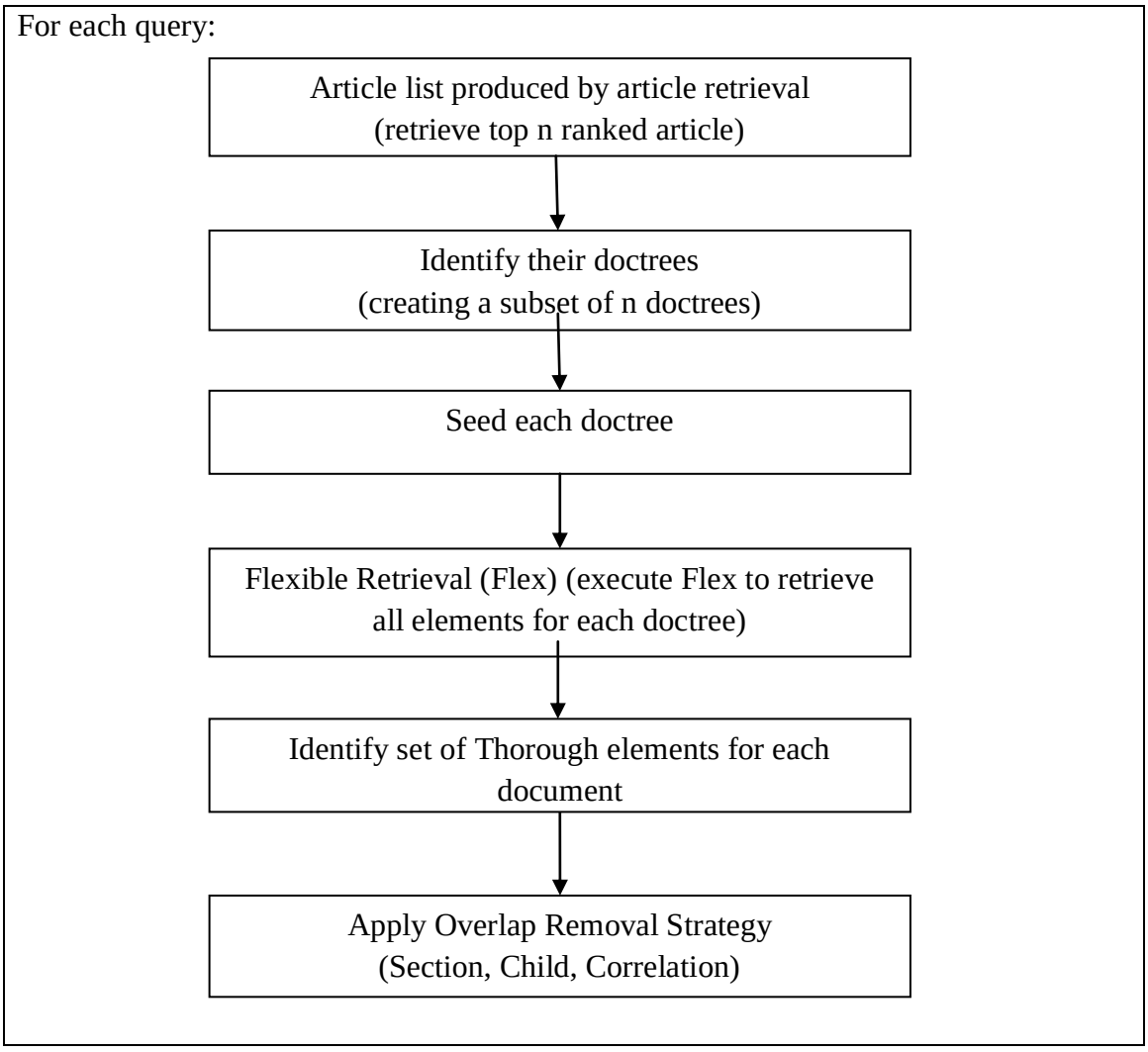

Figure 10: Steps in Browsing Phase

Create a Subset of Doctrees

For each query, the doctree (preorder traversal of a document tree produced in the parsing step) for each of the top-ranked *n* articles retrieved by that query is copied to a separate location, thus creating a subset of *n* doctrees. The doctrees associated with each article retrieved by the query are now grouped together.

Seeding Doctrees

The process of seeding populates the doctree; it fills all the terminal nodes of the doctree with content [5]. All terminal nodes of the doctree are populated at the end of seeding process. This stage requires the doctree subset and docid docpath mapping of the *para+mt* index in order to populate the all of terminal nodes of the doctree. Figure 11 shows a sample seeded document.

4049328/article[1]/	2	0	
4049328/article[1]/header[1]/	2	1	
4049328/article[1]/header[1]/title[1]/	0	1	12518253 0
4049328/article[1]/header[1]/categories[1]/	0	0	12518254 0
4049328/article[1]/bdy[1]/	3	0	
4049328/article[1]/bdy[1]/table[1]/	0	1	12518255 0
4049328/article[1]/bdy[1]/p[1]/	0	1	12518256 0
4049328/article[1]/bdy[1]/sec[1]/	3	0	
4049328/article[1]/bdy[1]/sec[1]/st[1]/	0	1	12518257 0
4049328/article[1]/bdy[1]/sec[1]/p[1]/	0	1	12518258 0
4049328/article[1]/bdy[1]/sec[1]/p[3]/	0	0	12518259 0

Figure 11: Sample Seeded Doctree for Document 4049328.xml

The first column in the seeded doctree represents the Wiki document identifier and the Xpath. The second column gives the number of children; the third column gives the number of siblings on the right. All terminal nodes have a fourth column which is the Smart identifier for that element.

Flexible Retrieval (Flex)

Flex [4] is used to retrieve the elements dynamically at run-time. The seeded trees, all-element values of slope and pivot [3], rank-ordered article list, query.*ltu*, and n_stats (used only for dynamic calculation of *ltu* parameters) are taken as input. Flex follows a bottom-up approach in which the correlation score for leaf nodes is calculated first and then the terminal node content is propagated to the parent. The parent vector is then correlated with the query. In Flex retrieval, based on the rank-ordered article list, a document is selected and all its positively correlated elements are "retrieved." These elements are then sorted by correlation. Hence the output of Flex is a set of (thorough) elements for each document. The *mts* are then removed from the Flex output. Figure 12 shows a typical Flex output after all *mts* are removed. See [5] for details.

```
1 16589150/article[1]/bdy[1]/sec[1]/p[1]/ 5.94792
1 16589150/article[1]/bdy[1]/sec[7]/ 4.94625
1 4049328/article[1]/bdy[1]/ 14.0401
1 4049328/article[1]/bdy[1]/sec[1]/ 13.9299
1 4049328/article[1]/bdy[1]/sec[1]/p[3]/ 10.2189
1 4049328/article[1]/bdy[1]/p[1]/ 5.39235
1 4049328/article[1]/bdy[1]/sec[1]/p[1]/ 4.87045
1 5507790/article[1]/bdy[1]/ 16.0082
1 5507790/article[1]/bdy[1]/sec[1]/p[1]/ 8.82004
```

Figure 12: Sample Flex Output

Overlap Removal Strategies

Flexible retrieval outputs all positively correlated elements within the rank-ordered list of articles. The parent node is built using the child. If a child is positively correlated, the parent is as well. This element set contains overlapping elements which are removed using three different overlap removal techniques, namely section, child and correlation strategies [10]. After overlap is removed, the output contains a ranked list of non-overlapping elements in the form of query id and Xpath. The correlation score is no longer needed and is thus dropped.

(i) Section Strategy

This strategy selects the non-body highest correlating (focused) element. There are three different cases in this strategy.

Case 1: If the parent has higher correlation than the child, and the parent is a non-body element, then the parent element is preferred to the child. Figure 13 shows this case.

Input:
1 4049328/article[1]/bdy[1]/sec[1]/ 13.9299
1 4049328/article[1]/bdy[1]/sec[1]/p[3]/ 10.2189

Output:
1 4049328/article[1]/bdy[1]/sec[1]/

Figure 13: Section Strategy – Case 1

Case 2: If the parent has higher correlation than the child, but the parent is a body element, then the child element is preferred to the parent. Figure 14 shows this case.

Input:
1 4049328/article[1]/bdy[1]/ 14.0401
1 4049328/article[1]/bdy[1]/sec[1]/ 13.9299

Output:
1 4049328/article[1]/bdy[1]/sec[1]/

Figure 14: Section Strategy – Case 2

Case 3: If the child has higher correlation than the parent, then the child element is always preferred and the parent is discarded. Figure 15 shows this case.

Input:
1 4049328/article[1]/bdy[1]/sec[1]/p[3]/ 13.9299
1 4049328/article[1]/bdy[1]/sec[1]/ 10.2189

Output:
1 4049328/article[1]/bdy[1]/sec[1]/p[3]/

Figure 15: Section Strategy – Case 3

(ii) Child Strategy

In this strategy, a child element is always preferred over its parent element even if its correlation is lower than that of the parent element. The two cases in this strategy are described below.

Case 1: If the parent has higher correlation than the child, then the child element is preferred to the parent. Figure 16 shows this case.

Input:
1 4049328/article[1]/bdy[1]/sec[1]/ 13.9299
1 4049328/article[1]/bdy[1]/sec[1]/p[3]/ 10.2189

Output:
1 4049328/article[1]/bdy[1]/sec[1]/p[3]/

Figure 16: Child Strategy – Case 1

Case 2: If the child has higher correlation than the parent, then the child element is preferred to the parent. Figure 17 shows this case.

Input:
1 4049328/article[1]/bdy[1]/sec[1]/p[3]/ 13.9299
1 4049328/article[1]/bdy[1]/sec[1]/ 10.2189

Output:
1 4049328/article[1]/bdy[1]/sec[1]/p[3]/

Figure 17: Child Strategy – Case 2

(iii) Correlation Strategy

This strategy gives preference to the highest correlating element. If there are two overlapping elements where one element is the child of the other, the element with higher correlation score appears in the output and the other element is discarded. Figure 18 shows an example of output produced by this strategy.

Case 1: If the parent has higher correlation than the child, then the parent element is preferred to the child. Figure 18 shows this case.

Input:
1 4049328/article[1]/bdy[1]/sec[1]/ 13.9299
1 4049328/article[1]/bdy[1]/sec[1]/p[3]/ 10.2189

Output:
1 4049328/article[1]/bdy[1]/sec[1]/

Figure 18: Correlation Strategy – Case 1

Case 2: If the child has higher correlation than the parent, then the child element is preferred to the parent. Figure 19 shows this case.

Input:
1 4049328/article[1]/bdy[1]/sec[1]/p[3]/ 13.9299
1 4049328/article[1]/bdy[1]/sec[1]/ 10.2189

Output:
1 4049328/article[1]/bdy[1]/sec[1]/p[3]/

Figure 19: Correlation Strategy – Case 2

3.1.3 Evaluation Phase

The RiC task is evaluated using Mean Average generalized Precision (MAgP), where the generalized score per article is based on the retrieval of highlighted text [9]. The per-article or per-document score reflects how well the retrieved text matches the relevant text within that document. Therefore, the per-document score is the harmonic mean of precision (fraction of text retrieved) and recall (fraction of text highlighted). We use a F_β score with β = 1/4, making precision four times more important than recall. The F-Score is calculated for each retrieved document *d* [9] [1]. Figure 20 shows the formula to calculate per-document score (F-Score) and Figures 3 and 4 show the formulas to calculate MAgP.

$$F_\beta = \frac{(1 + \beta^2) * P(d) * R(d)}{(\beta^2 * P(d)) + R(d)}$$

where

$P(d)$ is the precision for document *d.* See Figure 21.

$R(d)$ is the recall for document *d.* See Figure 22.

Figure 20: per-document Score (F-Score)

$$P(d) = \frac{|rel(d) \cap ret(d)|}{|ret(d)|}$$

where

$rel(d)$ is the amount of relevant text

$ret(d)$ is the amount of retrieved text

Figure 21: Formula to Calculate Precision

$$R(d) = \frac{|rel(d) \cap ret(d)|}{|rel(d)|}$$

where

$rel(d)$ is the amount of relevant text

$ret(d)$ is the amount of retrieved text

Figure 22: Formula to Calculate Recall

Now that the RiC output has no overlapping elements, we evaluate the output. Figure 23 shows the different steps in the evaluation process.

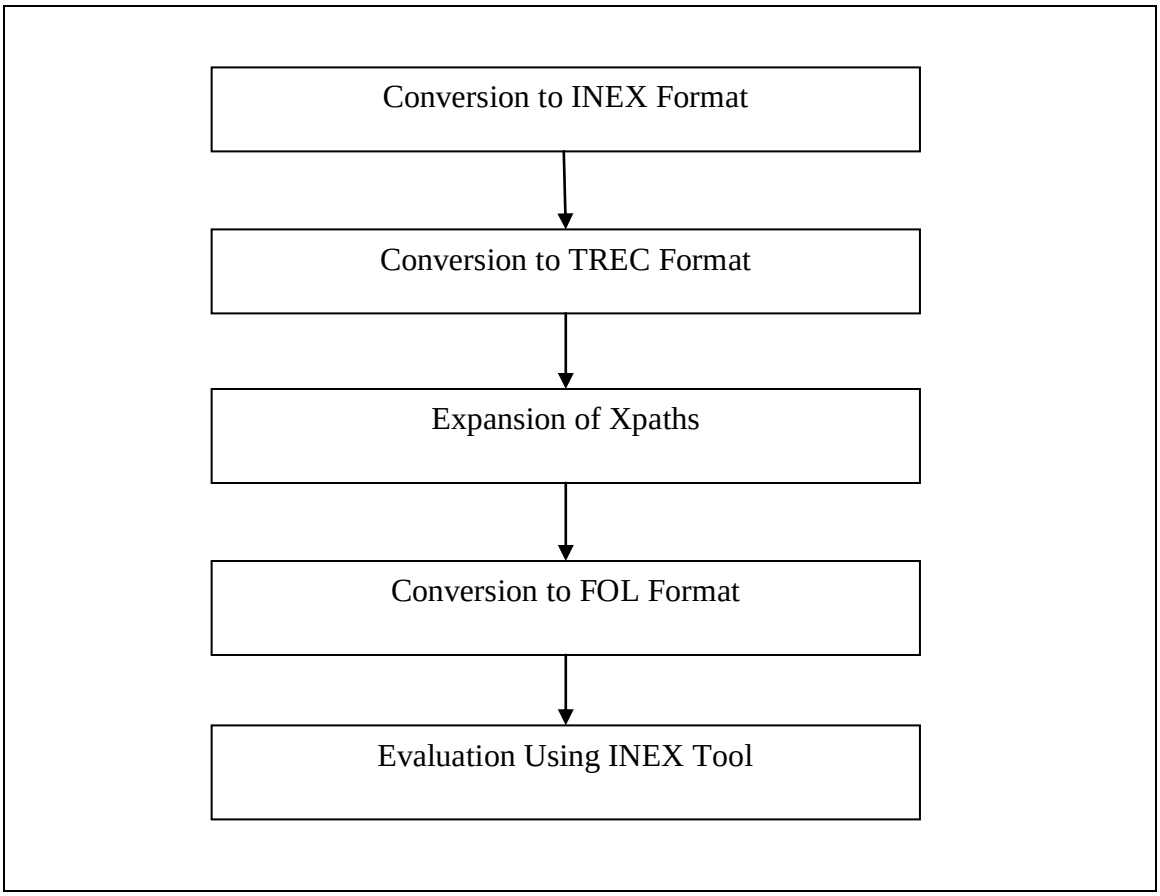

Figure 23: Steps in Evaluation Phase of RiC-2009

Conversion to INEX Format

The RiC output is now converted into XML form in the format approved by INEX. This conversion is achieved by the script provided by the INEX called *convert_to_inex_v2007.pl* [9]. Figure 24 shows a sample INEX format.

Conversion to TREC Format

The output is now converted into standard TREC format [9] as seen in Figure 25. Figure 26 shows a sample TREC file.

```
<inex-submission participant-id="72" run-id="ric" task="RIC" query="automatic" result-type="element">
      <topic-fields title="yes" mmtitle="no" castitle="no" description="no" narrative="no"/>
      <description></description>
      <collections>
            <collection>wikipedia</collection>
      </collections>
      <topic topic-id="2010001">
            <result>
                  <file>9586221</file>
                  <path>/article[1]/bdy[1]/sec[1]/p[1]/</path>
                  <rank>1</rank>
            </result>
            <result>
                  <file>9586221</file>
                  <path>/article[1]/bdy[1]/sec[4]/</path>
                  <rank>2</rank>
            </result>
            …
       </topic>
            …
</inex-submission>
```

Figure 24: Sample File in INEX XML Format

<qid> Q_0 <file> <rank> <rsv> <run_id> <Xpath>

where

<*qid*> is the topic number

Q_i refers to the query ID (Q_0 here)

<*file*> is the document ID of the file from which the element is retrieved

<*rank*> is the rank of the element retrieved for that query

<*rsv*> refers to the inverse rank which has to be unique and in decreasing (non-increasing) order, so that tied scores are handled

<*run_id*> is the name of the task

<*Xpath*> refers to the Xpath of the element within the document referred by <*file*>

Figure 25: Fields in TREC Format

```
2010001 Q0 16589150 120 79881 UMD_RIC /article[1]/bdy[1]/template[1]/
2010001 Q0 16589150 121 79880 UMD_RIC /article[1]/bdy[1]/sec[9]/p[3]/
2010001 Q0 16589150 122 79879 UMD_RIC /article[1]/bdy[1]/sec[3]/p[4]/
2010001 Q0 16589150 123 79878 UMD_RIC /article[1]/bdy[1]/p[3]/
2010001 Q0 16589150 124 79877 UMD_RIC /article[1]/bdy[1]/sec[7]/p[5]/
2010001 Q0 16589150 125 79876 UMD_RIC /article[1]/bdy[1]/sec[1]/p[1]/
2010001 Q0 4049328 126 79875 UMD_RIC /article[1]/bdy[1]/sec[1]/p[3]/
2010001 Q0 4049328 127 79874 UMD_RIC /article[1]/bdy[1]/p[1]/
2010001 Q0 4049328 128 79873 UMD_RIC /article[1]/bdy[1]/sec[1]/p[1]/
```

Figure 26: Sample TREC File

Expansion of Xpaths

The Xpath in the TREC format is not complete. It contains only the tags used in indexing (i.e., those following the *bdy* tag in the Xpath). The extended doctrees are used to patch these paths in the TREC file.

Extended doctrees are the preorder traversal of the document tree with the entire path as its Xpath. Figure 27 shows the expanded TREC file. (Figure 6 shows the extended doctrees.)

```
2010001 Q0 16589150 120 79881 UMD_RIC/article[1]/law_enforcement_agency[1]/bdy[1]/template[1]
2010001 Q0 16589150 121 79880 UMD_RIC /article[1]/law_enforcement_agency[1]/bdy[1]/sec[9]/p[3]
2010001 Q0 16589150 122 79879 UMD_RIC /article[1]/law_enforcement_agency[1]/bdy[1]/sec[3]/p[4]
2010001 Q0 16589150 123 79878 UMD_RIC /article[1]/law_enforcement_agency[1]/bdy[1]/p[3]
2010001 Q0 16589150 124 79877 UMD_RIC /article[1]/law_enforcement_agency[1]/bdy[1]/sec[7]/p[5]
2010001 Q0 16589150 125 79876 UMD_RIC /article[1]/law_enforcement_agency[1]/bdy[1]/sec[1]/p[1]
2010001 Q0 4049328 126 79875 UMD_RIC /article[1]/artifact[1]/way[1]/road[1]/bdy[1]/sec[1]/p[3]
2010001 Q0 4049328 127 79874 UMD_RIC /article[1]/artifact[1]/way[1]/road[1]/bdy[1]/p[1]
2010001 Q0 4049328 128 79873 UMD_RIC /article[1]/artifact[1]/way[1]/road[1]/bdy[1]/sec[1]/p[1]
```

Figure 27: Sample Expanded TREC File

Conversion to FOL format

INEX provides an evaluation tool for the evaluation of various 2009 Ad Hoc tasks. The SUB2FOL.jar converts the expanded TREC file into a FOL (File Offset Length) format [9]. (INEX requires the submission of files in FOL format which is a variant of TREC format.) This format has 8 fields; some are similar to the fields in the TREC format. Figure 28 shows the fields in the FOL format.

<qid> Q_0 <file> <rank> <rsv> <run_id> <offset> <length>

where

<*qid*> is the topic number

<Q_0> is the query number

<*file*> is the document ID of the file from which the element is retrieved

<*rank*> is the rank of the element

<*rsv*> is the inverse rank which has to be unique and in decreasing (non-increasing) order, so that tied scores are handled

<*run_id*> is the name of the task

<*offset*> and <*length*> are calculated in characters with respect to the textual contents of the XML file:

<*offset*> is the number of characters between the start of the file and start of the element

<*lengtht*> is the length of the element

Figure 28: Fields in FOL Format

The FOL file may contain <*offset*> and <*length*> as -1 if the element is invalid (i.e., too small). Hence these elements are removed from the element list and additional elements are considered. In these experiments the number of elements, may vary from 50 to 1500 (i.e., *n* = 50, 100, 150, 200, 250, 500, 1000 and 1500). Figure 29 shows the FOL file with invalid nodes, and Figure 30 shows same FOL file with replacement by valid nodes.

```
2010001 Q0 16589150 120 79881.0 UMD_RIC 216 375
2010001 Q0 16589150 121 79880.0 UMD_RIC 5041 1493
2010001 Q0 16589150 122 79879.0 UMD_RIC 2091 110
2010001 Q0 16589150 123 79878.0 UMD_RIC 874 220
2010001 Q0 16589150 124 79877.0 UMD_RIC 4828 163
2010001 Q0 16589150 125 79876.0 UMD_RIC 1108 371
2010001 Q0 4049328 126 79875.0 UMD_RIC 387 251
2010001 Q0 4049328 127 79874.0 UMD_RIC 679 363
2010001 Q0 4049328 128 79873.0 UMD_RIC 1078 2707
2010001 Q0 5507790 129 79872.0 UMD_RIC -1 -1
2010001 Q0 5507790 130 79871.0 UMD_RIC -1  -1
2010001 Q0 8884612 131 79870.0 UMD_RIC 102 639
```

Figure 29: Sample FOL File with Invalid Nodes

```
2010001 Q0 16589150 120 79881.0 1 216 375
2010001 Q0 16589150 121 79880.0 1 5041 1493
2010001 Q0 16589150 122 79879.0 1 2091 110
2010001 Q0 16589150 123 79878.0 1 874 220
2010001 Q0 16589150 124 79877.0 1 4828 163
2010001 Q0 16589150 125 79876.0 1 1108 371
2010001 Q0 4049328 126 79875.0 1 387 251
2010001 Q0 4049328 127 79874.0 1 679 363
2010001 Q0 4049328 128 79873.0 1 1078 2707
2010001 Q0 8884612 131 79870.0 1 102 639
```

Figure 30: Sample FOL File After Removing Invalid Nodes

Evaluation Using INEX Tool

INEX also provides an inex-eval.jar tool which uses MAgP (for the RiC task) to evaluate the FOL file. A qrels file for the 2009 queries with manual relevance assessments converted into FOL format is also available. The inex-eval tool uses the qrels file and FOL file as input and gives the result shown in Figure 31.

3.2 2010 Relevant in Context (RiC) Task

The 2010 RiC task is a variant of the 2009 RiC task which differs mainly in the evaluation metric. The scenario underlying the 2010 RiC task is to return the relevant information that is captured by a set of non-overlapping elements or passages within the context of the full document. At INEX 2010, this task is interpreted as a form of snippet retrieval and evaluation factors in result length or reading effort [1] [2].

This task also has the fetching and browsing phases in which the relevant articles are selected and the focused elements are identified. A few changes are made to the final results to be submitted to accommodate the snippets. After the relevant articles are retrieved, the browsing of the elements continues in document order. Hence in this case, flexible retrieval retrieves all the elements (from the documents in the rank-ordered article list). While evaluating the results, the elements which have fewer than 300 irrelevant characters between them are extracted and then evaluated.

```
<eval run-id="1" file="RIC_subtoFOL_1500.trec++">
num_q          all     52
num_ret        all     27271
num_rel        all     5471
num_rel_ret    all     3806
ret_size       all     66070932
rel_size       all     17641119
rel_ret_size   all     3263608
MAgP           all     0.17525218535220258
gP[1]          all     0.33649838639073526
gR[1]          all     0.01969663412898154
gP[3]          all     0.3554966826756972
gR[3]          all     0.04169196314697102
gP[4]          all     0.34019377257370353
gR[4]          all     0.055993758172551256
gP[6]          all     0.32628748829025744
gR[6]          all     0.08467317541705312
gP[11]         all     0.2824293592260972
gR[11]         all     0.14354119477451283
gP[25]         all     0.2261021758181687
gR[25]         all     0.25350799302679905
gP[50]         all     0.18997776022968157
gR[50]         all     0.3476831159037635
ircl_prn.0.00  all     0.4961703454876329
ircl_prn.0.10  all     0.3642929840944426
ircl_prn.0.20  all     0.3034059631781176
ircl_prn.0.30  all     0.26281956860386246
ircl_prn.0.40  all     0.21816137133609662
ircl_prn.0.50  all     0.17357460719563947
ircl_prn.0.60  all     0.13660816475742735
ircl_prn.0.70  all     0.09095661977121038
ircl_prn.0.80  all     0.056438394646239745
ircl_prn.0.90  all     0.028074514944758266
ircl_prn.1.00  all     0.015167171045945664
</eval>
```

Figure 31: Sample Output of the INEX Evaluation Tool

3.2.1 Evaluation Phase

The 2010 RiC task is evaluated using an effort-based measure. This metric uses a different per-document score that takes reading effort into consideration. Here, the per-document or per-article score is the character precision at a tolerance to irrelevance point (T2I) [2]. Tolerance to irrelevance is the point at which the user loses interest in the article due to the amount of irrelevant text.

In this task, the user is expected to read the elements returned in document order. The user stops reading only when he/she comes across 300 or more irrelevant characters or all characters of the document are read. The default number of irrelevant characters is 300 and hence this measure is known as T2I(300) [1] [2]. This score is used in the measure based on generalized precision and recall, penalizing the retrieval of irrelevant text. To measure overall performance of this task, the MAgP measure is used. See Section 2.5.3 for details. Figure 32 shows the steps in the evaluation phase.

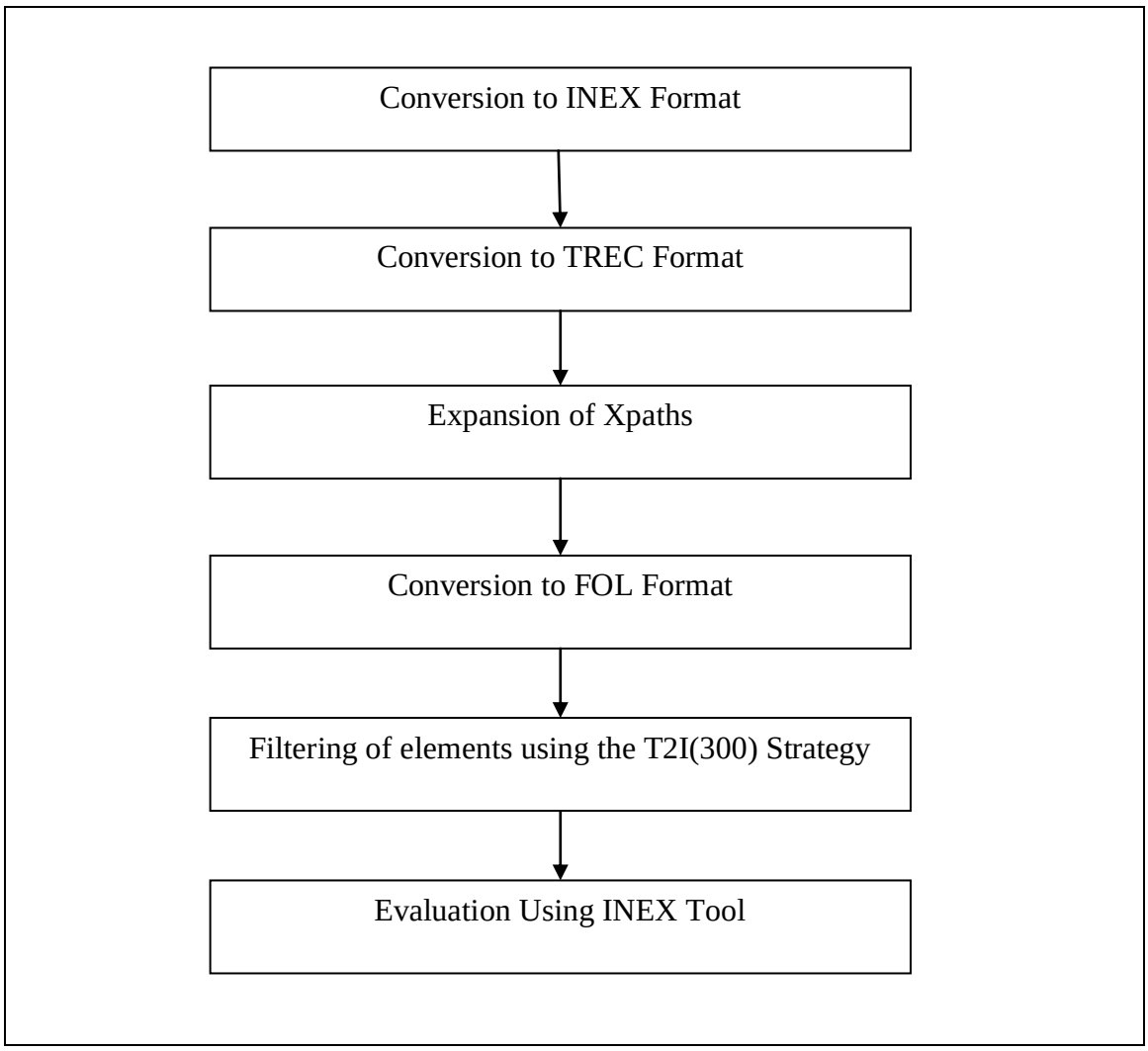

Figure 32: Steps in Evaluation Phase of RiC-2010

The steps mentioned in Section 3.1.3 are performed till a FOL file with all valid nodes is obtained. Before evaluation, the elements which contain more than 300 characters of irrelevant text between them are discarded using the T2I(300) strategy. The resultant file has nodes with fewer than 300 irrelevant characters between them. A qrels file for 2010 queries (with manual relevance assessments converted to FOL format) is also available. Figure 33 shows the file that is further evaluated using the inex-eval.jar tool, which takes the qrels file and FOL file as input. This file is further evaluated using the inex-eval.jar tool to get the MAgP score.

```
2010001 Q0 16589150 53 1448 UMD_RIC 216 375
2010001 Q0 16589150 54 1447 UMD_RIC 874 220
2010001 Q0 16589150 55 1446 UMD_RIC 1108 371
2010001 Q0 4049328 56 1445 UMD_RIC 387 251
2010001 Q0 4049328 57 1444 UMD_RIC 679 363
2010001 Q0 4049328 58 1443 UMD_RIC 1078 2707
```

Figure 33: Sample FOL File After Applying T2I(300) Strategy

3.3 2010 Restricted Relevant in Context (RRiC) Task

Snippets give an overview of the article. In resource-restricted conditions, it is essential to return only elements which give an overview of the document.

This task is a variant of the 2010 RiC task wherein only 500 characters per article are allowed [1]. It requires a ranked list of non-overlapping elements covering the relevant material in the article. A maximum of 500 characters per article are retrieved.

3.3.1 Evaluation Phase

The evaluation of the RRiC task is the same as that of the (unrestricted) RiC task. It also uses T2I(300) as per-document score and the main performance measure is MAgP based on T2I(300). (See Section 2.5.3 for details.) Figure 34 shows the steps of evaluation.

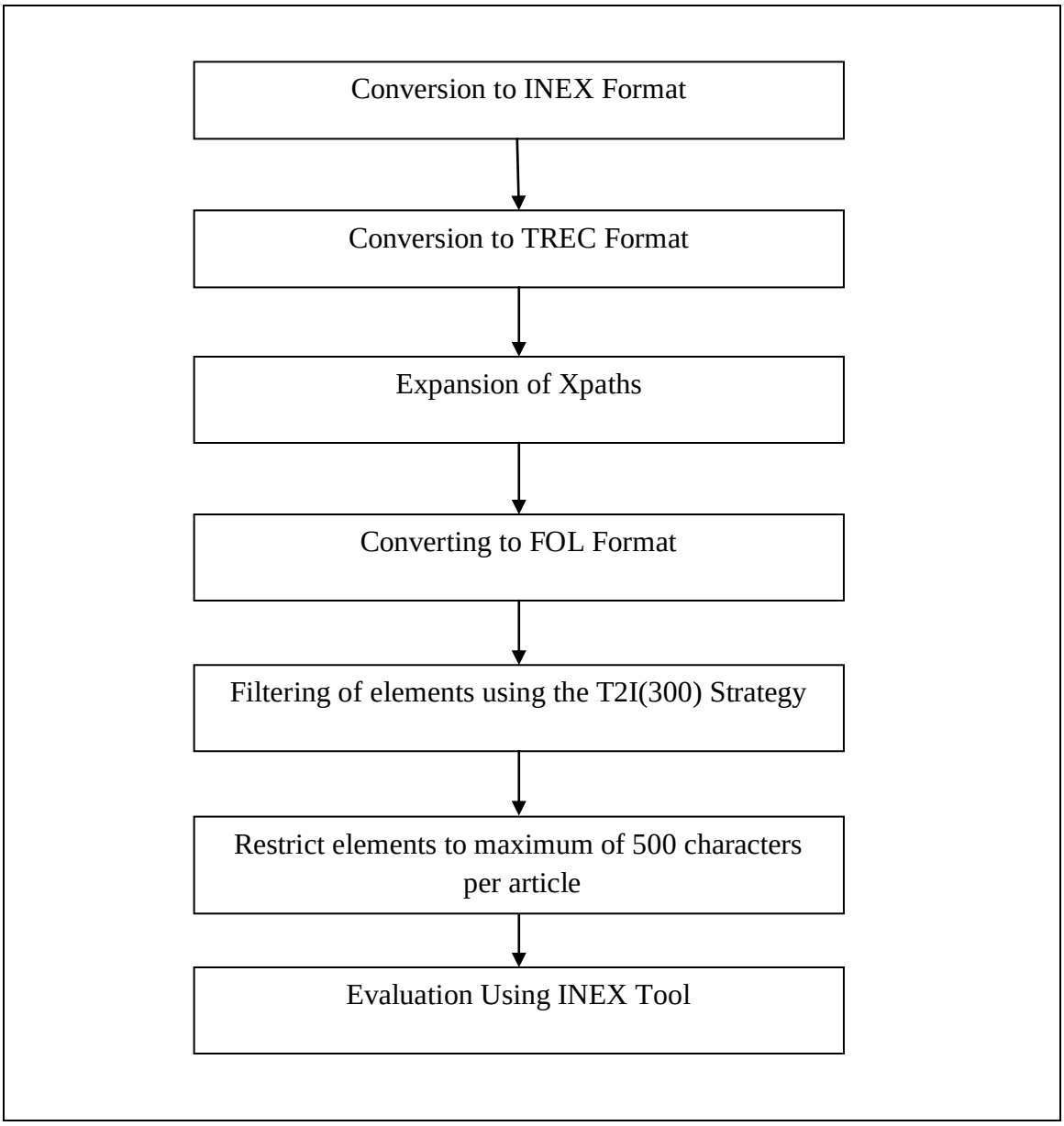

Figure 34: Steps in Evaluation Phase of RRiC-2010

After all elements from the documents in the rank-ordered article list are extracted, the steps in Section 3.1.3 are followed till a FOL with all valid nodes is generated. The T2I(300) strategy is applied on the FOL file. As the main specification of RRiC is to meet resource-restricted conditions, the length of the element (per article) is maximally 500 characters. Therefore the FOL obtained after applying the T2I(300) strategy is taken and the length column in the FOL is chopped to accommodate 500 character per article.

Figure 35 shows a sample FOL file after extracting the snippet of 500 characters.

```
2010001 Q0 16589150 22 1479 UMD_RIC 216 375
2010001 Q0 16589150 23 1478 UMD_RIC 874 125
2010001 Q0 4049328 24 1477 UMD_RIC 387 251
2010001 Q0 4049328 25 1476 UMD_RIC 679 249
2010001 Q0 8884612 26 1475 UMD_RIC 102 500
```

Figure 35: Sample FOL File After Chopping to 500 Characters

This file along with 2010 qrels is further evaluated using the inex-eval.jar tool to get the MAgP score.

4. Experiments, Results and Analysis

This chapter discusses the experiments performed for the 2009 RiC task, the 2010 RiC task, and the 2010 RRiC task. The experiments are presented and a detailed analysis of the experiments is also provided. There are 115 queries for 2009 task and 107 queries for 2010 task. The same slope and pivot values are used for all the tasks as the document collection is the same (see Table 2). In each case, the reference run provided by INEX [8] is used for article retrieval. Flex generates all Thorough elements for the top *n* documents, so any invalid nodes (these are judged too small to be meaningful) are eliminated so the next element in rank order can take its place in the element list.

Table 2: Slope and Pivot Values

	Slope	**Pivot**
All-elements Retrieval	0.11	38

4.1 Methodology for Relevant in Context (RiC) Task-2009

For this task, the focused elements retrieved for a document are sorted by correlation. Therefore, for each document the elements with positive correlation are arranged in decreasing order of their correlation. This is normal Flex output. The overlap removal techniques are applied to this output, which is then converted to INEX format and finally to TREC format. The XPaths are expanded and the TREC file is converted to FOL. This file contains the focused elements for *n* documents, where *n* ranges from 25 to 1500. Then for each query, *m* elements (where *m* ranges from 50 to 1500) are extracted and are further evaluated to obtain the MAgP score.

4.1.1 Relevant in Context Task-2009 Experiments

These experiments are performed using the 2009 Wiki document collection, the 2009 query set and each of the three overlap removal techniques. Results are evaluated using the 2009 INEX evaluation tool, and best results are highlighted.

Experiment 1 (2009 RiC)

In this experiment, all the three overlap removal techniques, namely, section, child and correlation [10], are used to obtain non-overlapping elements for RiC 2009 task. (See Section 3.1.2 for details). Tables 3 - 5 give the results of this experiment.

Table 3: MAgP Section Strategy RiC 2009

# of Docs	# of Elements							
	50	**100**	**150**	**200**	**250**	**500**	**1000**	**1500**
25	0.0622	0.0885	0.1017	0.1052	0.1054	0.1054	0.1054	0.1054
50	0.0622	0.0896	0.1063	0.1184	0.1247	0.1293	0.1293	0.1293
100	0.0622	0.0896	0.1067	0.1194	0.1286	0.1485	0.1510	0.1510
150	0.0622	0.0896	0.1067	0.1194	0.1286	0.1512	0.1599	0.1600
200	0.0622	0.0896	0.1067	0.1194	0.1286	0.1514	0.1647	0.1666
250	0.0622	0.0896	0.1067	0.1194	0.1286	0.1515	0.1664	0.1706
500	0.0622	0.0896	0.1067	0.1194	0.1286	0.1515	0.1672	0.1747
1000	0.0622	0.0896	0.1067	0.1194	0.1286	0.1515	0.1672	**0.1749**
1500	0.0622	0.0896	0.1067	0.1194	0.1286	0.1515	0.1672	**0.1749**

Table 4: MAgP Child Strategy RiC 2009

# of Docs	# of Elements							
	50	**100**	**150**	**200**	**250**	**500**	**1000**	**1500**
25	0.0438	0.0662	0.0802	0.0912	0.0978	0.1050	0.1052	0.1052
50	0.0438	0.0662	0.0813	0.0941	0.1037	0.1257	0.1292	0.1292
100	0.0438	0.0662	0.0813	0.0945	0.1041	0.1346	0.1486	0.1507
150	0.0438	0.0662	0.0813	0.0945	0.1042	0.1349	0.1538	0.1584
200	0.0438	0.0662	0.0813	0.0945	0.1042	0.1350	0.1553	0.1624
250	0.0438	0.0662	0.0813	0.0945	0.1042	0.1350	0.1556	0.1644
500	0.0438	0.0662	0.0813	0.0945	0.1042	0.1350	0.1560	0.1655
1000	0.0438	0.0662	0.0813	0.0945	0.1042	0.1350	0.1560	**0.1656**
1500	0.0438	0.0662	0.0813	0.0945	0.1042	0.1350	0.1560	**0.1656**

Table 5: MAgP Correlation Strategy RiC 2009

# of Docs	# of Elements							
	50	**100**	**150**	**200**	**250**	**500**	**1000**	**1500**
25	0.1064	0.1076	0.1078	0.1078	0.1078	0.1078	0.1078	0.1078
50	0.1206	0.1313	0.1317	0.1320	0.1321	0.1321	0.1321	0.1321
100	0.1206	0.1420	0.1515	0.1537	0.1538	0.1542	0.1542	0.1542
150	0.1206	0.1420	0.1548	0.1608	0.1621	0.1631	0.1634	0.1634
200	0.1206	0.1420	0.1548	0.1617	0.1654	0.1693	0.1703	0.1703
250	0.1206	0.1420	0.1548	0.1617	0.1657	0.1732	0.1748	0.1748
500	0.1206	0.1420	0.1548	0.1617	0.1657	0.1786	0.1845	0.1847
1000	0.1206	0.1420	0.1548	0.1617	0.1657	0.1786	0.1869	0.1885
1500	0.1206	0.1420	0.1548	0.1617	0.1657	0.1786	0.1869	**0.1889**

Observations (2009 RiC)

The observations made for the results of RiC 2009 task are discussed below.

Figure 36 shows a comparison of the best MAgP score obtained for all the three overlap removal strategies for the 2009 RiC task.

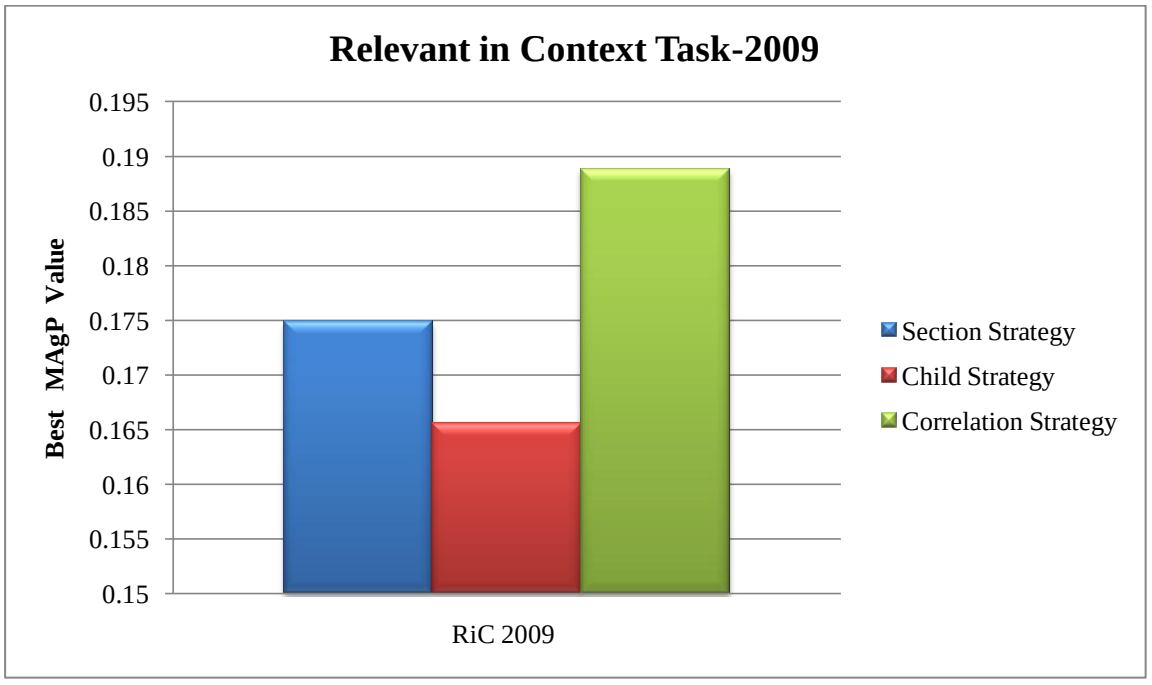

Figure 36: Comparison of all Strategies for 2009 RiC Task

1. It is observed that amongst the three overlap removal strategies, correlation strategy performed well when compared to other two strategies.

2. The **correlation strategy** has the best MAgP score of **0.1889** placing UMD in **1st place.** Table 6 gives the MAgP value and ranking of the top-ranked participants [9] for comparison. There were 19 participants that submitted results for the 2009 Ad Hoc Track.

3. Using a confidence interval of 95% in a one-tailed t-test, we found that the INEX 2009 RiC Task results were significantly different from the results of the participants ranked 2, 3, 5, 8 and 10. There were no significant differences between our results and those of the participants at ranks 1, 4, 6, 7, and 9.

Table 6: Ranking of Top-10 2009 RiC Task Participants

Participant	**MAgP Value**	**Rank**
University of Minnesota Duluth* (Correlation Strategy)	**0.1889**	-
Queensland University of Technology	0.1885	1
University of Otago	0.1847	2
University of Amsterdam	0.1773	3
LIG	0.1760	4
University of Minnesota Duluth* (Section Strategy)	**0.1749**	-
University of Tampere	0.1720	5
University of Minnesota Duluth* (Child Section)	**0.1656**	-
University of Twente	0.1188	6
Saint Etienne University	0.1075	7
School of Electronic Engineering and Computer Science	0.1045	8
Renmin University of China	0.1028	9
University of Minnesota Duluth** (early run)	0.0424	10

4.2 Methodology for Relevant in Context (RiC) Task-2010

For this task, the focused elements retrieved for a document are sorted by correlation. This is normal Flex output. The overlap removal techniques are applied to this output, which is then converted to INEX format and finally to TREC format. The XPaths are expanded and the TREC file is converted to FOL. The same values of *n* and *m* are used. The T2I strategy is applied to all elements. Using this strategy, the elements which have fewer than 300 irrelevant characters between them are retained and all other elements within that document are discarded. For each query, *m* elements are extracted from the resultant file and further evaluated to obtain the MAgP score.

4.2.1 Relevant in Context Task-2010 Experiments

The experiments described in this section are performed using the 2009 Wiki document collection, the 2010 query set and each of the three overlap removal techniques. Results are evaluated using the INEX 2009 F-Score as well as INEX 2010 T2I-Score provided by the 2010 INEX evaluation tool. Best results are highlighted.

Experiment 2 (2010 RiC using F-Score)

In this experiment, all three overlap removal techniques, namely, section, child and correlation [10] are used to obtain non-overlapping elements for RiC 2010 task. (See Section 3.1.2 for details). While evaluating the elements, the F-Score metric is used. Tables 7 - 9 give the results of this experiment.

Table 7: MAgP Section Strategy RiC 2010 F-Score

# of Docs	# of Elements							
	50	**100**	**150**	**200**	**250**	**500**	**1000**	**1500**
25	0.0768	0.0931	0.0933	0.0933	0.0933	0.0933	0.0933	0.0933
50	0.0768	0.1001	0.1114	0.1159	0.1160	0.1160	0.1160	0.1160
100	0.0768	0.1001	0.1126	0.1245	0.1333	0.1418	0.1418	0.1418
150	0.0768	0.1001	0.1126	0.1245	0.1337	0.1551	0.1556	0.1556
200	0.0768	0.1001	0.1126	0.1245	0.1337	0.1594	0.1633	0.1633
250	0.0768	0.1001	0.1126	0.1245	0.1337	0.1601	0.1694	0.1694
500	0.0768	0.1001	0.1126	0.1245	0.1337	0.1601	0.1779	0.1825
1000	0.0768	0.1001	0.1126	0.1245	0.1337	0.1601	0.1780	**0.1838**
1500	0.0768	0.1001	0.1126	0.1245	0.1337	0.1601	0.1780	**0.1838**

Table 8: MAgP Child Strategy RiC 2010 F-Score

# of Docs	# of Elements							
	50	**100**	**150**	**200**	**250**	**500**	**1000**	**1500**
25	0.0688	0.0868	0.0881	0.0881	0.0881	0.0881	0.0881	0.0881
50	0.0688	0.0898	0.1032	0.1097	0.1104	0.1104	0.1104	0.1104
100	0.0688	0.0898	0.1036	0.1139	0.1227	0.1357	0.1357	0.1357
150	0.0688	0.0898	0.1036	0.1139	0.1230	0.1473	0.1491	0.1491
200	0.0688	0.0898	0.1036	0.1139	0.1230	0.1497	0.1567	0.1567
250	0.0688	0.0898	0.1036	0.1139	0.1230	0.1499	0.1625	0.1625
500	0.0688	0.0898	0.1036	0.1139	0.1230	0.1499	0.1686	0.1747
1000	0.0688	0.0898	0.1036	0.1139	0.1230	0.1499	0.1686	0.1752
1500	0.0688	0.0898	0.1036	0.1139	0.1230	0.1499	0.1686	**0.1753**

Table 9: MAgP Correlation Strategy RiC 2010 F-Score

# of Docs	# of Elements							
	50	**100**	**150**	**200**	**250**	**500**	**1000**	**1500**
25	0.0879	0.0879	0.0879	0.0879	0.0879	0.0879	0.0879	0.0879
50	0.1035	0.1085	0.1085	0.1085	0.1085	0.1085	0.1085	0.1085
100	0.1035	0.1253	0.1315	0.1319	0.1319	0.1319	0.1319	0.1319
150	0.1035	0.1253	0.1382	0.1436	0.1443	0.1445	0.1445	0.1445
200	0.1035	0.1253	0.1382	0.1464	0.1507	0.1516	0.1516	0.1516
250	0.1035	0.1253	0.1382	0.1464	0.1520	0.1571	0.1571	0.1571
500	0.1035	0.1253	0.1382	0.1464	0.1520	0.1672	0.1697	0.1697
1000	0.1035	0.1253	0.1382	0.1464	0.1520	0.1672	0.1740	0.1753
1500	0.1035	0.1253	0.1382	0.1464	0.1520	0.1672	0.1740	**0.1764**

Observations (2010 RiC using F-Score)

Figure 37 shows a comparison of the best MAgP score obtained using the F-Score for all the three overlap removal strategies, namely section, child and correlation for the 2010 RiC task.

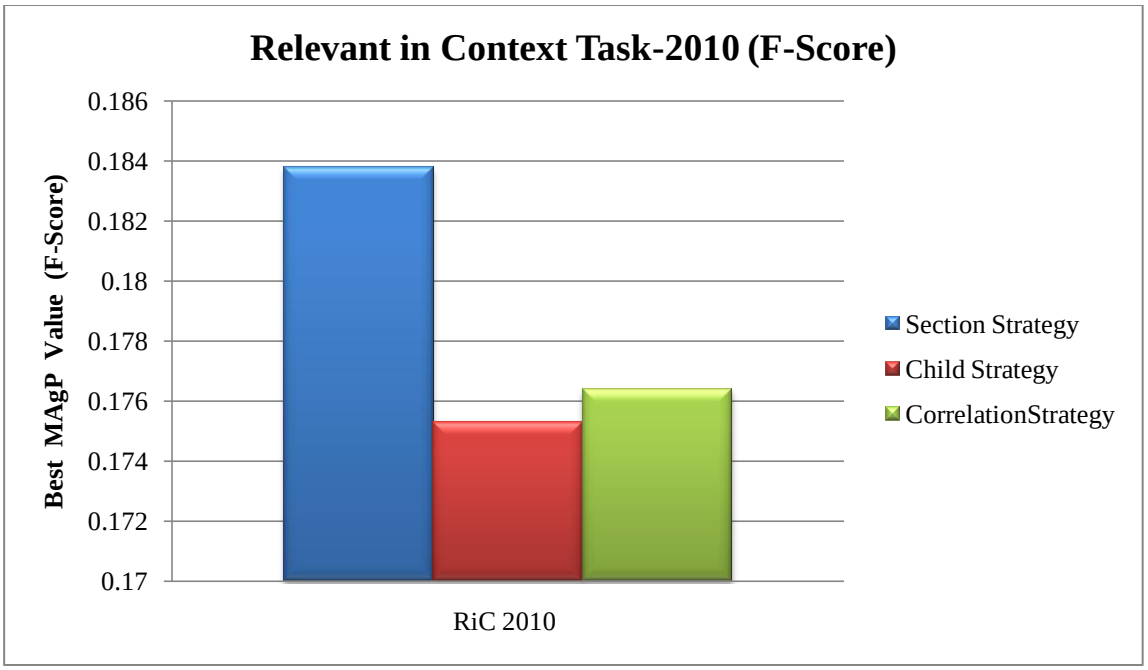

Figure 37: Comparison of all Strategies for 2010 RiC Task (F-Score)

1. It is observed that the section strategy performed well when compared to other two overlap removal strategies.

2. The **section strategy** has the best MAgP score of **0.1838** placing UMD in **2nd place.** Table 10 gives the MAgP value and ranking of the top ten participants [1] for comparison. There were 18 participants that submitted results for the 2010 Ad Hoc Track.

We are still awaiting data for significance testing of our results for the INEX 2010 RiC Task (using F-Score).

Table 10: Ranking of Top-10 2010 RiC Task (F-Score) Participants

Participant	MAgP Value	Rank
ENSM-SE	0.1970	1
University of Minnesota Duluth* (Section Strategy)	**0.1838**	-
University of Minnesota Duluth* (Correlation Strategy)	**0.1764**	-
University of Minnesota Duluth* (Child Strategy)	**0.1753**	-
Peking University	0.1726	2
University of Otago	0.1710	3
Renmin University of China	0.1671	4
Radboud University Nijmegen	0.1623	5
RMIT University	0.1541	6
LIA - University of Avignon	0.1298	7
Doshisha University	0.1122	8
Indian Statistical Institute	0.0693	9
Queensland University of Technology	0.0634	10

Experiment 3 (2010 RiC using T2I-Score)

In this experiment, all three overlap removal techniques, namely, section, child and correlation, are used to obtain non-overlapping elements for RiC 2010 task. (See Section 3.1.2 for details). While evaluating the elements, the T2I-Score metric is used. Tables 11 - 13 give the results of this experiment.

Table 11: MAgP Section Strategy RiC 2010 T2I-Score

# of Docs	# of Elements							
	50	**100**	**150**	**200**	**250**	**500**	**1000**	**1500**
25	0.0949	0.0965	0.0965	0.0965	0.0965	0.0965	0.0965	0.0965
50	0.0969	0.1178	0.1192	0.1192	0.1192	0.1192	0.1192	0.1192
100	0.0969	0.1217	0.1369	0.1435	0.1448	0.1448	0.1448	0.1448
150	0.0969	0.1217	0.1374	0.1480	0.1550	0.1580	0.1580	0.1580
200	0.0969	0.1217	0.1374	0.1481	0.1562	0.1650	0.1650	0.1650
250	0.0969	0.1217	0.1374	0.1481	0.1563	0.1703	0.1706	0.1706
500	0.0969	0.1217	0.1374	0.1481	0.1563	0.1747	0.1827	0.1828
1000	0.0969	0.1217	0.1374	0.1481	0.1563	0.1747	0.1847	0.1876
1500	0.0969	0.1217	0.1374	0.1481	0.1563	0.1747	0.1847	**0.1877**

Table 12: MAgP Child Strategy RiC 2010 T2I-Score

# of Docs	# of Elements							
	50	**100**	**150**	**200**	**250**	**500**	**1000**	**1500**
25	0.0914	0.0946	0.0946	0.0946	0.0946	0.0946	0.0946	0.0946
50	0.0930	0.1140	0.1170	0.1170	0.1170	0.1170	0.1170	0.1170
100	0.0930	0.1170	0.1320	0.1398	0.1415	0.1419	0.1419	0.1419
150	0.0930	0.1170	0.1323	0.1434	0.1504	0.1547	0.1547	0.1547
200	0.0930	0.1170	0.1323	0.1434	0.1511	0.1615	0.1616	0.1616
250	0.0930	0.1170	0.1323	0.1434	0.1511	0.1658	0.1669	0.1669
500	0.0930	0.1170	0.1323	0.1434	0.1511	0.1693	0.1786	0.1789
1000	0.0930	0.1170	0.1323	0.1434	0.1511	0.1693	0.1800	**0.1833**
1500	0.0930	0.1170	0.1323	0.1434	0.1511	0.1693	0.1800	**0.1833**

Table 13: MAgP Correlation Strategy RiC 2010 T2I-Score

# of Docs	# of Elements							
	50	**100**	**150**	**200**	**250**	**500**	**1000**	**1500**
25	0.0852	0.0852	0.0852	0.0852	0.0852	0.0852	0.0852	0.0852
50	0.1043	0.1066	0.1066	0.1066	0.1066	0.1066	0.1066	0.1066
100	0.1043	0.1275	0.1298	0.1298	0.1298	0.1298	0.1298	0.1298
150	0.1043	0.1275	0.1405	0.1426	0.1426	0.1426	0.1426	0.1426
200	0.1043	0.1275	0.1405	0.1483	0.1498	0.1498	0.1498	0.1498
250	0.1043	0.1275	0.1405	0.1483	0.1535	0.1553	0.1553	0.1553
500	0.1043	0.1275	0.1405	0.1483	0.1535	0.1670	0.1675	0.1675
1000	0.1043	0.1275	0.1405	0.1483	0.1535	0.1670	0.1721	0.1723
1500	0.1043	0.1275	0.1405	0.1483	0.1535	0.1670	0.1721	**0.1733**

Observations (2010 RiC using T2I-Score)

Figure 38 compares the best MAgP score obtained using the T2I-Score for all three overlap removal strategies.

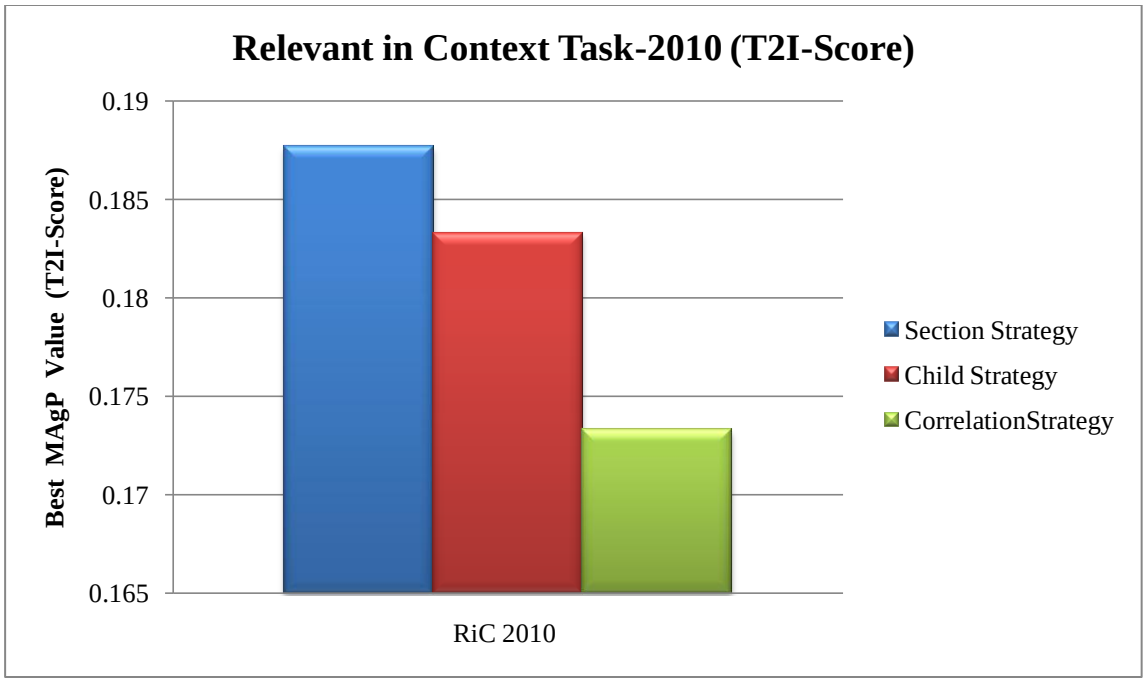

Figure 38: Comparison of all Strategies for 2010 RiC Task (T2I-Score)

1. It is observed that the section strategy performed well when compared to other two overlap removal strategies.

2. The **section strategy** has the best MAgP score of **0.1877** placing UMD in **2nd place.** Table 14 gives the MAgP value and ranking of the top ten participants [1] for comparison. There were 18 participants that submitted results for the 2010 Ad Hoc Track.

We are still awaiting data for significance testing of our results for the INEX 2010 RiC Task (using T2I-Score).

Table 14: Ranking of Top-10 2010 RiC Task (T2I-Score) Participants

Participant	MAgP Value	Rank
ENSM-SE	0.1977	1
University of Minnesota Duluth* (Section Strategy)	**0.1877**	-
University of Minnesota Duluth* (Child Strategy)	**0.1833**	-
University of Minnesota Duluth* (Child Strategy)	**0.1733**	-
Peking University	0.1615	2
LIA - University of Avignon	0.1588	3
Queensland University of Technology	0.1521	4
University of Otago	0.1436	5
Radboud University Nijmegen	0.1377	6
Renmin University of China	0.1372	7
RMIT University	0.1335	8
Doshisha University	0.1014	9
University of Amsterdam	0.0695	10

4.3 Methodology for Restricted Relevant in Context (RRiC) Task-2010

For this task, the focused elements retrieved for a document are sorted by correlation. This is normal Flex output. The overlap removal techniques are applied to this output, which is converted to INEX format and finally to TREC format. The XPaths are expanded and the TREC file is converted to FOL. The values of *n* and *m* are as before. The T2I strategy, wherein the elements having fewer than 300 irrelevant characters between them are retained and all other elements within that document are discarded, is applied to all elements.

As the RRiC task is aimed at retrieving 500 characters per article, the top-most element for each article is considered. In the FOL file, if the length of the element (last column) is greater than or equal to 500, then the length is chopped to 500 to accommodate the snippet. If the length is less than 500, then that element is considered along with the next element in the rank-ordered list of elements for that document. Hence, in every article a

group of elements whose lengths sum to 500 are chosen. For each query *m* elements are extracted from the result file and further evaluated to obtain the MAgP score.

4.3.1 Restricted Relevant in Context Task-2010 Experiments

The experiments described in this section are performed using the 2009 Wiki document collection, the 2010 query set, and each of three overlap removal technique. All results are evaluated using the INEX 2009 F-Score as well as INEX 2010 T2I-Score provided in the 2010 INEX evaluation tool. Best results are highlighted.

Experiment 4 (2010 RRiC using F-Score)

In this experiment, all the three overlap removal techniques, namely, section, child and correlation are used to obtain non-overlapping elements for the RRiC 2010 task. (See Section 3.1.2 for details). F-Score metric is used to evaluate the elements. Tables 15 - 17 give the results.

Table 15: MAgP Section Strategy RRiC 2010 F-Score

# of Docs	# of Elements							
	50	**100**	**150**	**200**	**250**	**500**	**1000**	**1500**
25	0.0686	0.0686	0.0686	0.0686	0.0686	0.0686	0.0686	0.0686
50	0.0813	0.0866	0.0866	0.0866	0.0866	0.0866	0.0866	0.0866
100	0.0813	0.1016	0.1079	0.1079	0.1079	0.1079	0.1079	0.1079
150	0.0813	0.1016	0.1138	0.1190	0.1190	0.1190	0.1190	0.1190
200	0.0813	0.1016	0.1138	0.1209	0.1249	0.1254	0.1254	0.1254
250	0.0813	0.1016	0.1138	0.1209	0.1258	0.1302	0.1302	0.1302
500	0.0813	0.1016	0.1138	0.1209	0.1258	0.1385	0.1413	0.1413
1000	0.0813	0.1016	0.1138	0.1209	0.1258	0.1385	0.1447	0.1461
1500	0.0813	0.1016	0.1138	0.1209	0.1258	0.1385	0.1447	**0.1466**

Table 16: MAgP Child Strategy RRiC 2010 F-Score

# of Docs	# of Elements							
	50	**100**	**150**	**200**	**250**	**500**	**1000**	**1500**
25	0.0727	0.0728	0.0728	0.0728	0.0728	0.0728	0.0728	0.0728
50	0.0827	0.0916	0.0916	0.0916	0.0916	0.0916	0.0916	0.0916
100	0.0827	0.1022	0.1129	0.1138	0.1138	0.1138	0.1138	0.1138
150	0.0827	0.1022	0.1150	0.1228	0.1249	0.1253	0.1253	0.1253
200	0.0827	0.1022	0.1150	0.1233	0.1288	0.1318	0.1318	0.1318
250	0.0827	0.1022	0.1150	0.1233	0.1289	0.1367	0.1367	0.1367
500	0.0827	0.1022	0.1150	0.1233	0.1289	0.1431	0.1482	0.1482
1000	0.0827	0.1022	0.1150	0.1233	0.1289	0.1431	0.1511	0.1531
1500	0.0827	0.1022	0.1150	0.1233	0.1289	0.1431	0.1511	**0.1534**

Table 17: MAgP Correlation Strategy RRiC 2010 F-Score

# of Docs	# of Elements							
	50	**100**	**150**	**200**	**250**	**500**	**1000**	**1500**
25	0.0574	0.0574	0.0574	0.0574	0.0574	0.0574	0.0574	0.0574
50	0.0725	0.0732	0.0732	0.0732	0.0732	0.0732	0.0732	0.0732
100	0.0725	0.0907	0.0911	0.0911	0.0911	0.0911	0.0911	0.0911
150	0.0725	0.0907	0.1009	0.1014	0.1014	0.1014	0.1014	0.1014
200	0.0725	0.0907	0.1009	0.1072	0.1075	0.1075	0.1075	0.1075
250	0.0725	0.0907	0.1009	0.1072	0.1117	0.1121	0.1121	0.1121
500	0.0725	0.0907	0.1009	0.1072	0.1117	0.1222	0.1225	0.1225
1000	0.0725	0.0907	0.1009	0.1072	0.1117	0.1222	0.1267	0.1267
1500	0.0725	0.0907	0.1009	0.1072	0.1117	0.1222	0.1267	**0.1277**

Observations (2010 RRiC using F-Score)

The observations made for the results of the 2010 RRiC task using the F-Score are discussed below. Figure 39 compares the best MAgP scores obtained using the F-Score for the three overlap removal strategies.

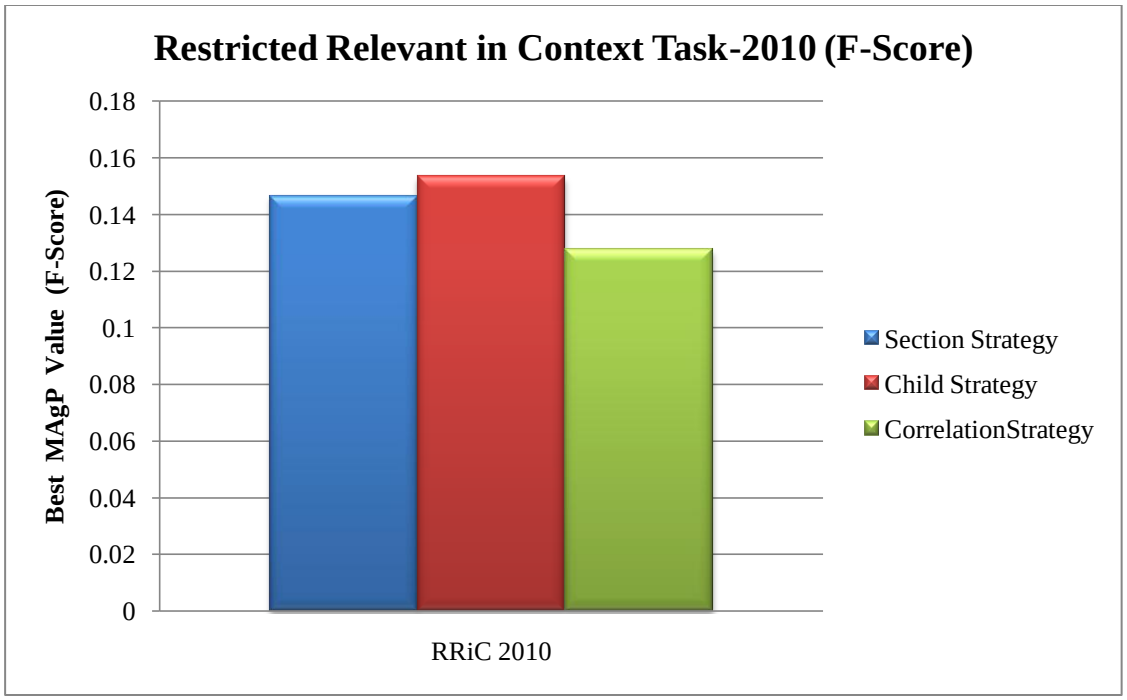

Figure 39: Comparison of all Strategies for 2010 RRiC Task (F-Score)

1. It is observed that the child strategy performed well when compared to other two overlap removal strategies.

2. The **child strategy** has the best MAgP score of **0.1534** placing UMD in **1st place.** Table 18 gives the MAgP value and ranking of the top ten participants [1] for comparison. There were 18 participants that submitted results for the 2010 Ad Hoc Track.

We are still awaiting data for significance testing of our results for the INEX 2010 RRiC Task (using F-Score)

Table 18: Ranking of Top-10 2010 RRiC Task (F-Score) Participants

Participant	MAgP Value	Rank
University of Minnesota Duluth* (Child Strategy)	**0.1534**	-
University of Minnesota Duluth* (Section Strategy)	**0.1466**	-
University of Minnesota Duluth* (Correlation Strategy)	**0.1277**	-
Queensland University of Technology	0.1064	1
LIA - University of Avignon	0.1053	2
Peking University	0.1030	3
University of Otago	0.0953	4
Radboud University Nijmegen	0.0945	5
Doshisha University	0.0537	6
University of Waterloo	0.0497	7
University of Amsterdam	0.0462	8
Indian Statistical Institute	0.0327	9

Experiment 5 (2010 RRiC using T2I-Score)

In this experiment, all three overlap removal techniques are used to obtain non-overlapping elements for the RRiC 2010 task. (See Section 3.1.2 for details). The T2I-Score metric is used for evaluation. Tables 19 - 21 give the results of this experiment.

Table 19: MAgP Section Strategy RRiC 2010 T2I-Score

# of	# of Elements							
Docs	**50**	**100**	**150**	**200**	**250**	**500**	**1000**	**1500**
25	0.0898	0.0898	0.0898	0.0898	0.0898	0.0898	0.0898	0.0898
50	0.1051	0.1112	0.1112	0.1112	0.1112	0.1112	0.1112	0.1112
100	0.1051	0.1285	0.1352	0.1352	0.1352	0.1352	0.1352	0.1352
150	0.1051	0.1285	0.1420	0.1474	0.1474	0.1474	0.1474	0.1474
200	0.1051	0.1285	0.1420	0.1496	0.1537	0.1541	0.1541	0.1541
250	0.1051	0.1285	0.1420	0.1496	0.1547	0.1593	0.1593	0.1593
500	0.1051	0.1285	0.1420	0.1496	0.1547	0.1682	0.1708	0.1708
1000	0.1051	0.1285	0.1420	0.1496	0.1547	0.1682	0.1743	0.1757
1500	0.1051	0.1285	0.1420	0.1496	0.1547	0.1682	0.1743	**0.1762**

Table 20: MAgP Child Strategy RRiC 2010 T2I-Score

# of Docs	# of Elements							
	50	**100**	**150**	**200**	**250**	**500**	**1000**	**1500**
25	0.0909	0.0910	0.0910	0.0910	0.0910	0.0910	0.0910	0.0910
50	0.1027	0.1128	0.1128	0.1128	0.1128	0.1128	0.1128	0.1128
100	0.1027	0.1249	0.1362	0.1370	0.1370	0.1370	0.1370	0.1370
150	0.1027	0.1249	0.1388	0.1471	0.1491	0.1494	0.1494	0.1494
200	0.1027	0.1249	0.1388	0.1477	0.1533	0.1561	0.1561	0.1561
250	0.1027	0.1249	0.1388	0.1477	0.1534	0.1613	0.1613	0.1613
500	0.1027	0.1249	0.1388	0.1477	0.1534	0.1682	0.1730	0.1730
1000	0.1027	0.1249	0.1388	0.1477	0.1534	0.1682	0.1760	0.1779
1500	0.1027	0.1249	0.1388	0.1477	0.1534	0.1682	0.1760	**0.1782**

Table 21: MAgP Correlation Strategy RRiC 2010 T2I-Score

# of Docs	# of Elements							
	50	**100**	**150**	**200**	**250**	**500**	**1000**	**1500**
25	0.0811	0.0811	0.0811	0.0811	0.0811	0.0811	0.0811	0.0811
50	0.1005	0.1012	0.1012	0.1012	0.1012	0.1012	0.1012	0.1012
100	0.1005	0.1223	0.1228	0.1228	0.1228	0.1228	0.1228	0.1228
150	0.1005	0.1223	0.1342	0.1346	0.1346	0.1346	0.1346	0.1346
200	0.1005	0.1223	0.1342	0.1410	0.1413	0.1413	0.1413	0.1413
250	0.1005	0.1223	0.1342	0.1410	0.1459	0.1464	0.1464	0.1464
500	0.1005	0.1223	0.1342	0.1410	0.1459	0.1574	0.1577	0.1577
1000	0.1005	0.1223	0.1342	0.1410	0.1459	0.1574	0.1620	0.1621
1500	0.1005	0.1223	0.1342	0.1410	0.1459	0.1574	0.1620	**0.1631**

Observations (2010 RRiC using T2I-Score)

Observations regarding results of the RRiC 2010 task using the T2I-Score are discussed below. Figure 40 shows a comparison of the best MAgP scores obtained using T2I-Score for all three overlap removal strategies.

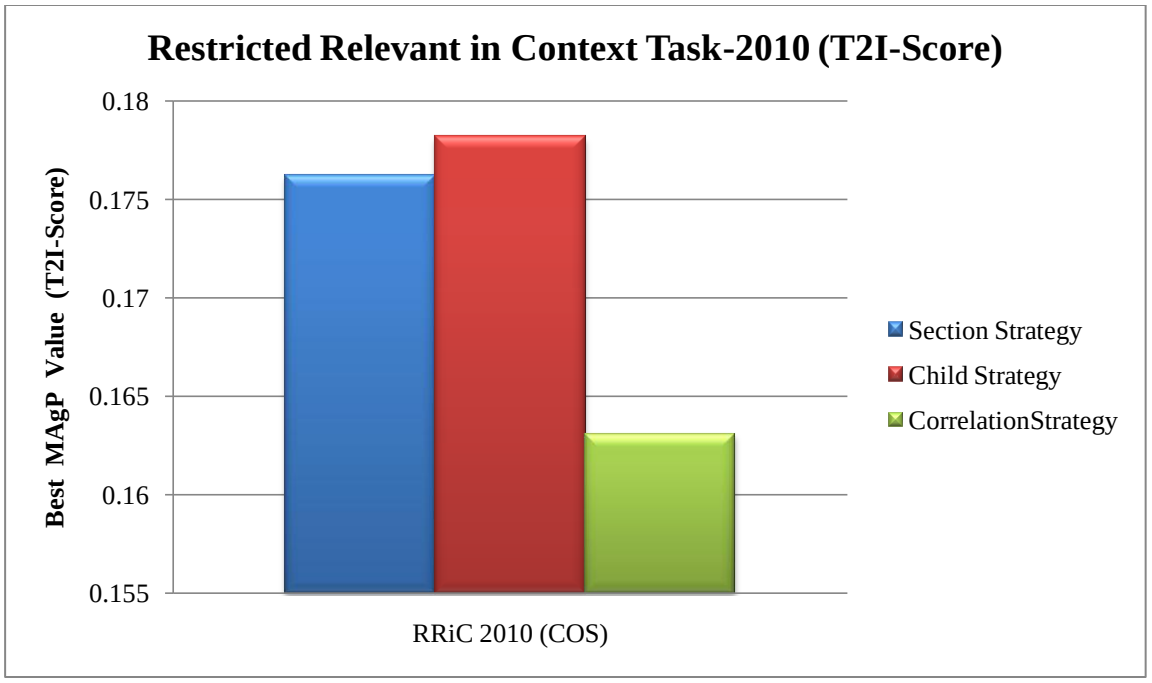

Figure 40: Comparison of all Strategies for 2010 RRiC Task (T2I-Score)

1. It is observed that the child strategy performed well when compared to other two overlap removal strategies. The MAgP score obtained using the section strategy is almost the same as the MAgP score produced using the child strategy.

2. The **child strategy** has the best MAgP score of **0.1782** placing UMD in **1st place.** Table 22 gives the MAgP value and rankings of the top ten participants [1] for comparison. There were 18 participants that submitted results for the 2010 Ad Hoc Track.

We are still awaiting data for significance testing of our results for the INEX 2010 RRiC Task (using T2I-Score)

Table 22: Ranking of Top-10 2010 RRiC Task (T2I-Score) Participants

Participant	MAgP Value	Rank
University of Minnesota Duluth* (Child Strategy)	**0.1782**	**-**
University of Minnesota Duluth* (Section Strategy)	**0.1762**	-
University of Minnesota Duluth* (Correlation Strategy)	**0.1632**	-
Peking University	0.1580	1
LIA - University of Avignon	0.1541	2
Queensland University of Technology	0.1508	3
University of Otago	0.1436	4
Radboud University Nijmegen	0.1375	5
University of Waterloo	0.0650	6
Doshisha University	0.0600	7
University of Amsterdam	0.0576	8
Indian Statistical Institute	0.0485	9

Table 23 shows the rank of UMD for the different RiC tasks.

Table 23: Scores and Ranks for All Tasks

Task	MAgP Score	Rank
2009 RiC Task	0.1889	1
2010 RiC Task (F-Score)	0.1838	2
2010 RiC Task (T2I-Score)	0.1877	2
2010 RRiC Task (F-Score)	0.1535	1
2010 RRiC Task (T2I-Score)	0.1782	1

*** - Current Runs**

**** - Previous Runs**

5. Conclusions and Future Work

Results have improved significantly over the past year largely, due to the improved parsing techniques. The experiments show that the three overlap removal techniques behave differently depending on the task.

The correlation strategy performs well for the 2009 RiC task. The section strategy performs well for the 2010 RiC task for both F-Score and T2I-Score evaluation metrics. From this we infer that returning the highly correlated element and all the elements around it (such that fewer than 300 irrelevant characters are read) works well. The child strategy performed well for the 2010 RRiC task (for both F-Score and T2I-Score). In this case, returning 500 characters from a highly correlated element or 500 characters from top correlated elements produces a good result.

Due to a parsing error, we found that a small set of documents from the reference runs were missing from our document set. Results may be improved if these documents are included. The present parser considers *image, categories,* and *header* as terminal nodes. Most of these are considered invalid nodes by the INEX evaluation tool. Recognizing these nodes as invalid so that they can be automatically removed from the result may also improve the results.

More research in the area of snippet retrieval may also improve results. Investigation into returning a subset of the element (such as a sentence or two) rather than entire element may be feasible.

References

[1] Arvola P., Geva S., Kamps J., Trotman A. Overview of the INEX 2010 Ad Hoc Track http://www.cs.otago.ac.nz/homepages/andrew/2010-13.pdf

[2] Arvola P., Keka”la”inen J., Junkkari M. Expected Reading Effort in Focused retrieval evaluation, *Information Retrieval*, Volume 13, Number 5, 460-484, May 2010

[3] Banhatti, R. Improving Results for the INEX 2009 Thorough and 2010 Efficiency Tasks, Department of Computer Science, University of Minnesota Duluth, Aug 2011. http://www.d.umn.edu/cs/thesis/banhatti.pdf

[4] Crouch, C. Dynamic element retrieval in structured environment. *ACM TOIS*, 24(4): 437-454, 2006.

[5] Deepak Acquilla, N. Improving Results for the INEX 2009 and 2010 Focused Tasks, Department of Computer Science, University of Minnesota Duluth, Aug 2011. http://www.d.umn.edu/cs/thesis/deepakacquilla.pdf

[6] Geva S., Kamps J., Trotman A. About INEX https://inex.mmci.uni-saarland.de/about.html

[7] Geva S., Kamps J., Trotman A. INEX 2009 Guidelines for Topic Development http://www.inex.otago.ac.nz/tracks/adhoc/gtd.asp

[8] Geva S., Kamps J., Trotman A. INEX Document Collection http://www.inex.otago.ac.nz/data/documentcollection.asp

[9] Geva S., Kamps J., Trotman A. Overview of the INEX 2009 Ad Hoc Track http://www.cs.otago.ac.nz/homepages/andrew/2009-13.pdf

[10] Poluri, P. Focused Retrieval using Exact Methodology, Department of Computer Science, University of Minnesota Duluth, Aug 2009. http://www.d.umn.edu/cs/thesis/poluri.pdf

[11] Salton, G., ed. *The Smart Retrieval System - Experiments in Automatic Document Processing*. Prentice-Hall,1971.

[12] Salton, G., Wong, A., Yang, C. A Vector Space Model for Automatic Indexing, *Comm. ACM*, 18(11), 613-620, 1975.

[13] Singhal, A., Buckley, C., Mitra, M. Pivoted Document Length Normalization, *Proceedings Of the 19th Annual International ACM SIGIR Conference,* Zurich, Switzerland, 19-21, 1996.