Enriching Textbooks with Images

Rakesh Agrawal Sreenivas Gollapudi Anitha Kannan Krishnaram Kenthapadi
Search Labs, Microsoft Research
Mountain View, CA, USA
{rakesha, sreenig, ankannan, krisken}@microsoft.com

ABSTRACT

Textbooks have a direct bearing on the quality of education imparted to the students. Therefore, it is of paramount importance that the educational content of textbooks should provide rich learning experience to the students. Recent studies on understanding learning behavior suggest that the incorporation of digital visual material can greatly enhance learning. However, textbooks used in many developing regions are largely text-oriented and lack good visual material. We propose techniques for finding images from the web that are most relevant for augmenting a section of the textbook, while respecting the constraint that the same image is not repeated in different sections of the same chapter. We devise a rigorous formulation of the image assignment problem and present a polynomial time algorithm for solving the problem optimally. We also present two image mining algorithms that utilize orthogonal signals and hence obtain different sets of relevant images. Finally, we provide an ensembling algorithm for combining the assignments. To empirically evaluate our techniques, we use a corpus of high school textbooks in use in India. Our user study utilizing the Amazon Mechanical Turk platform indicates that the proposed techniques are able to obtain images that can help increase the understanding of the textbook material.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

General Terms

Algorithms, Experimentation, Human Factors

Keywords

Education, Textbooks, Data Mining, Image Mining, Text Augmentation

1. INTRODUCTION

Education is the key for improving the economic well-being of people living in developing regions [1, 24]. While the problem of providing high-quality education is multi-faceted and complex [8, 50], textbooks have been shown to provide the most cost-effective means of improving the educational quality [11, 16, 26]. They are the primary conduits for delivering content knowledge to the students and the teachers base their lesson plans primarily on the material given in textbooks [18, 55]. Unfortunately, textbooks in many developing countries often suffer from the lack of clarity as well as the inadequacy of information provided [2, 44].

It has been shown that the linking of encyclopedic information to educational material can improve both the quality of the knowledge acquired and the time needed to obtain such knowledge [12]. In [4], a decision model has been proposed for diagnosing deficient sections of a textbook. In [5, 40, 41, 43], techniques have been proposed for finding articles from the web with which such sections should be augmented.

We extend prior work and propose techniques for augmenting textbooks with links to selective, highly relevant images. Our work has been instigated by the research in the learning literature showing that the use of visual materials enhances learning, not only by enabling retention of information but also by promoting comprehension and transfer [35, 39, 48]. Some researchers even suggest that visual materials should be given as much attention as text and not viewed as merely supporting the text [36].

The learning literature informs us that any supplementary material is most effective when it is presented in close proximity to the main material [9]. We therefore find images that are relevant to a particular section of the textbook. In order to avoid undue cognitive burden, we select only a small number of images that collectively best aid the understanding, shunning repetition of an image in different sections of the same chapter.

Our solution has three components:

1. *Image Assignment*. Given a set of candidate images and their relevance scores for every section of a chapter, this component assigns images to various sections in a way that maximizes the relevance score for the chapter while maintaining the constraints that no section has been assigned more than a certain maximum number of images and no image is used more than once in the chapter. We provide a polynomial time algorithm for implementing the optimizer.

2. *Image Mining*. This component comprises of algorithms that mine the web for images relevant to a particular section and provide a ranked list of top k images along with their relevance scores. It is preferable to have algorithms that make use of orthogonal signals in their search for images in order to have a broad selection of images to choose from. We provide two specific algorithms, namely AFFINITY and COMITY, which satisfy the above properties.

CIKM'11, October 24–28, 2011, Glasgow, Scotland, UK.

3. *Image Ensembling*. Since the relevance scores provided by different image mining algorithms will in general be incomparable, the assignment of images to different sections of a chapter needs to be performed separately for each algorithm. The image ensembling component aggregates the ranked lists of image assignments to produce the final result.

   While it is possible to use any rank aggregation algorithm, we wanted a voting scheme that considers all the elements of a ranked list and provides consensus ranking. The popular Borda's method fits the bill [52]. Ensembling is done sequentially within a chapter, starting from the first section. Top images selected for a section are eliminated from the pool of available images for the remaining sections. The image assignment is then rerun, followed by an ensembling for the next section.

In order to assess the usefulness of our proposal, we conducted a user study employing the Amazon Mechanical Turk platform [29]. The corpus used in this evaluation comprises of the high school textbooks published by the Indian National Council of Educational Research and Training (NCERT). We wanted to use a corpus from a developing country and settled on the NCERT books because of their ready availability online. However, our techniques are not country-dependent. We consider fourteen textbooks from grades IX to XII, covering eight subjects, seventy chapters, and over a thousand sections. The results of this study indicate that the proposed techniques are able to discover images that can help increase the understanding of the textbook material.

Scope of the paper. We present techniques for proposing images with which a section of a textbook can be augmented, but do not discuss specific mechanisms for integrating the augmentations into the textbook. Our techniques could be integrated into authoring tools for helping textbook authors decide what images to use when writing or revising a book. They can also be used for creating supplementary material that is distributed with the paper version of the books. Furthermore, there are ongoing efforts aimed at creating platforms and inexpensive devices for distributing books in a digital form (see, for example, the use of interactive DVDs as an educational platform [17], inexpensive e-book readers [3, 7], and mobile learning devices [31]). Our work fits quite naturally with these efforts.

Complementary approaches and issues that merit serious investigation, but are beyond the scope of this paper include: (a) refining and enhancing the results produced by our techniques using collaboration and crowdsourcing [2, 53], (b) implications for royalty sharing and intellectual property rights [14, 20], and (c) integration with other interventions for improving the learning outcomes [19, 45].

Organization of the paper. The rest of the paper is structured as follows. We discuss related work in §2. The particular algorithms used in our solution are described in §3. The results of experimental evaluation are presented in §4. We conclude with a summary and directions for future work in §5.

2. RELATED WORK

There has been research on translating short textual sentences to pictures [42]. The work in [10] uses semantic representation of the text to synthesize three-dimensional scenes from English text. Such translations work well only with descriptive sentences that can be effectively mapped to corresponding object models. Our focus is on finding a few representative images relevant for a section of a textbook, not on synthesizing scenes from sentences, which leads to different techniques.

The goal of research in story picturing is to find a set of pictures to describe a fragment of a text. In [32], key phrases are extracted from the text to retrieve matching images, which are then ranked by weighting image region similarity and phrase overlaps. In [59], the picturable key phrases are extracted from a given text; the picturability of a phrase is measured as the ratio of the frequencies under regular web vs. image search. Each picturable phrase is then associated with an image using image search coupled with computer vision and finally graphics is used to spatially arrange the images to represent the story. In [15], the meaning of images and their surrounding texts are represented jointly as probability distributions over a set of latent topics, which are then used to find pictures to illustrate a story. While close in spirit, the setting and the solution techniques in these works and ours are quite different.

The rich literature on image retrieval in response to key word queries is relevant for the image mining component of our work. Indeed, one of our mining algorithms (COMITY) issues queries to the commercial search engines as a subroutine. We refer the reader to [13] for a comprehensive survey of recent advances in image retrieval. While there is ground breaking research on retrieving images solely based on the content of the images, most image search engines work by indexing associated metadata and matching keyword queries with the stored metadata. The metadata may include a description of the image, the filename of the image, the anchor text pointing to the image, the queries that led to a click on the image, or the text adjacent to the image. The ImageCLEF competition, for example, uses an image repository consisting of Wikipedia images that have been tagged with the associated title and description of the image extracted from the Wikipedia page for the image [46].

Several web based applications (e.g. Flickr) provide facilities for any web user to tag pictures hosted on their websites with textual descriptions that can serve as metadata for the image. Lui von Ahn developed an interactive ESP game in which the players labeled images while playing the game [56]. There is rich body of image processing research on automatically linking text to image. The survey in [6] distinguishes the task of annotation from that of correspondence. The former attempts to predict the annotation of an entire image using all information present, whereas the latter attempts to associate particular words with particular image substructures. There is also work on extracting image regions containing text that can then be fed to an optical character recognition module [30, 38].

Related work also includes the proposal in [2] to create an education network to harness the collective efforts of educators, parents, and students to collaboratively enhance the quality of educational material. Some websites (e.g. Notemonk.con) allow students to download textbooks, ask questions on a topic, and annotate books for quick reference. Several institutions are making the videos of the course lectures available through Internet and there are websites (e.g. EducationPortal.com) that aggregate links to them. We view these crowdsourcing efforts as complementary approaches to improving the quality of textbooks.

3. SYSTEM DESIGN

We now describe the design of our system. We first present the image assignment component consisting of an optimization problem, followed by two algorithms for mining relevant images from the web, and finally the image ensembling component.

3.1 Image Assignment

Given a set of candidate images relevant to the various sections of a chapter and their relevance scores, the goal of the image assignment component is to allocate to each section the most relevant images, while respecting the constraints that each section is not augmented with too many images and that each image is used no more than once in a chapter. The rationale for these constraints is that an augmentation of a section with too many images will put undue cognitive burden on the reader while the repetition of an image across sections in the same chapter would be redundant for the reader.

First, a few notations. Let $I = \{1, 2, \ldots, n\}$ denote the set of images and $S = \{1, 2, \ldots, m\}$ denote the set of sections in a chapter. Let λ_{ij} denote the (non-negative) relevance score of image $i \in I$ for section $j \in S$ ($\lambda_{ij} = 0$ if the image i is not present in the candidate set of images for section j). Let K_j denote the maximum number of images that can be associated with section j. K_j could be either a fixed integer for all sections or a function of the length of the section j.

This problem admits a natural greedy algorithm. Sort the λ_{ij} values in decreasing order and go through them. At each step, the greedy algorithm picks the highest λ_{ij} value such that an image can still be assigned to section j (that is, less than K_j images have so far been assigned to j) and then assigns image i to section j. This process ends when either all sections have been assigned the maximum number of images or there are no more images to be assigned.

At a first glance, the greedy algorithm might seem optimal in terms of the sum of relevance scores of all assigned images. But the following counterexample shows that the optimal value can be substantially larger. Consider a chapter consisting of two sections and suppose that we want two images each for a section ($K_1 = K_2 = 2$). Represent by (i, λ) that image i's relevance score is λ. Let the top images and their relevance scores obtained by an image mining algorithm for various sections be as follows: $s_1 \leftarrow \langle (i_1, 1), (i_2, 1-\epsilon), (i_3, 1-3\epsilon) \rangle, s_2 \leftarrow \langle (i_2, 1-2\epsilon), (i_4, \epsilon), (i_5, \epsilon) \rangle$, where $\epsilon = 0.01$. Then the greedy assignment would be $s_1 \leftarrow \langle i_1, i_2 \rangle, s_2 \leftarrow \langle i_4, i_5 \rangle$ with a total score of $2 + \epsilon$. On the other hand, an optimal assignment is $s_1 \leftarrow \langle i_1, i_3 \rangle, s_2 \leftarrow \langle i_2, i_4 \rangle$ with a total score of $3 - 4\epsilon$.

Therefore, we instantiate the image assignment component as an optimization problem. We show that this optimization problem can be solved optimally in polynomial time and provide an efficient algorithm as part of the proof. The following is the statement of the optimization problem:

MAXRELEVANTIMAGEASSIGNMENT

$$\max \sum_{i \in I} \sum_{j \in S} x_{ij} \cdot \lambda_{ij} \tag{1}$$

s.t.

$$x_{ij} \in \{0, 1\} \;\; \forall i \in I \, \forall j \in S \tag{2}$$

$$\sum_{i \in I} x_{ij} \leq K_j \;\; \forall j \in S \tag{3}$$

$$\sum_{j \in S} x_{ij} \leq 1 \;\; \forall i \in I \tag{4}$$

Here, x_{ij} is an indicator variable that takes value 1 if image i is selected for section j and 0 otherwise. Eq. 2 captures this binary constraint. Eq. 3 ensures that the number of images assigned to a section is at most K_j. Eq. 4 enforces that each image is assigned to at most one section in a chapter. The optimization objective (Eq. 1) is the total relevance score for the chapter, defined as the sum over all sections of relevance scores of the images assigned to the section. Thus the goal of the optimization is to compute the binary variables x_{ij} such that the total relevance score for the chapter is maximized.

Algorithm 1 AFFINITY

Input: A textbook section j; Number of desired image results k; Number of desired closest articles from an authoritative external source t'; Number of desired concept phrases c.

Output: A list of top k image results from the authoritative source, along with value scores.

```
Obtain (up to) top $c$ concept phrases from section $j$.
Obtain (up to) top $t'$ closest articles from the authoritative external source, based on content similarity with section $j$.
Extract the set of images present in these $t'$ articles, as well as the metadata associated with each image aggregated over all occurrences of the image.
For each image $i$, let $n_{ij}$ := Number of articles in which image $i$ appears, $d_{ij}$ := Number of concept phrases contained in the metadata for image $i$, $w_{ij}$ := Number of matching words from all concepts in the metadata for image $i$.
Assign the relevance score $\lambda_{ij} := n_{ij}^{w_1} \cdot d_{ij}^{w_2} \cdot w_{ij}^{w_3}$ for image $i$ ($w_1, w_2, w_3$ determine the relative weight given to the three counts above).
Return top $k$ images along with their $\lambda_{ij}$ values.
```

THEOREM 3.1. MAXRELEVANTIMAGEASSIGNMENT *can be solved optimally in polynomial time.*

PROOF. The proof follows by showing an efficient reduction from MAXRELEVANTIMAGEASSIGNMENT to the MAXIMUM WEIGHTED BIPARTITE MATCHING problem [49], which admits an efficient polynomial time solution. Given an instance of MAXRELEVANTIMAGEASSIGNMENT, form a complete weighted bipartite graph $G = (V, E)$ as follows. Associate a node u_i with each image $i \in I$ and associate K_j nodes, $v_{j1}, v_{j2}, \ldots, v_{jK_j}$, with each section j. Create an edge between every image node and every section node copy. Weight of the edge (u_i, v_{jk}) is set to λ_{ij} for each $k \in \{1, 2, \ldots, K_j\}$, that is, each of the K_j edges joining an image i to the section j has the same weight, equal to the corresponding relevance score.

We observe that any feasible solution to MAXRELEVANTIMAGEASSIGNMENT corresponds to selecting a matching in G. Given a satisfying assignment of x_{ij}'s, we can obtain a matching in G by picking one of the K_j edges corresponding to any x_{ij} that is set to 1. Similarly, given any matching in G, there is a corresponding feasible solution. Further the objective of MAXRELEVANTIMAGEASSIGNMENT can be maximized by obtaining the maximum weight bipartite matching in G. As the MAXIMUM WEIGHTED BIPARTITE MATCHING problem can be solved optimally in $O(nm(n+m))$ time, it follows that MAXRELEVANTIMAGEASSIGNMENT can also be solved optimally in $O(nm(n+m))$ time. □

3.2 Image Mining

Here we give particulars of the AFFINITY and COMITY algorithms, the two algorithms used for obtaining the ranked list of top k images along with their relevance scores for a given section. Note that our system design admits various possible variants of these algorithms as well as additional image mining algorithms one could conceive.

Algorithm AFFINITY

The intuition behind this algorithm is the observation that the images included in an authoritative article relevant to a topic are often illustrative of the key concepts underlying the topic. We therefore find authoritative articles whose contents have high textual similarity with a given section of the book. We then extract images contained in these articles and use their relevance scores to find top k images for the section.

Algorithm 1 (AFFINITY) first obtains the key concept phrases present in a section as well as the closest authoritative articles from the web. Thus the key topics discussed in the section are available in the form of the concept phrases while the search space for images is refined to the set of articles with high document similarity to the section. The relevance score for an image is computed by analyzing the overlap between the concept phrases and the cumulative metadata associated with the various copies of the image present in the narrowed set of articles. The metadata for an image comprises of text adjacent to the image including caption and alternative text, filename of the image, anchor texts pointing to the image, and queries that led to clicks on the image. The scoring has desirable properties such as: (a) an image occurring in multiple articles gets a higher score, (b) an image whose metadata contains multiple concept phrases gets a higher score, and (c) an image whose metadata contains words from many concepts gets a higher score.

In the rest of this subsection, we provide implementation details of steps 1 and 2. If a textbook includes a back-of-the-book index [47], it can be used for obtaining concept phrases for a section. There is also rich literature on algorithmically extracting key phrases from a document that can inform the task of extracting key concepts from a section [33]. The current approaches primarily involve detection of the key phrases based on rules or statistical and learning methods. In the former, the structural properties of phrases form the basis for the rule generation. In the latter, the importance of a phrase is computed based on statistical properties (e.g., relative frequency, document frequency) of the phrase.

Our study of several textbooks from the developing regions revealed that they often do not include a back-of-the-book index, but the concept phrases typically consist of terminological noun phrases containing adjectives, nouns, and sometimes prepositions. It is uncommon for concepts to contain other parts of speech such as adverbs, conjunctions, or verbs. Following [34], we adopt the linguistic pattern A^*N^+, where A refers to an adjective and N a noun and use the algorithm from [5] to determine the top c concepts. Examples of concepts satisfying this pattern include “cumulative distribution function”, “fiscal policy”, and “electromagnetic radiation”.

Early research on finding documents similar to a given document includes techniques based on relevance feedback [51], which were subsequently used to show similar results by some of the search engines [54]. Recent research includes the query by document work [58] that extracts key phrases and uses them as queries to search for similar documents (see also references therein). We index the corpus of authoritative articles using Lucene [25] and use its in-built scoring to retrieve the closest articles to a given textbook section.

Algorithm COMITY

One might think that one could simply use the text string of a section to query a commercial image search engine and obtain the relevant images. However, the current search engines do not perform well with long queries [28, 37]. Indeed, when we queried the search engines using even the first paragraph of a section, we got none or meaningless results. In one major stream of research on information retrieval with long queries, the focus is on selecting a subset of the query, while in another it is on weighting the terms of the query [57]. This body of research however is not designed to work for queries consisting of arbitrary textbook sections.

Algorithm 2 (COMITY) is based on using the key concepts present in a section to query the commercial image search engines. However, each concept phrase in isolation may not be representative of the section as a typical book section can discuss multiple concepts. Hence we form $\binom{c}{2} + \binom{c}{3}$ image search queries by combining two and three concept phrases each, in order to provide more context about the section. A relevant image for the section is likely to occur among the top results for many such queries. Thus, by aggregating the image result lists over all the combination queries, we end up boosting the relevance scores of very relevant images for the section. We further increase the coverage by obtaining and merging results across e different search engines. We treat each search engine as a blackbox [23, 28], that is, we have access to the ranking of results but do not have access to the internals of the search engine such as the score given to a document with respect to a query.

Algorithm 2 COMITY

Input: A textbook section j; Number of desired image results k; Number of desired image search results per query t; Number of desired concept phrases c.

Output: A list of top k image results from web, along with relevance scores.

```
Obtain (up to) top $c$ concept phrases from section $j$.
Form queries consisting of two and three concepts phrases each ($\binom{c}{2} + \binom{c}{3}$ queries in total).
Obtain (up to) top $t$ image search results for each of the queries from $e$ different search engines.
Aggregate over (potentially $e(\binom{c}{2} + \binom{c}{3})$) lists of images, to obtain $\lambda_{ij}$ values for each image.
Return top $k$ images along with their $\lambda_{ij}$ values.
```

Aggregation across multiple lists is performed as follows. Each of (up to) t images in a result list is assigned a position-discounted score equal to $1/(p + \theta)$ where p denotes the position and θ is a smoothing parameter. For the same image occurring in multiple lists, the scores are added, weighted by a function f of the importance of the concept phrase present in the underlying query: $\lambda_{ij} := \sum_q f(\text{Importance scores of concept phrases used in } q) \times (1/(p(i, q, R(q)) + \theta))$. Here the summation is over $e(\binom{c}{2} + \binom{c}{3})$ queries issued and $p(i, q, R(q))$ denotes the position of image i in the result list $R(q)$ for query q if i is present in $R(q)$ and ∞ otherwise. This choice is based on our empirical observation that an image occurring among the top results for multiple queries was more relevant to the section than an image that occurred among the top results for only one query.

3.3 Image Ensembling

We next describe our ensembling algorithm for combining the different image assignments. Since the relevance scores computed by the image mining algorithms will be incomparable in general, we combine the results *after* the MAXRELEVANTIMAGEASSIGNMENT optimization has been performed independently for each algorithm. We use only the ordering returned by these algorithms and do rank aggregation without considering the magnitudes of the scores.

We employ Borda’s method to merge l ranked lists corresponding to l different image mining algorithms. Borda’s method tries to achieve a consensus ranking and satisfies certain desirable properties such as reversal symmetry [52]. It assigns a score corresponding to the positions in which an image appears within each ranked list of preferences, and the images are sorted by their total score.

However, a consequence of performing rank aggregation for each section independently is that the same image may appear more than once in a chapter. Consider a chapter consisting of two sections

Algorithm 3 ENSEMBLE

Input: Set of sections $S = \{1, 2, \ldots, m\}$ in a textbook chapter; Set of images $I = \{1, 2, \ldots, n\}$; Number of desired images K_j for each section $j \in S$; Scores assigned by l different image mining algorithms for each image $i \in I$; Orderings produced after the optimization for these l algorithms.

Output: A new assignment of images to sections.

```
Let $I_0 := I$ and $S_0 := S$. For each of $l$ image mining algorithms, perform MAXRELEVANTIMAGEASSIGNMENT optimization over $I$ and $S$ to get an assignment of images for all sections in $S$.
for section $j = 1$ to $m$ do
    Merge $l$ ranked lists (corresponding to $l$ algorithms) for section $j$ using Borda's method, and assign the top $K_j$ images from the merged list to section $j$. Let $A_j$ denote the set of assigned images.
    Remove the assigned images from consideration for subsequent sections, that is, $I_j := I_{j-1} \setminus A_j$ and $S_j := S_{j-1} \setminus \{j\}$.
    For each of $l$ image mining algorithms, perform MAXRELEVANTIMAGEASSIGNMENT optimization over $I_j$ as the set of images and $S_j$ as the set of sections, and thereby obtain the new assignment of images for sections $j + 1$ through $m$.
```

and suppose that we want two images for every section. Assume that the optimal assignments (ranked lists) corresponding to the two image mining algorithms are as follows. Alg_1(OPT): $s_1 \leftarrow \langle i_1, i_2 \rangle, s_2 \leftarrow \langle i_3, i_4 \rangle$ (that is, image i_1 has the highest relevance score and i_2 has the second highest score for section s_1 and similarly $\langle i_3, i_4 \rangle$ in that order are the top two images for section s_2), and Alg_2(OPT): $s_1 \leftarrow \langle i_3, i_4 \rangle, s_2 \leftarrow \langle i_1, i_2 \rangle$. Then the rank aggregation would give: $s_1 \leftarrow \langle i_1, i_3 \rangle, s_2 \leftarrow \langle i_1, i_3 \rangle$.

Algorithm 3 (ENSEMBLE) avoids this problem by taking advantage of the logical linear organization of sections within a chapter. It considers sections in a chapter sequentially from the first section to the last, ensembling at a section level, and then removing images selected for this section from the pool of available images for the remaining sections. Before moving to a subsequent section, it reruns the image assignment optimization for the remaining sections over the remaining images. Thus images discarded due to merging for a section are taken into account for consideration in subsequent sections as such images may be more relevant than any of the candidate images for a section.

Consider a chapter consisting of three sections and suppose that we want two images for every section. Assume that the images and their relevance scores for different sections found by the two image mining algorithms are as follows. Alg_1: $s_1 \leftarrow \langle (i_1, 1), (i_2, 0.9) \rangle$, $s_2 \leftarrow \langle (i_7, 0.7), (i_8, 0.6) \rangle, s_3 \leftarrow \langle (i_2, 0.5), (i_3, 0.4), (i_5, 0.3) \rangle$, and Alg_2: $s_1 \leftarrow \langle (i_3, 1), (i_4, 0.9) \rangle, s_2 \leftarrow \langle (i_7, 0.6), (i_8, 0.4) \rangle$, $s_3 \leftarrow \langle (i_4, 0.5), (i_1, 0.4), (i_6, 0.3) \rangle$. The optimal assignments would be: Alg_1(OPT): $s_1 \leftarrow \{i_1, i_2\}, s_2 \leftarrow \{i_7, i_8\}, s_3 \leftarrow \{i_3, i_5\}$, and Alg_2(OPT): $s_1 \leftarrow \{i_3, i_4\}, s_2 \leftarrow \{i_7, i_8\}, s_3 \leftarrow \{i_1, i_6\}$. The rank aggregation for the first section would give: $s_1 \leftarrow \{i_1, i_3\}$, thereby dropping i_2 from Alg_1 and i_4 from Alg_2 respectively. We note that i_2 is more relevant than current assignments for section s_3 under Alg_1 and similarly, i_4 is more relevant than current assignments for section s_3 under Alg_2. The benefit of rerunning the optimization is that such dropped images can be assigned to later sections (s_3 in our example). ENSEMBLE would result in the final assignment: $s_1 \leftarrow \{i_1, i_3\}, s_2 \leftarrow \{i_7, i_8\}, s_3 \leftarrow \{i_2, i_4\}$, which is more desirable than an assignment that excludes assigned images from later sections but does not rerun optimization ($s_3 \leftarrow \{i_5, i_6\}$).

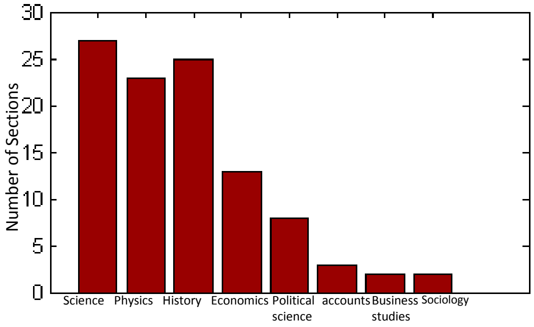


Figure 1: Subject-wise distribution of sections

4. EXPERIMENTS

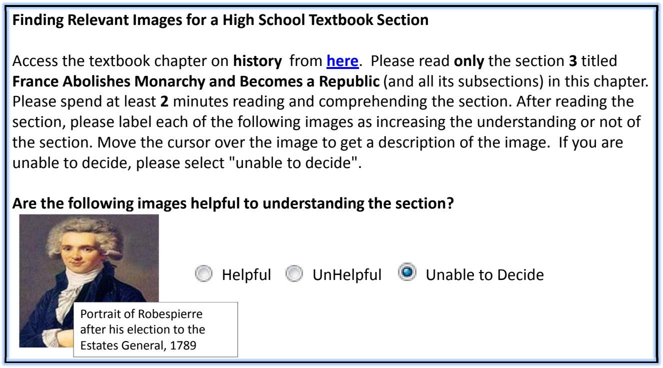
The difficulty of evaluating a system like ours has been earlier experienced in works with similar goals (e.g. story picturing [32, 59], scene synthesis from text [10], translating sentences into pictorial representations [42]). We believe that defining standard benchmarks will greatly accelerate research into these topics and represents an interesting research opportunity. Meanwhile, we resort to conducting a user study to understand the performance of our proposal. Ideally, we would have liked to have high school students from India participate in this study. In the absence of the availability of this subject population to us, we use the Amazon Mechanical Turk platform, which is becoming increasingly mainstream for carrying out user studies [29].

4.1 Methodology

Data Sets

We wanted to evaluate our techniques on textbooks from a developing country. We chose a corpus of high school textbooks published by the National Council of Educational Research and Training, India because of the online availability of these books. We considered fourteen books from grades IX to XII, covering eight subjects: Accounting, Business Studies, Economics, History, Physics, Political Science, (General) Science, and Sociology. They covered seventy chapters and over a thousand sections. Out of these, we randomly selected 100 sections for the user study. This sample size keeps the cost of carrying out the study within a reasonable budget, and yet it is big enough for having meaningful results. Figure 1 shows the subject-wise distribution of sections.

For mining images, AFFINITY limits itself to finding similar Wikipedia articles and extracts images only from them. This constraint allows us to contain the crawling and image extraction effort. Similarly, when querying the search engines, COMITY uses the site restriction feature to obtain only Wikipedia images, which allows the behavior of the two image mining algorithms to be compared. It also reduces the determination of whether the two images are the same to comparing their Wikipedia URLs. We note that the Wikipedia articles are known to differ from the general web pages in characteristics such as the encyclopedic orientation and the style of writing [22], and restricting to Wikipedia reduces the available pool of relevant images to choose from. In spite of these handicaps, our techniques emerge as quite effective in the user study.



Finding Relevant Images for a High School Textbook Section

Access the textbook chapter on **history** from here. Please read **only** the section **3** titled **France Abolishes Monarchy and Becomes a Republic** (and all its subsections) in this chapter. Please spend at least **2** minutes reading and comprehending the section. After reading the section, please label each of the following images as increasing the understanding or not of the section. Move the cursor over the image to get a description of the image. If you are unable to decide, please select "unable to decide".

Are the following images helpful to understanding the section?

○ Helpful ○ UnHelpful ● Unable to Decide

Portrait of Robespierre after his election to the Estates General, 1789

Figure 2: A sample Mechanical Turk HIT

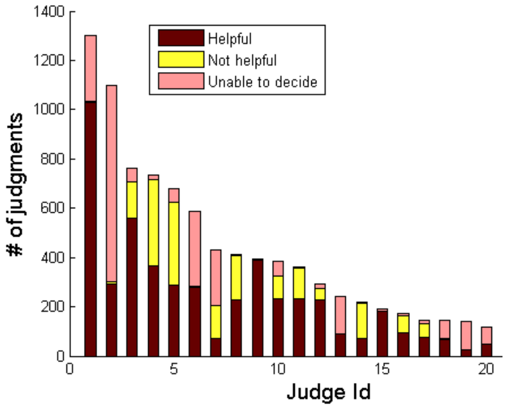


Figure 3: Distribution of judgments for top 20 judges who participated in the most number of HITs

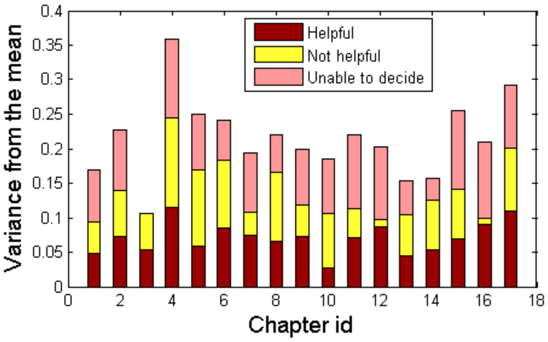


Figure 4: Variance from the mean for the three types of judgments for the Science books

Parameters

In the image mining component, we set the number of images per section parameter k to 5 and the number of desired concept phrases parameter c to 20. For Comity, we set the number of desired image search results per query t to 10. We use two different search engines ($e = 2$): Google and Bing. We use $\theta = 10$ as the smoothing parameter and give equal importance to all the concepts. For Affinity, we obtain $t' = 10$ closest articles for every book section and set $w_1 = 1$, $w_2 = 1$, and $w_3 = 1/3$. During the image assignment, we set the maximum number of images that can be associated with any section to 5 (that is, $K_j = 5$ for all sections).

Judges

Each section along with the candidate images produced by Affinity, Comity, and Ensemble (five per algorithm) was provided to Mechanical Turk as a HIT (Human Intelligence Task). Images within a HIT were intermixed and randomly ordered. Fig. 2 illustrates a sample HIT. Each judge was asked to read the section (provided as a link in the HIT) and rate every image separately as helpful or unhelpful in understanding the material. We also provided the 'unable to decide' option. Images were also provided with the associated captions that the judge could use in making the decision.

We also specified the minimum time a judge was required to spend on a given HIT. This time was arrived at by considering the length of the section and the statistics on average reading speed [21, 27]. We rejected any HIT (and resubmitted it) where the time spent was less than the minimum. We found that different judges took varying amount of time on a particular section which is what we expected (and hoped).

Each HIT was judged by 7 judges. There were 58 judges who took part in the study. Fig. 3 shows the distribution of judgments in terms of helpful, unhelpful, and undecided for the top 20 judges. This figure shows that the distribution of judgments varied across the judges, which indicates the absence of impostor judges who repeatedly judged a HIT using multiple identities. The same figure also indicates that judges were not providing random judgments; otherwise the distribution would have been more uniform.

We also computed the three-dimensional normalized mean vector using all the available judgments for a given chapter. This vector captures the probability of each of the three types of judgments for the images shown for that chapter. Fig. 4 shows the averaged variance from the mean for chapters in the Science books (similar trends held for other subjects). Here the average was performed across all the judges. The deviation is not large, indicating the absence of outlier judges.

Helpfulness Index

We propose *helpfulness index* as the measure for understanding the effectiveness of an image augmentation system. Given the total number of images an image augmentation system can propose, the helpfulness index is defined to be the total number of images deemed helpful for understanding the corresponding section divided by the total number of images.

For an image to be considered helpful for understanding a section, we require that the majority of judges find it helpful. Although we have odd number of judges for every image, judges are allowed to vote 'unable to decide'. Thus, there may be a split vote on an image. We can therefore compute the majority in two different ways:

- *Conservative:* #Helpful > (#Helpful + #Unhelpful)/2,
- *Liberal:* #Helpful $\geq$ (#Helpful + #Unhelpful)/2.

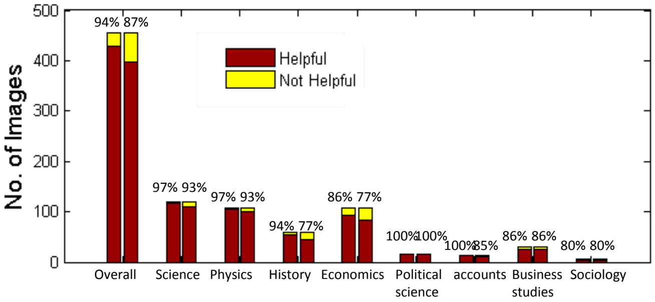


Figure 5: Performance of our system. The first bars correspond to the *Liberal* and the second to the *Conservative* definition of majority. The numbers above the bar give the corresponding helpfulness index.

Conservative requires that an image has a clear helpful majority. *Liberal* breaks ties in the favor of helpfulness. We can thus have two definitions of helpfulness index. By default, we use the *Conservative*.

Note that the helpfulness index is analogous to the notion of *precision* used in information retrieval. One could similarly define a measure analogous to *recall* in terms of the number of images for which a helpfulness decision could be made. This measure turns out to be 100% in our case as there was no image on which all judges were undecided.

4.2 Main Results

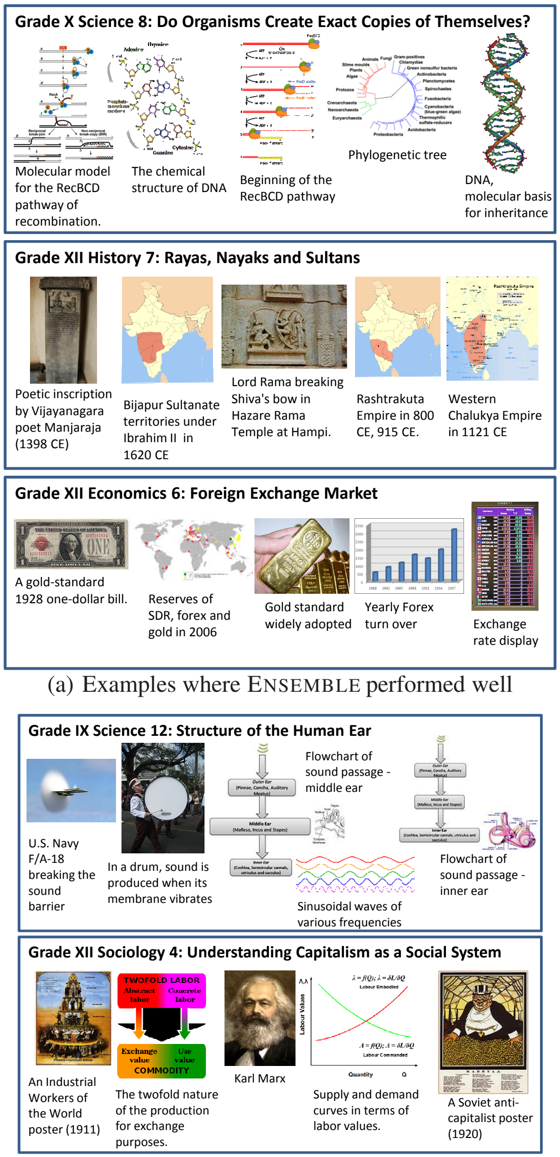
Fig. 5 shows the performance of our system, both at the aggregate level ("overall") as well as at the individual subject level. These results were produced by the ENSEMBLE algorithm.

The results are quite encouraging. We see that when we use *Conservative* definition of the majority of judgments, the judges considered 87% of the images assigned to various textbook sections to be helpful. This number increases to 94% under the *Liberal* definition. We can also see that the performance is maintained across all the subjects.

We also manually inspected the results. We discuss next some anecdotal examples.

Anecdotal Examples

Fig. 6(a) shows some examples where ENSEMBLE performed well. We show top five images produced by ENSEMBLE for three different sections from three different subjects. We can see that the images are quite relevant. We discuss the first example in more depth. This example shows the proposed augmentations for the section on how organisms create exact copies of themselves, appearing in the eighth chapter titled 'How do organisms reproduce' in the grade X Science book. This section discusses three main points: (1) due to evolution, organisms are similar in their blueprint; (2) DNA replicates to pass on genetic material; and (3) DNA copying during reproduction should be consistent so that the organism is well adjusted to its ecosystem. We observe that proposed images convey related information. The image on Phylogenetic tree captures the evolutionary relationships among biological species. The two images of DNA (chemical and physical structure) are illustrative of how the DNA can be easily replicated by breaking its double Helix structure. The section describes the consistency requirement of DNA copying using bacteria as the example organism. The images of RecBCD pathway in E. coli bacterium are complementary as it



Grade X Science 8: Do Organisms Create Exact Copies of Themselves?
Molecular model for the RecBCD pathway of recombination.
The chemical structure of DNA
Beginning of the RecBCD pathway
Phylogenetic tree
DNA, molecular basis for inheritance
Grade XII History 7: Rayas, Nayaks and Sultans
Poetic inscription by Vijayanagara poet Manjaraja (1398 CE)
Bijapur Sultanate territories under Ibrahim II in 1620 CE
Lord Rama breaking Shiva's bow in Hazare Rama Temple at Hampi.
Rashtrakuta Empire in 800 CE, 915 CE.
Western Chalukya Empire in 1121 CE
Grade XII Economics 6: Foreign Exchange Market
A gold-standard 1928 one-dollar bill.
Reserves of SDR, forex and gold in 2006
Gold standard widely adopted
Yearly Forex turn over
Exchange rate display
Grade IX Science 12: Structure of the Human Ear
U.S. Navy F/A-18 breaking the sound barrier
In a drum, sound is produced when its membrane vibrates
Flowchart of sound passage - middle ear
Sinusoidal waves of various frequencies
Flowchart of sound passage - inner ear
Grade XII Sociology 4: Understanding Capitalism as a Social System
An Industrial Workers of the World poster (1911)
TWOFOLD LABOR
Abstract labor
Concrete labor
Exchange value
Use value
COMMODITY
The twofold nature of the production for exchange purposes.
Karl Marx
Labour Values
Quantity
Supply and demand curves in terms of labor values.
A Soviet anti-capitalist poster (1920)

(a) Examples where ENSEMBLE performed well

(b) Examples where ENSEMBLE did not perform well

Figure 6: Anecdotal examples

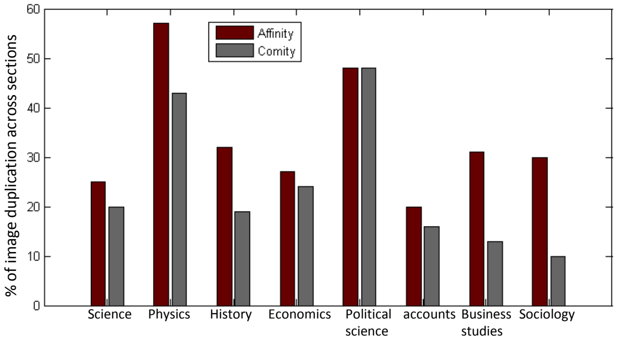
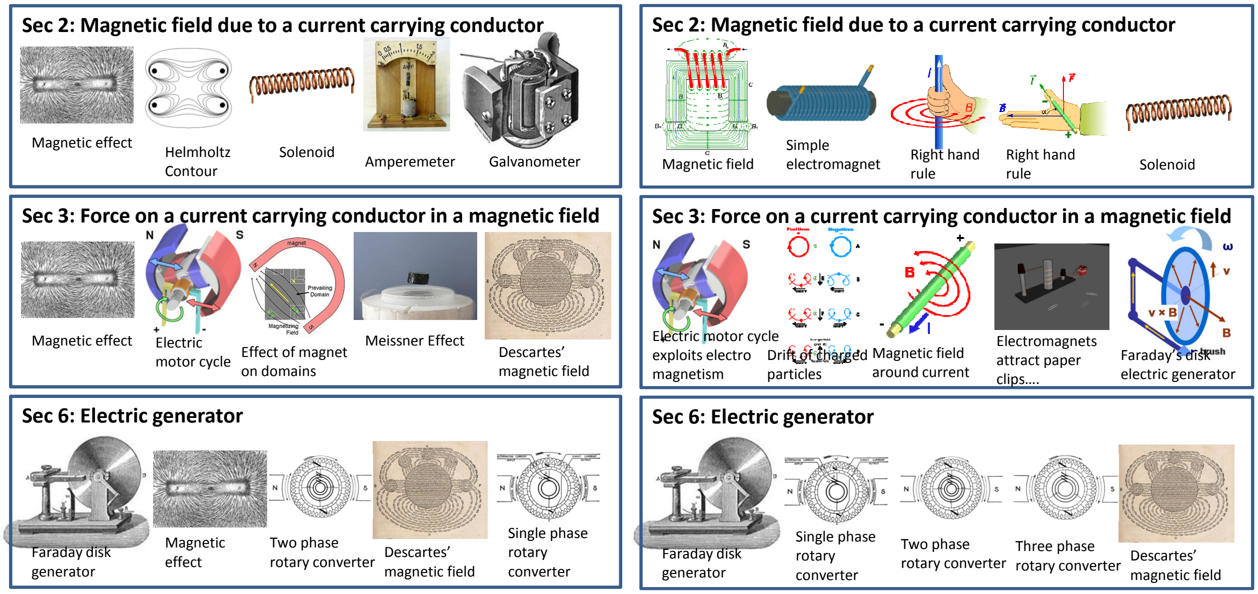


Figure 7: Chapter level duplication in images for AFFINITY and COMITY


Sec 2: Magnetic field due to a current carrying conductor
Magnetic effect
Helmholtz Contour
Solenoid
Amperemeter
Galvanometer
Sec 3: Force on a current carrying conductor in a magnetic field
Magnetic effect
Electric motor cycle
Effect of magnet on domains
Meissner Effect
Descartes' magnetic field
Sec 6: Electric generator
Faraday disk generator
Magnetic effect
Two phase rotary converter
Descartes' magnetic field
Single phase rotary converter
Sec 2: Magnetic field due to a current carrying conductor
Magnetic field
Simple electromagnet
Right hand rule
Right hand rule
Solenoid
Sec 3: Force on a current carrying conductor in a magnetic field
Electric motor cycle exploits electro magnetism
Drift of charged particles
Magnetic field around current
Electromagnets attract paper clips....
Faraday's disk electric generator
Sec 6: Electric generator
Faraday disk generator
Single phase rotary converter
Two phase rotary converter
Three phase rotary converter
Descartes' magnetic field

(a) AFFINITY before image assignment (b) AFFINITY after image assignment

Figure 8: Effect of optimization: Comparison of top five images for three sections from the chapter titled 'Magnetic effects of electric current' in the grade X Science book

plays crucial role of initiating recombinational repair of potentially lethal double strand breaks in DNA.

Fig. 6(b) shows the cases that illustrate opportunity for further work. The first example is from a section in the twelfth chapter titled 'Sound' in the grade IX Science book. This section describes the structure of the human ear and how it enables conversion of pressure variations in air with audible frequencies into electric signals that travel to the brain via the auditory nerve. The last two images that depict the passage of sound through the middle and inner ears are obviously useful. The image illustrating the sinusoidal waves of various frequencies is also helpful. While the first two images are also relevant for the chapter as a whole, they are not particularly suited for this section. The reason these images were ranked high was that they have overlapping words; an important component of human ear is the thin membrane called the ear drum, which can be confused with a playing drum or membrane on it.

4.3 Virtue of the Image Assignment Optimization

The image mining algorithms retrieve images that can potentially repeat across multiple sections of the same chapter. In fact, as can be seen in Fig. 7, this duplication is quite high for both AFFINITY and COMITY. The duplication in this figure has been computed at the subject level by averaging the overlaps computed at the chapter level for all the chapters in that subject. The image assignment algorithm reorders the images and uses more of the available ordered set of images so that the images selected within a chapter are distinct.

We illustrate this point further with an example from the chapter on 'Magnetic effects of electric current' in the grade X Science book. Fig. 8a shows the top five images for three sections in this chapter proposed by AFFINITY before image assignment. Clearly, while all the sets of images shown are relevant to the chapter and can serve as illustrations for the corresponding sections, there is redundancy amongst them. Fig. 8b shows the top five images for the same sections, after applying the image assignment optimization.

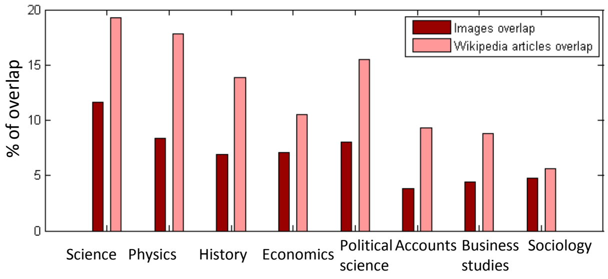


Figure 9: Overlap between AFFINITY and COMITY

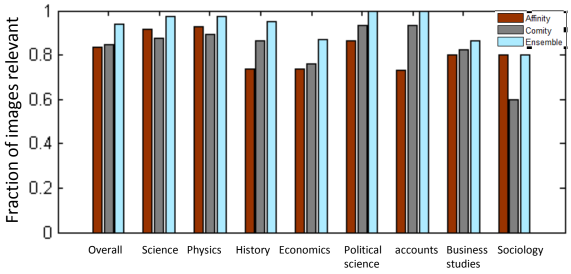


Figure 10: Comparison between AFFINITY, COMITY, and ENSEMBLE

Besides removing redundancy, the optimization also enabled selection of images more relevant to the sections. For instance, Section 2 on 'Magnetic field due to current carrying conductor' now includes images for the right hand rule, which is a guide to finding the direction of magnetic field around a current-carrying wire. Section 6 on 'Electric generator' now includes wiring schematic for each of (single, two, three) phase to rotary converter, which is a motor generator that leverages magnetic field to generate power.

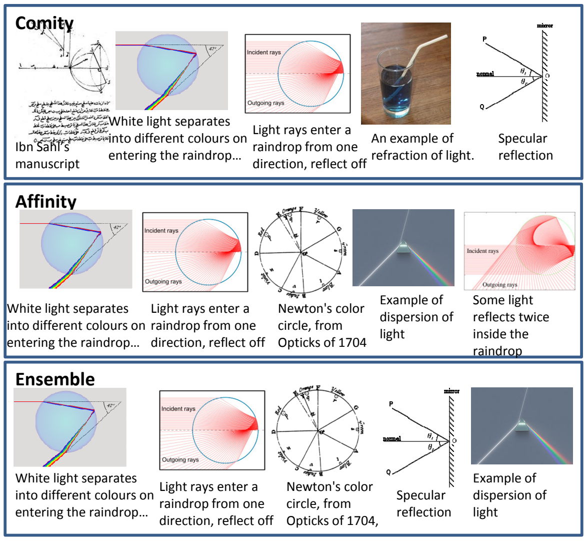

Comity
Ibn Sahl's manuscript
White light separates into different colours on entering the raindrop...
Light rays enter a raindrop from one direction, reflect off
An example of refraction of light.
Specular reflection
Affinity
White light separates into different colours on entering the raindrop...
Light rays enter a raindrop from one direction, reflect off
Newton's color circle, from Opticks of 1704
Example of dispersion of light
Some light reflects twice inside the raindrop
Ensemble
White light separates into different colours on entering the raindrop...
Light rays enter a raindrop from one direction, reflect off
Newton's color circle, from Opticks of 1704,
Specular reflection
Example of dispersion of light
Incident rays
Outgoing rays
mirror
normal

Figure 11: Illustration of the benefit of ENSEMBLE for the section on 'Dispersion of white light by a glass prism' in grade X Science textbook

4.4 Orthogonality of AFFINITY and COMITY

ENSEMBLE is best suited when the ensembled algorithms use largely non-overlapping signals. In order to understand the overlap between AFFINITY and COMITY, we compute two statistics: one using the overlap in the retrieved images and the other using the overlap in the articles from which these images are retrieved.

Fig. 9 provides the two statistics, subject-wise. The image overlap is computed by measuring the ratio of intersection to union in the top five images retrieved by the two algorithms. We compute overlap at the chapter level, and average it across all chapters for a subject. To compute article overlap, we pool the articles from which the top images for a particular chapter were retrieved for each algorithm. Then, we compute the ratio of intersection to union of these articles for every chapter. We can see that the overlap is small. These numbers are substantially lower when analyzed at the section level.

4.5 Benefit of Ensembling

Fig. 10 compares ENSEMBLE with AFFINITY and COMITY. We can see that ENSEMBLE consistently produces a larger fraction of the helpful images than either of the other two, thereby confirming the value of ensembling.

We use Fig. 11 to provide an illustration, using the results for a section from the chapter on 'Human eye and colourful world' in the grade X science book. This chapter describes the optical phenomena in nature, including rainbow formation and the splitting of white light and blue color of the sky. The section we discuss introduces the concept of dispersion and assumes that the students are already familiar with refraction. The ensembled results are clearly superior than the results of the other two algorithms.

5. CONCLUSIONS AND FUTURE WORK

We studied the feasibility of enriching textbooks with links to relevant images and presented a solution for this problem. The solution comprises of three components: a component for optimizing the assignment of images to different sections within a chapter on a per algorithm basis, another for mining images from the web using multiple algorithms, and finally a component for ensembling the results of the per algorithm assignments. We devised a rigorous formulation of the image assignment problem and gave a polynomial time algorithm for solving the problem optimally. We also presented two particular image mining algorithms that utilize orthogonal signals and hence obtain different sets of relevant images. Finally, we provided an ensembling algorithm for combining the assignments.

We conducted a user study employing a corpus of fourteen high school textbooks, published by the central body for education in India. We used the Amazon Mechanical Turk platform for this purpose. Seven judges each, coming from a population of 56 judges, judged the results produced by our implementation for a random sample of 100 textbook sections. The results demonstrate the promise of the proposed system: the judges conservatively considered 87% of the images assigned to various sections to be helpful for understanding the corresponding section and the performance was maintained across the subjects.

The directions for future work include investigating what new issues arise if the ideas from this paper were to be extended for enriching textbook material with other media types. Another important research direction is developing benchmarks for measuring the performance of these systems. Finally, it is worthwhile to examine synergies between algorithmic approaches such as the one presented in this paper and the collaborative approaches such as the one proposed in [2].

6. REFERENCES

[1] *Knowledge for Development: World Development Report 1998/99*. World Bank, 1998.

[2] Improving India's education system through information technology. *IBM*, 2005.

[3] A. Adams and J. van der Gaag. First step to literacy: Getting books in the hands of children. *The Brookings Institution*, January 2011.

[4] R. Agrawal, S. Gollapudi, A. Kannan, and K. Kenthapadi. Identifying enrichment candidates in textbooks. In *WWW*, 2011.

[5] R. Agrawal, S. Gollapudi, K. Kenthapadi, N. Srivastava, and R. Velu. Enriching textbooks through data mining. In *First Annual ACM Symposium on Computing for Development (ACM DEV)*, 2010.

[6] K. Barnard, P. Duygulu, D. Forsyth, N. de. Freitas, D. M. Blei, and M. I. Jordan. Matching words and pictures. In *Journal of Machine Learning Research*, volume 3, 2003.

[7] S. Boss. What's next: Curling up with e-readers. *Stanford Social Innovation Review*, Winter 2011.

[8] J. P. G. Chimombo. Issues in basic education in developing countries: An exploration of policy options for improved delivery. *Journal of International Cooperation in Education*, 8(1), 2005.

[9] J. Coiro, M. Knobel, C. Lankshear, and D. Leu, editors. *Handbook of research on new literacies*. Lawrence Erlbaum, 2008.

[10] B. Coyne and R. Sproat. WordsEye: An automatic text-to-scene conversion system. In *SIGGRAPH*, 2001.

[11] M. Crossley and M. Murby. Textbook provision and the quality of the school curriculum in developing countries: Issues and policy options. *Comparative Education*, 30(2), 1994.

[12] A. Csomai and R. Mihalcea. Linking educational materials to encyclopedic knowledge. In *AIED*, 2007.

[13] R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys*, 40, September 2008.

[14] L. Downes. *The laws of disruption: Harnessing the new forces that govern life and business in the digital age*. Basic Books, 2009.

[15] Y. Feng and M. Lapata. Topic models for image annotation and text illustration. In *HLT-NAACL*, 2010.

[16] B. Fuller. What school factors raise achievement in the third world? *Review of educational research*, 57(3), 1987.

[17] K. Gaikwad, G. Paruthi, and W. Thies. Interactive DVDs as a platform for education. In *ICTD*, 2010.

[18] J. Gillies and J. Quijada. Opportunity to learn: A high impact strategy for improving educational outcomes in developing countries. *USAID Educational Quality Improvement Program (EQUIP2)*, 2008.

[19] P. Glewwe, M. Kremer, and S. Moulin. Many children left behind? Textbooks and test scores in Kenya. *American Economic Journal: Applied Economics*, 1(1), 2009.

[20] R. Gorman and J. Ginsburg. *Copyright: Cases and materials*. Foundation Press, 2006.

[21] W. Grabe. Efficiency in reading – thirty fiver years later. In *Reading in a Foreign Language*, 2010.

[22] G. Grefenstette. Comparing the language used in Flickr, general web pages, Yahoo images and Wikipedia. In *LREC Workshop on Language Resources for Content-Based Image Retrieval (OntoImage)*, 2008.

[23] G. Guo, G. Xu, H. Li, and X. Cheng. A unified and discriminative model for query refinement. In *SIGIR*, 2008.

[24] E. A. Hanushek and L. Woessmann. The role of education quality for economic growth. *Policy Research Department Working Paper 4122, World Bank*, 2007.

[25] E. Hatcher and O. Gospodnetic. *Lucene in Action*. Manning, 2004.

[26] S. Heyneman, J. Farrell, and M. Sepulveda-Stuardo. Textbooks and achievement in developing countries: What we know. *Journal of Curriculum Studies*, 13(3), 1981.

[27] K. Holmqvist, J. Holsanova, M. Barthelson, and D. Lundqvist. Reading or Scanning? A study of newspaper and net paper reading. In J. R. Hyönä and H. Deubel, editors, *The mind's eye: Cognitive and applied aspects of eye movement research*. Elsevier Science, 2003.

[28] S. Huston and W. B. Croft. Evaluating verbose query processing techniques. In *SIGIR*, 2010.

[29] P. G. Ipeirotis. Analyzing the Amazon mechanical turk marketplace. *ACM Crossroads*, 17(2), 2010.

[30] A. Jain and B. Yu. Automatic text location in images and video frames. *Pattern recognition*, 31(12), 1998.

[31] A. Jawa, S. Datta, S. Nanda, V. Garg, V. Varma, S. Chande, and M. K. P. Venkata. Smeo: A platform for smart classrooms with enhanced information access and operations automation. In *10th International Conference on Next Generation Wired/Wireless Advanced Networking*, 2010.

[32] D. Joshi. The story picturing engine: Finding elite images to illustrate a story using mutual reinforcement. In *ACM SIGMM International Workshop on Multimedia Information Retrieval*, 2004.

[33] D. Jurafsky and J. Martin. *Speech and language processing*. Prentice Hall, 2008.

[34] J. S. Justeson and S. M. Katz. Technical terminology: Some linguistic properties and an algorithm for indentification in text. *Natural Language Engineering*, 1(1), 1995.

[35] P. Katsioloudis. *Identification of Quality Indicators of Visual-Based Learning Material in Technology Education Programs for Grades 7-12*. PhD thesis, North Carolina State University, 2007.

[36] E. Kuiper, M. Volman, and J. Terwel. The Web as an information resource in K-12 education: Strategies for supporting students in searching and processing information. *Review of Educational Research*, 75(3), 2005.

[37] G. Kumaran and V. R. Carvalho. Reducing long queries using query quality predictors. In *SIGIR*, 2009.

[38] R. Lienhart and A. Wernicke. Localizing and segmenting text in images and videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 12(4), 2002.

[39] R. Mayer. *Multimedia Learning*. Cambridge University Press, 2001.

[40] O. Medelyan. *Human-competitive automatic topic indexing*. PhD thesis, The University of Waikato, 2009.

[41] R. Mihalcea and A. Csomai. Wikify!: Linking documents to encyclopedic knowledge. In *CIKM*, 2007.

[42] R. Mihalcea and C. W. Leong. Toward communicating simple sentences using pictorial representations. *Machine Translation*, 22(3), 2008.

[43] D. Milne. *Applying Wikipedia to Interactive Information Retrieval*. PhD thesis, University of Waikato, 2010.

[44] R. Mohammad and R. Kumari. Effective use of textbooks: A neglected aspect of education in Pakistan. *Journal of Education for International Development*, 3(1), 2007.

[45] J. Moulton. How do teachers use textbooks and other print materials: A review of the literature. *The Improving Educational Quality Project, South Africa*, 1994.

[46] H. Müller, P. Clough, T. Deselaers, and B. Caputo. *ImageCLEF: Experimental Evaluation in Visual Information Retrieval*. Springer, 2010.

[47] N. Mulvany. *Indexing books*. University of Chicago Press, 2005.

[48] S. Panjwani, L. Micallef, K. Fenech, and K. Toyama. Effects of integrating digital visual materials with textbook scans in the classroom. *International Journal of Education and Development using Information and Communication Technology*, 5(3), 2009.

[49] C. Papadimitriou and K. Steiglitz. *Combinatorial optimization: Algorithms and complexity*. Dover, 1998.

[50] A. Riddell. *Factors influencing educational quality and effectiveness in developing countries: A review of research*. Deutsche Gesellschaft fur Technische Zusammenarbeit (GTZ), Germany, 2008.

[51] J. Rocchio. Relevance feedback in information retrieval. In *The SMART Retrieval System – Experiments in Automatic Document Processing*, 1971.

[52] D. Saari. *Decisions and elections: Explaining the unexpected*. Cambridge University Press, 2001.

[53] B. W. Speck, T. R. Johnson, C. P. Dice, and L. B. Heaton. *Collaborative writing: An annotated bibliography*. Greenwood Press, 1999.

[54] A. Spink, B. Jansen, and H. Ozmultu. Use of query reformulation and relevance feedback by Excite users. In *Internet Research: Electronic Networking Applications and Policy*, 2000.

[55] M. Stein, C. Stuen, D. Carnine, and R. M. Long. Textbook evaluation and adoption. *Reading & Writing Quarterly*, 17(1), 2001.

[56] L. Von Ahn and L. Dabbish. Labeling images with a computer game. In *CHI*, 2004.

[57] X. Xue, S. Huston, and W. B. Croft. Improving verbose queries using subset distribution. In *CIKM*, 2010.

[58] Y. Yang, N. Bansal, W. Dakka, P. Ipeirotis, N. Koudas, and D. Papadias. Query by document. In *WSDM*, 2009.

[59] X. Zhu, A. B. Goldberg, M. Eldawy, C. R. Dyer, and B. Strock. A text-to-picture synthesis system for augmenting communication. In *AAAI*, 2007.