Democratizing Learning from Textbooks

Rakesh Agrawal
Joint work with S. Gollapudi, A. Kannan, K. Kenthapadi, et al.

July 15, 2014
Research Expedition

Assuming

- Availability of inexpensive cloud-connected electronic devices
- Migration of the content of the current textbooks to those devices

How to enhance the quality of the electronic textbooks?
Data Mining for Enhancing Electronic Textbooks

Diagnostic tools for identifying weaknesses in textbooks

Within section deficiencies
- Syntactic complexity of writing and dispersion of key concepts in the section [AGK+11a]

Across sections deficiencies
- Comprehension burden due to non-sequential presentation of concepts [ACG+12]

Algorithmic enhancement of textbooks for enriching reading experience

References to selective web content
- Links to authoritative articles [AGK+10], images [AGK+11b] and videos [ACG+14] based on the focus of the section

References to prerequisites
- Links to concepts necessary for understanding the present section, derived using a model of how students read textbooks [AGK+13]

- Validation on textbooks from U.S.A and India, on different subjects, across grades
- Prototypes and research papers (see References)

Joint work with Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi, et al.
A Peek Under the Hood

Diagnostic tools for identifying weaknesses in textbooks

Within section deficiencies
- Syntactic complexity of writing and dispersion of key concepts in the section [AGK+11a]

Across sections deficiencies
- Comprehension burden due to non-sequential presentation of concepts [ACG+12]

Algorithmic enhancement of textbooks for enriching reading experience

References to selective web content
- Links to authoritative articles [AGK+10], images [AGK+11b] and videos [ACG+14] based on the focus of the section

References to prerequisites
- Links to concepts necessary for understanding the present section, derived using a model of how students read textbooks [AGK+13]

• Validation on textbooks from U.S.A and India, on different subjects, across grades
• Prototypes and research papers (see References)

Joint work with Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi, et al.
Identification of Deficient Sections

**Decision Variables**
- Dispersion of key concepts
- Syntactic complexity of writing

**Probabilistic Decision Model**

**Algorithmically Generated Training Set**
- Map a section to closest Wikipedia article version
- Impute immaturity score to section
- Perform thresholding to get labels

Textbooks

Deficient / Good / Examine
Dispersion of Key Concepts

Many unrelated concepts ➔ Hard to understand section

• $V = \text{set of key concepts discussed in section } s$
  – Terminological noun phrases: Linguistic pattern $A^*N^+$ ($A$: adjective; $N$: noun)
  – “concepti” Wikipedia titles

• $\text{Related}(x,y) = \text{Concept } x \text{ is related to concept } y$
  – Co-occurrence
  – $true$ if Wikipedia article for $x$ links to the article for $y$

• $\text{Dispersion}(s) := \text{Fraction of unrelated concept pairs}$
  – $(1 – \text{Edge Density})$ of the concept graph
Many unrelated concepts [high dispersion]:

- Factors like capital contribution and risk vary with the size and nature of business, and hence a form of business organisation that is suitable from the point of view of the risks for a given business when run on a small scale might not be appropriate when the same business is carried on a large scale.
3. Direct post is for direct advertising. It can be both addressed as well as unaddressed.
4. International Money Transfer through collaboration with Western Union financial services, USA, which enables remittance of money from 185 countries to India.
5. Passport facilities — A unique partnership with the ministry of external affairs for facilitating passport application.
6. Speed Post: It has over 1000 destinations in India and links with 97 major countries across the globe.
7. e-bill post is the latest offering of the department to collect bill payment across the counter for BSNL and Bharti Airtel.

**Telecom Services**

World class telecommunications infrastructure is the key to rapid economic and social development of the country. It is in fact the backbone of every business activity. In today’s world the dream of doing business across continents will remain a dream in the absence of telecom infrastructure. There have been far reaching developments in the convergence of telecom, IT, consumer electronics and media industries worldwide. Recognising the potential in enhancing quality of life and to facilitate India’s vision of becoming IT super power by the year 2025, new Telecom Policy Framework 1999 and Broadband Policy 2004 were developed by the Government of India. Through this framework the government intends to provide both universal services to all uncovered areas and high-level services for meeting the needs of the country’s economy. The various types of telecom services are:

(i) **Cellular mobile services:** These are all types of mobile telecom services including voice and non-voice messages, data services and PCO services utilising any type of network equipment within their service area. They can also provide direct interconnectivity with any other type of telecom service provider.

(ii) **Radio paging services:** Radio Paging Service is an affordable means of transmitting information to persons even when they are mobile. It is a one-way information broadcasting solution, and has spread its reach far and wide. Radio paging services are available including tone only, numeric only and alpha/numeric paging.

(iii) **Fixed line services:** These are all types of fixed services including voice and non-voice messages and data services to establish linkages for long distance traffic. These utilise any type of network equipment primarily connected through fibre optic cables laid across the length and breadth of the country. The also provide interconnectivity with other types of telecom services.

(iv) **Cable services:** These are linkages and switched services within a licensed area of operation to operate media services, which are essentially one way entertainment related services. The two way communication including voice.
Diagnostic tools for identifying weaknesses in textbooks

Within section deficiencies

Syntactic complexity of writing and dispersion of key concepts in the section [AGK+11a]

Across sections deficiencies

Comprehension burden due to non-sequential presentation of concepts [ACG+12]

Algorithmic enhancement of textbooks for enriching reading experience

References to selective web content

Links to authoritative articles [AGK+10], images [AGK+11b] and videos [ACG+14] based on the focus of the section

References to prerequisites

Links to concepts necessary for understanding the present section, derived using a model of how students read textbooks [AGK+13]

• Validation on textbooks from U.S.A and India, on different subjects, across grades
• Prototypes and research papers (see References)

Another Peek
Comity

- Intuition: Combine results of a large number of short, but relevant queries
  - Search engines barf on long queries (such as entire section content)
- Identify key concepts present in a section, $C$
- Form two-concept and three-concept queries, $Q$
- For each $q \in Q$, obtain ranked list of objects $I(q)$
- Relevance score($i$) of object $i = \sum_q f($position of object in $I(q)$, importance of concepts in $q$)
Avoid repetition across sections:

\[
\max \sum_{i \in I} \sum_{j \in S} x_{ij} \cdot \lambda_{ij}
\]

Relevance score of object \(i\) to section \(j\)

s.t.

\[x_{ij} \in \{0, 1\} \quad \forall i \in I \forall j \in S\]

=1 if image \(i\) is selected for section \(j\) else 0

\[\sum_{i \in I} x_{ij} \leq K_j \quad \forall j \in S\]

Constraint: At most \(K_j\) images can be assigned to section \(j\)

\[\sum_{j \in S} x_{ij} \leq 1 \quad \forall i \in I\]

Constraint: An image can belong to at most one section

**Can be solved optimally in polynomial time**
Video Augmentation: Make inaccessible accessible

Table of contents for navigating the book (automatically extracted)

3.1: Atoms and Molecules: Laws of Chemical Combination

Chapter 3

ATOMS AND MOLECULES

Ancient Indian and Greek philosophers have always wondered about the unknown and unseen form of matter. The idea of divisibility of matter was considered long back in India, around 500 BC. An Indian philosopher, Maharshi Kanada, postulated that if we go on dividing matter (prakriti), we shall get smaller and smaller particles. Ultimately, a time will come when we shall come across the smallest particles beyond which further division will not be possible. He named these particles Pramanis. Another Indian philosopher, Pashudha Katyayana, elaborated this doctrine and said that these particles normally exist in a combined form which gives us various forms of matter.

Around the same era, ancient Greek philosophers – Democritus and Leucippus suggested that if we go on dividing matter, a stage will come when particles obtained cannot be divided further. Democritus called these indivisible particles atoms (meaning indivisible). All this was based on philosophical considerations and not much experimental work to validate these ideas could be done till the eighteenth century.

By the end of the eighteenth century, scientists recognised the difference between elements and compounds and naturally became interested in finding out how and why elements combine and what happens when they combine. Antoine L. Lavoisier laid the foundation of chemical sciences by establishing two important laws of chemical combination.

3.1 Laws of Chemical Combination

The following two laws of chemical combination were established after much experimentation by Lavoisier and Joseph L. Proust.

Activity 3.1

Take one of the following sets, X and Y, of chemicals:

- Copper sulphate – sodium carbonate
- Barium chloride – sodium sulphate
- Small solid tube – small liquid tube

Prepare separately a 5% solution of any one pair of substances listed under X and Y in water. Take a little amount of solution of Y in a conical flask and some solution of X in an ignition tube. Hang the ignition tube in the flask carefully; see that the solutions do not get mixed. Put a cork on the flask (see Fig. 3.1).

Augmentations panel: Video demonstrates the reaction for the second set of chemicals prescribed

Win8 Surface Prototype
This section is about magnetic field lines created by bar magnet. Section contains static images of magnetic field for bar magnet, solenoid and dipole.

The videos describes step-by-step magnetic field creation in bar magnet.
Ongoing Research

• Inferring learning units and dependence between them from current educational material (knowledge graph)
• Improvement in educational material based on data on student interactions with the material
• Synergies with crowdsourcing approaches
• Individualized learning plans
• Dynamic formation of classes and study groups
• Performance evaluation methodologies and benchmarks
• Issues related to privacy, security, confidentiality, copyright, attribution, revenue sharing, …


