



Microsoft
Research Silicon Valley

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 a 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](#))

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

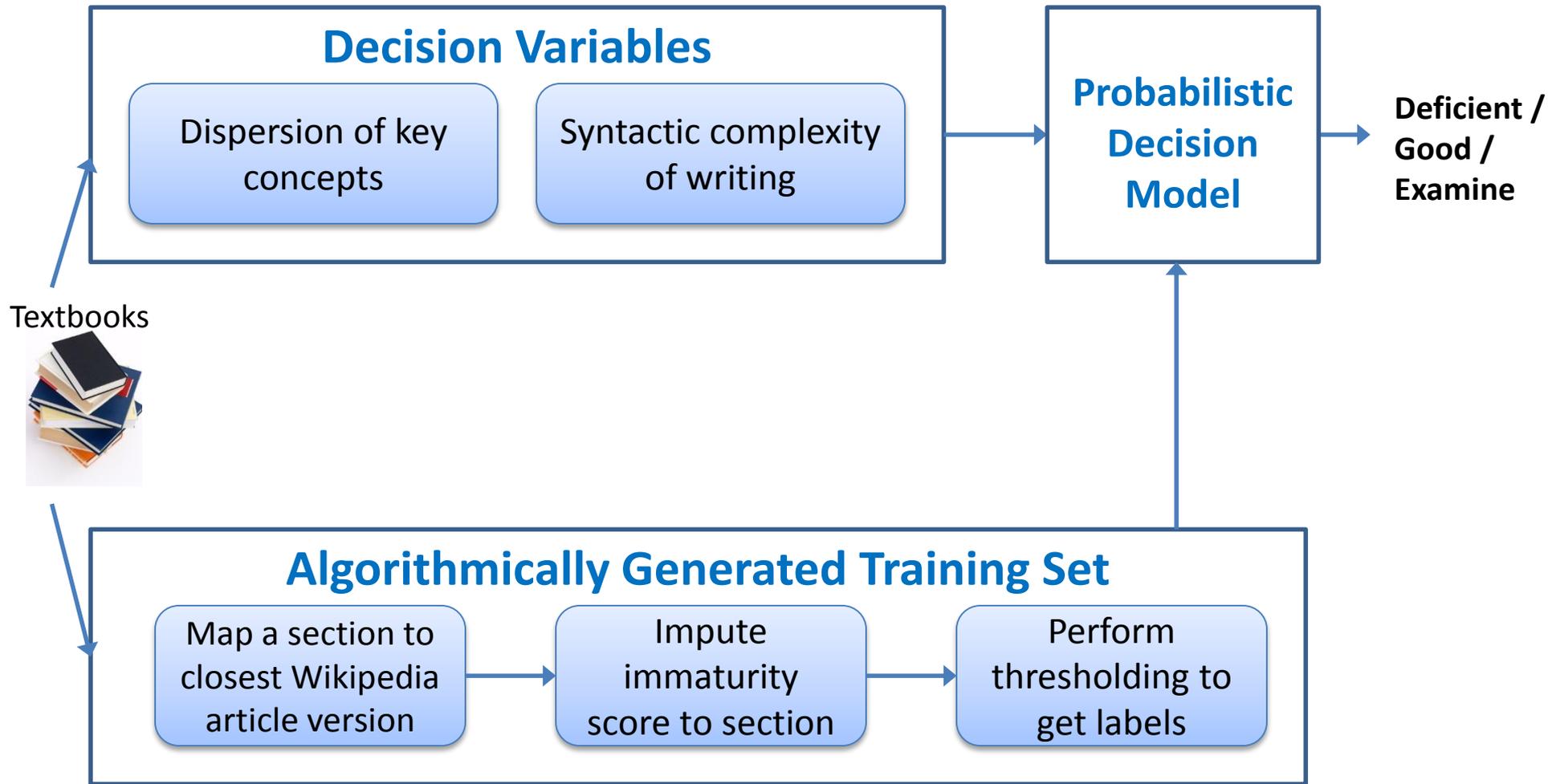
Links to authoritative articles [AGK+10], images [AGK+11b] and videos [ACG+14] based on the focus of the section

References to prerequisites

Links to concepts necessary for understanding the present section, derived using a model of a 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](#))

Identification of Deficient Sections



Dispersion of Key Concepts

Many unrelated concepts → Hard to understand section

- V = set of key concepts discussed in section s
 - *Terminological noun phrases*: Linguistic pattern A^*N^+ (A: adjective; N: noun)
 - *“concepti” Wikipedia titles*
- $Related(x,y)$ = Concept x is related to concept y
 - *Co-occurrence*
 - *true* if Wikipedia article for x links to the article for y
- $Dispersion(s)$:= Fraction of unrelated concept pairs
 - $(1 - \text{Edge Density})$ of the concept graph

Illustrative Result: Deficient Section

CHAPTER 2

FORMS OF BUSINESS ORGANISATION

2.7 CHOICE OF FORM OF BUSINESS ORGANISATION

After studying various forms of business organisations, it is evident that each form has certain advantages as well as disadvantages. It, therefore, becomes vital that certain basic considerations are kept in mind while choosing an appropriate form of

(ii) Liability: In case of sole proprietorship and partnership firms, the liability of the owners/partners is unlimited. This may call for paying the debt from personal assets of the owners. In joint Hindu family business, only the *karta* has unlimited liability. In cooperative societies and companies, however, liability is limited and creditors can force payment of their claims only to the extent of the company's assets.

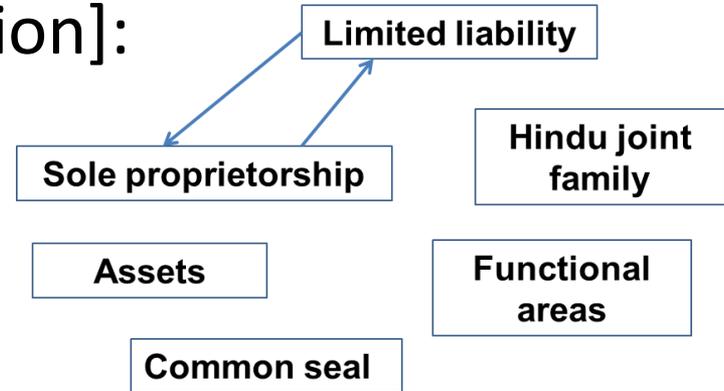
above are inter-related. Factors like capital contribution and risk vary with the size and nature of business, and hence a form of business organisation that is suitable from the point of view of the risks for a given business when run on a small scale might not be

operations. Cooperative societies and companies have to be compulsorily registered. Formation of a company involves a lengthy and expensive legal procedure. From the point of view of initial cost, therefore, sole proprietorship is the preferred form as it involves least expenditure. Company form of organisation, on the other hand, is more complex and involves greater costs.

in nature and require professionalised management, company form of organisation is a better alternative. Proprietorship or partnership may be suitable, where simplicity of operations allow even people with limited skills to run the business. Thus, the nature of operations and the need for professionalised management affect the choice of the form of organisation.

(v) Capital considerations: Companies organisations one by one. In Table 2.5, we analysed characteristics of different forms of organisations taken together so as to enable you to understand on a comparative basis as to where a form of organisation stands in comparison to others in respect of select features.

- Many unrelated concepts [high dispersion]:



- Long sentences, e.g.,
 - *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.*

Win 8 Surface Prototype

← Class 11 Business Studies

CHAPTER 4 BUSINESS SERVICES

Hide Augmentation

Show Highlights

videos articles images diagnostics navigation

BUSINESS SERVICES 100

- Direct post is for direct advertising. It can be both addressed as well as unaddressed.
- International Money Transfer through collaboration with Western Union financial services, USA, which enables remittance of money from 185 countries to India.
- Passport facilities — A unique partnership with the ministry of external affairs for facilitating passport application.
- Speed Post: It has over 1000 destinations in India and links with 97 major countries across the globe.
- 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 inter connectivity 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 fiber optic cables laid across the length and breadth of the country. They also provide inter connectivity 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,

Communication services

Sentence Complexity

Percentile %	Complexity
0	10
20	14
40	15
50	18.0
60	17
80	18
100	22

Word Complexity

Percentile %	Complexity
0	1.4
20	1.5
40	1.6
60	1.7
80	1.8
100	2.2

Dispersion

Percentile %	Dispersion
0	0.2
20	0.3
40	0.4
50	0.72
60	0.6
80	0.7
100	0.8

Sentence Count: 80 Word Count: 1,485

High School Textbook from National Council of Educational Research and Training (NCERT), India

Another Peek

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 a 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](#))

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 $l(q)$
- Relevance score(i) of object i =
$$\sum_q f(\text{position of object in } l(q), \text{ importance of concepts in } q)$$

From Section Level to Book Level Assignments

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

Total relevance score for the chapter: sum of relevance scores of objects assigned

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

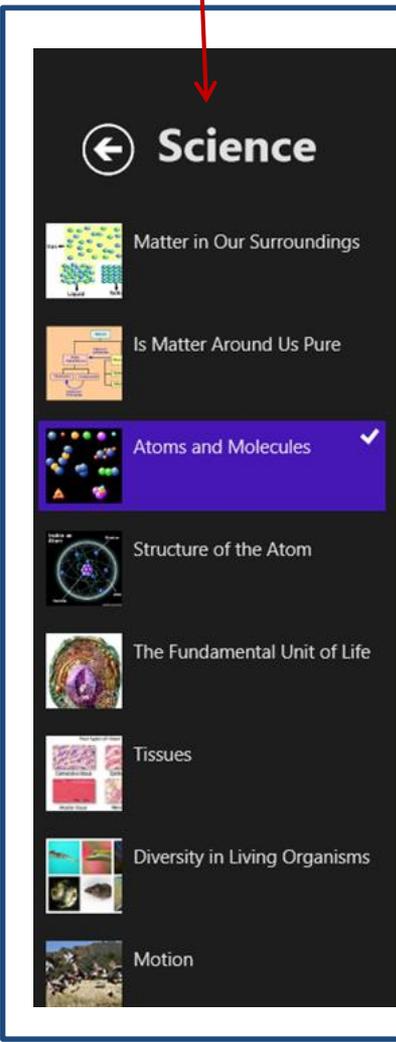
Can be solved optimally in polynomial time

Video Augmentation: Make inaccessible accessible

Table of contents for navigating the book (automatically extracted)

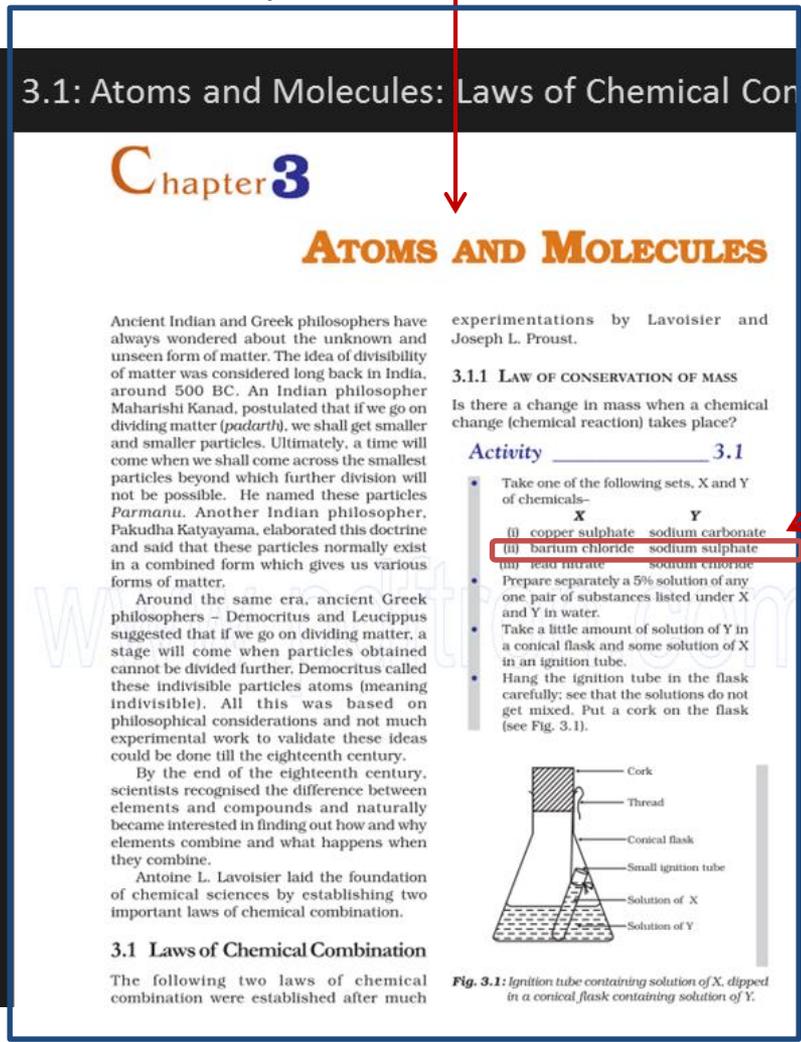
Re-rendered section: This section, about the laws of chemical combination, prescribes an activity for the chemistry lab, but the school might lack the lab to do the experiments

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



← Science

- Matter in Our Surroundings
- Is Matter Around Us Pure
- Atoms and Molecules** ✓
- Structure of the Atom
- The Fundamental Unit of Life
- Tissues
- Diversity in Living Organisms
- Motion



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 Maharishi Kanad, postulated that if we go on dividing matter (*padarth*), 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 *Parmanu*. Another Indian philosopher, Pakudha Katyayama, elaborated this doctrine and said that these particles normally exist in a combined form which gives us various forms of matter.

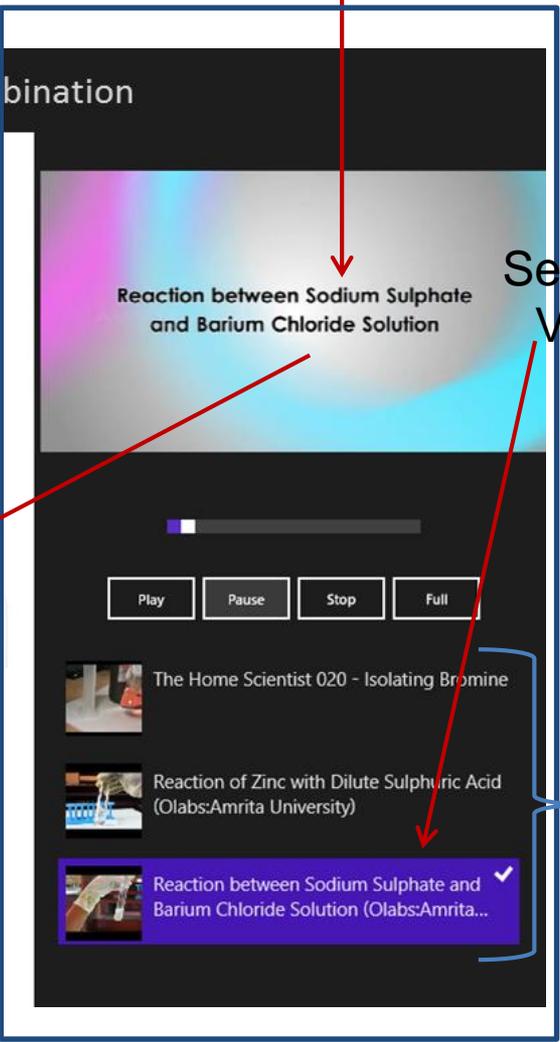
Around the same era, ancient Greek philosophers – Democritus and Leucippus suggested that if we go on dividing matter, a stage will come when particles obtained cannot be divided further. Democritus called these indivisible particles atoms (meaning indivisible). All this was based on philosophical considerations 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

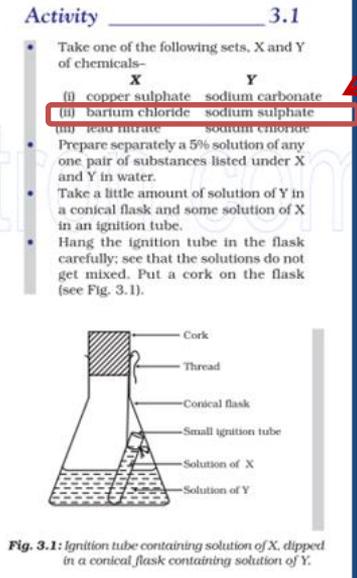


Reaction between Sodium Sulphate and Barium Chloride Solution

Selected Video

Reaction between Sodium Sulphate and Barium Chloride Solution (Olabs:Amrita University) ✓

videos



Activity 3.1

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

X	Y
(i) copper sulphate	sodium carbonate
(ii) barium chloride	sodium sulphate
(iii) lead nitrate	sodium chromate

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).

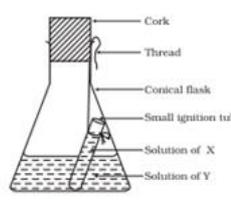


Fig. 3.1: Ignition tube containing solution of X, dipped in a conical flask containing solution of Y.

Video Augmentation: Assist in understanding content

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.

5.2 Magnetism and Matter: Bar Magnet

- ← Physics
- Electric Charges and Fields
- Electrostatic Potential and Capacitance
- Current Electricity
- Moving Charges and Magnetism
- Magnetism and Matter**
- Electromagnetic Induction
- Alternating Current
- Electromagnetic Waves

Magnetism and Matter

We begin our study by examining iron filings sprinkled on a sheet of glass placed over a short bar magnet. The arrangement of iron filings is shown in Fig. 5.2.

The pattern of iron filings suggests that the magnet has two poles similar to the positive and negative charge of an electric dipole. As mentioned in the introductory section, one pole is designated the *North pole* and the other, the *South pole*. When suspended freely, these poles point approximately towards the geographic north and south poles, respectively. A similar pattern of iron filings is observed around a current carrying solenoid.

5.2.1 The magnetic field lines

The pattern of iron filings permits us to plot the magnetic field lines*. This is shown both for the bar-magnet and the current-carrying solenoid in Fig. 5.3. For comparison refer to the Chapter 1, Figure 1.7(d). Electric field lines of an electric dipole are also displayed in Fig. 5.3(c). The magnetic field lines are a visual and intuitive realisation of the magnetic field. Their properties are:

- (i) The magnetic field lines of a magnet (or a solenoid) form continuous closed loops. This is unlike the electric dipole where these field lines begin from a positive charge and end on the negative charge or escape to infinity.
- (ii) The tangent to the field line at a given point represents the direction of the net magnetic field **B** at that point.

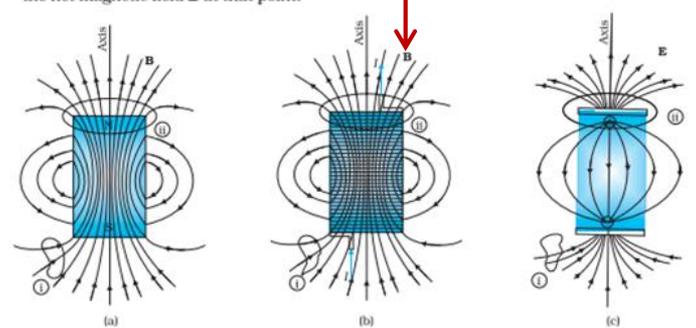


FIGURE 5.3 The field lines of (a) a bar magnet, (b) a current-carrying finite solenoid and (c) electric dipole. At large distances, the field lines are very similar. The curves labelled ⊕ and ⊖ are closed Gaussian surfaces.

$$\vec{B} = \mu_0 \vec{H}$$

Flux \nearrow Field

permeability μ_0/m

$\mu_0 = 4\pi \times 10^{-7}$

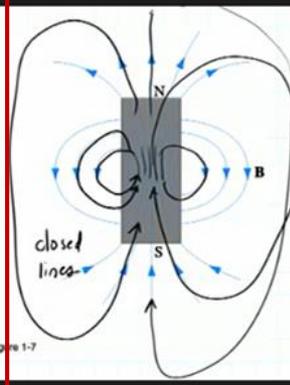


Figure 1-7

Play Pause Stop Full

- ECE3300 Lecture 2-5 Magnetic Fields
- ECE3300 Lecture 2-2 Superposition of charges
- ECE3300 Lecture 2211 Faraday's law

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, ...

References

- [AGK+10] Rakesh Agrawal, Sreenivas Gollapudi, Krishnaram Kenthapadi, Nitish Srivastava, Raja Velu. "[Enriching Textbooks Through Data Mining](#)". [DEV 2010](#).
- [AGK+11a] Rakesh Agrawal, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi. "[Identifying Enrichment Candidates in Textbooks](#)". [WWW 2011](#).
- [AGK+11b] Rakesh Agrawal, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi. "[Enriching Textbooks With Images](#)". [CIKM 2011](#).
- [ACG+12] Rakesh Agrawal, Sunandan Chakraborty, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi. "[Empowering Authors to Diagnose Comprehension Burden in Textbooks](#)". [KDD 2012](#).
- [AGK+13] Rakesh Agrawal, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi. "[Studying from Electronic Textbooks](#)". [CIKM 2013](#).
- [AJK14] Rakesh Agrawal, M. Hanif Jhaveri, and Krishnaram Kenthapadi. "[Evaluating Educational Interventions at Scale](#)". [LAS 2014](#).
- [ACG+14] Rakesh Agrawal, Maria Christoforaki, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi, Adith Swaminathan. "[Augmenting Textbooks with Videos](#)". [ICFCA 2014](#).
- [AGT14] Rakesh Agrawal, Behzad Golshan, Evimaria Terzi. "[Grouping Students in Educational Settings](#)". [KDD 2014](#).