ICADL 2013

Microsoft Research Silicon Valley

Computational Education: The Next Frontier for Digital Libraries?

Rakesh Agrawal

Microsoft Technical Fellow

Microsoft Research, Mountain View, California

December 9, 2013

Microsoft Research

15th Int'l Conf. on Asia-Pacific Digital Libraries

Outline

- What has changed?
- The central role of technology in rethinking education!
- Whither digital libraries?

Outline

- What has changed?
- The central role of technology in rethinking education!
- Whither digital libraries?

Tom Friedman: Flat World 4.0*

When I wrote "The world is flat" in 2004:

- ✓ Facebook didn't have an index entry
- ✓ Twitter was a sound,
- ✓ Cloud was still in the sky
- ✓ 4G was a parking place
- ✓ Linked-in was a prison
- ✓ Application was what you sent to college
- ✓ Big data was a rapper
- ✓ Skype was a typo

Something really big just happened in last 7-8 years

*The Role of Education in Hyper-Connected, Global World. Teach For All Global Conf. Oct. 2013.

Thinking About Education

Three key questions:

- What is being taught
  - Curriculum, syllabus, educational material
- How it is being delivered
  - Teachers, classes, assessments
- How it is funded
  - Business models

Emergent Perfect Storm

- Electronic textbooks
  - Fast adoption of cloud-connected electronic devices (worldwide)
  - Open content (e.g. OpenStax, ck12.org, NCERT, Crowdsourcing)
- Internet-based classes
  - MOOCs (e.g. Coursera, EdX, Udacity, Khan, TED-Ed)
  - Small virtual classes (e.g. Shankar Mahadevan Academy)
  - Electronic certification (e.g. Mozilla's OpenBadges)
- New models of funding education
  - Recipients give back to the seed fund for future recipients at their pace (e.g. Dakshana)
  - Market for options on future earnings (e.g. Oregon legislation)

Outline

- What has changed?
- The central role of technology in rethinking education!
- Whither digital libraries?

Data Mining for Enriching 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+13] 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)

Microsoft Research

Joint work with Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi, *et al.*

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.

[ACG+13] Rakesh Agrawal, Maria Christoforaki, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi, Adith Swaminathan. "Augmenting Textbooks with Videos". Under submission, 2013.

[AGT13] Rakesh Agrawal, Behzad Golshan, Evimaria Terzi. "Forming Beneficial Teams of Students in Massive Online Classes", Under submission, 2013.

Data Mining for Enriching 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+13] based on the focus of the section

References to prerequisites

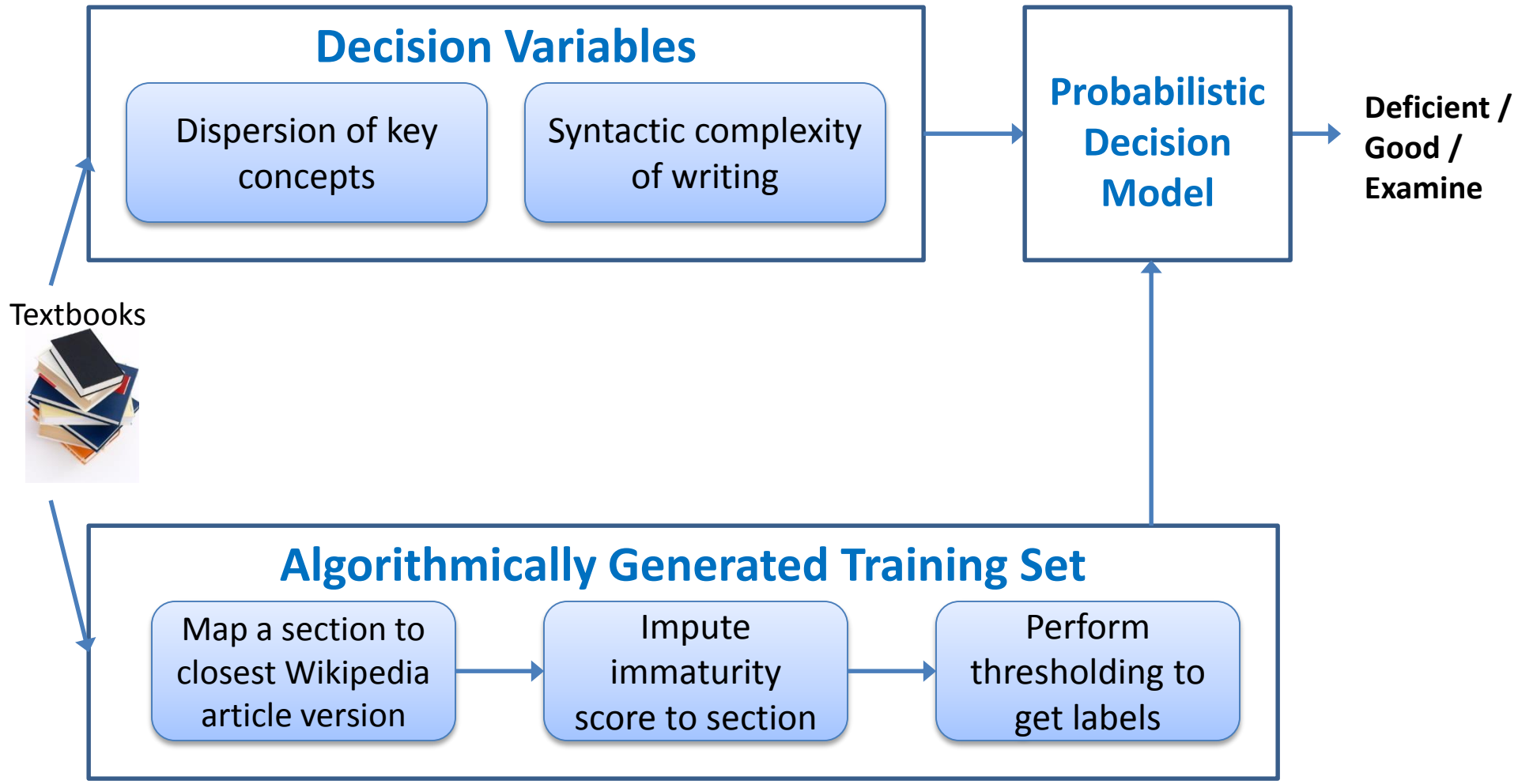
Links to concepts necessary for understanding the present section, derived using a model of a how students read textbooks [AGK+13]

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

Microsoft Research

Joint work with Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi, *et al.*

Identification of Deficient Sections

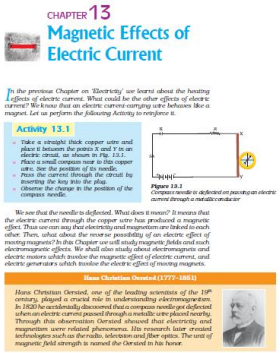


Dispersion of Key Concepts

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*) = *C*oncept *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 – Edge Density) of the concept graph

A Tale of Two Sections


CHAPTER 13
Magnetic Effects of Electric Current

In the previous Chapter on 'Electricity' we learnt about the heating effects of electric current. What could be the other effects of electric current? We know that an electric current-carrying wire behaves like a magnet. Let us perform the following Activity to reinforce it.

Activity 13.1

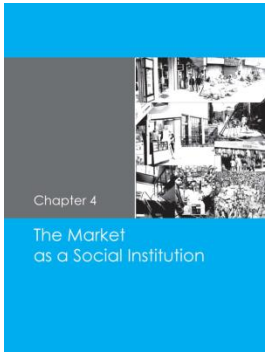
- Take a straight thick copper wire and place it between the points X and Y in an electric circuit, as shown in Fig. 13.1.
- Place a small compass in this copper wire near one of its ends.
- Pass the current through the circuit by inserting the key into the plug.
- Observe the change in the position of the compass needle.

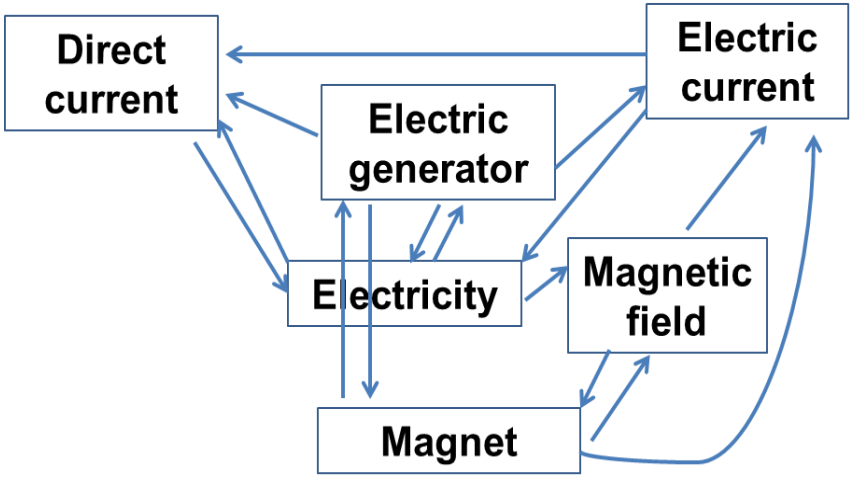
Figure 13.1 Compass needle is deflected on passing an electric current through a metallic conductor

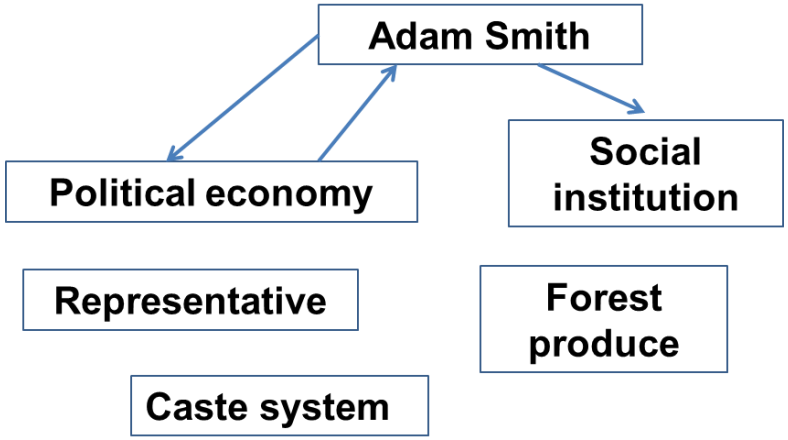
We see that the needle is deflected. What does it mean? It means that the electric current through the copper wire has produced a magnetic effect. Thus we can say that electricity and magnetism are linked to each other. Then, what about the reverse possibility of an electric effect of moving magnets? In this Chapter we will study magnetic fields and such electromagnetic effects. We shall also study about electromagnets and electric motors which involve the magnetic effect of electric current, and electric generators which involve the electric effect of moving magnets.

Hans Christian Oersted (1777–1851)

Hans Christian Oersted, one of the leading scientists of the 19th century, played a crucial role in understanding electromagnetism. In 1820 he accidentally discovered that a compass needle got deflected when an electric current passed through a metallic wire placed nearby. Through this observation Oersted showed that electricity and magnetism were related phenomena. His research later created technologies such as the radio, television and fiber optics. The unit of magnetic field strength is named the Oersted in his honor.







Dispersion = 1 – 15/30 = 0.5

Dispersion = 1 – 3/30 = 0.9

Larger dispersion ➔ Harder to understand section

Readability Formulas

- 100+ years of readability research
- 200+ Readability formulas
  - In widespread use (notwithstanding limitations)
- Popular formulas:

Flesch Reading Ease Score [17]	206.835	−	84.6	×	S/W	−	1.015	×	W/T
Flesch-Kincaid Grade Level [31]	−15.59	+	11.8	×	S/W	+	0.39	×	W/T
Dale-Chall Grade Level [14]	14.862	−	11.42	×	D/W	+	0.0512	×	W/T
Gunning Fog Index [23]			40	×	C/W	+	0.4	×	W/T
SMOG Index [37]	3.0	+	$\sqrt{30}$	×	$\sqrt{C/T}$				
Coleman-Liau Index [10]	−15.8	+	5.88	×	L/W	−	29.59	×	T/W
Automated Readability Index [46]	−21.43	+	4.71	×	L/W	+	0.50	×	W/T

C	=	Number of words with three syllables or more
D	=	Number of words on the Dale Long List
L	=	Number of letters
S	=	Number of syllables
T	=	Number of sentences
W	=	Number of words

- Regression coefficients learned over specific datasets
  - McCall-Crabbs Standard Test Lessons

Syntactic Complexity

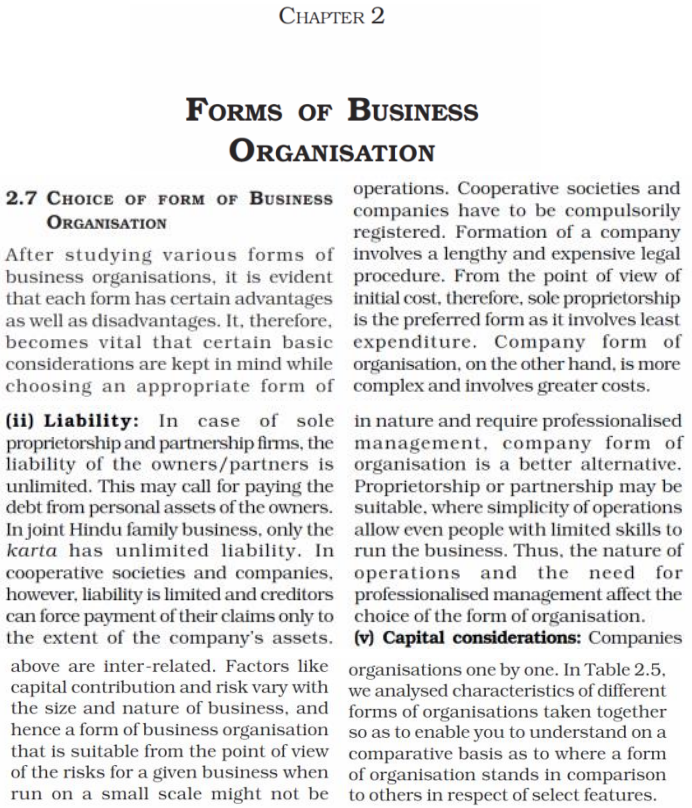
- Direct use of *Readability formulas* yielded poor results
- Variables abstracted from readability formulas:
  - Word length: Average syllables per word (S/W)
  - Sentence length: Average words per sentence (W/T)
- Larger syntactic complexity ➔ Harder to understand

Aakash Prototype

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

Microsoft Research

Illustrative Result: Deficient Section



CHAPTER 2

FORMS OF BUSINESS ORGANISATION

2.7 CHOICE OF FORM OF BUSINESS ORGANISATION

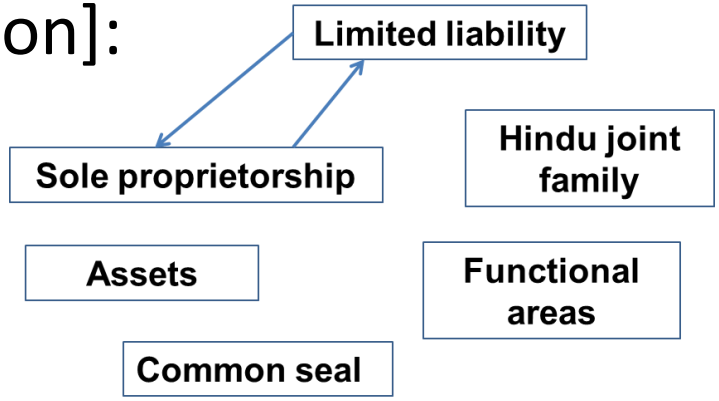
After studying various forms of business organisations, it is evident that each form has certain advantages as well as disadvantages. It, therefore, becomes vital that certain basic considerations are kept in mind while choosing an appropriate form of operations. Cooperative societies and companies have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

(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. 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

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 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.*

Microsoft Research

Data Mining for Enriching 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+13] based on the focus of the section

References to prerequisites

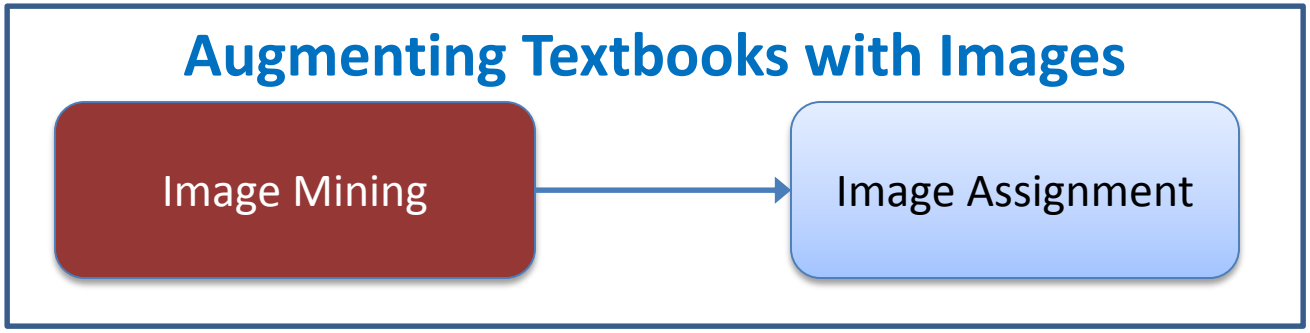
Links to concepts necessary for understanding the present section, derived using a model of a how students read textbooks [AGK+13]

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

Microsoft Research

Joint work with Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi, *et al.*

Comity



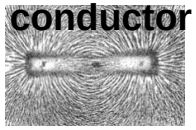
- Intuition: Combine results of a large number of short, but relevant queries
  - Search engines barf on long queries (such as entire section content)
- Identify key concepts present in a section, *C*
- Form two-concept and three-concept queries, *Q*
- For each $q \in Q$, obtain ranked list of images $I(q)$ using image search
- Relevance score(*i*) of image i =

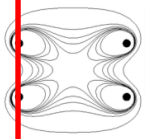
  $\sum_q f$(position of image in $I(q)$, importance of concepts in q)

From Section Level to Book Level Image Assignments

BEFORE IMAGE ASSIGNMENT

Sec 2: Magnetic field due to a current carrying conductor


Magnetic effect

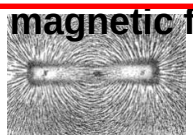

Helmholtz Contour

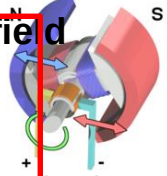

Solenoid


Amperemeter

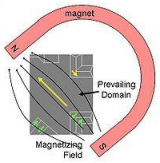

Galvanometer

Sec 3: Force on a current carrying conductor in a magnetic field


Magnetic effect



Electric motor cycle

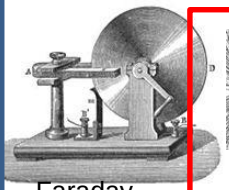


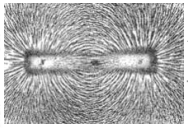
Effect of magnet on domains

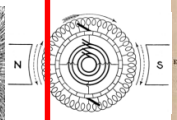

Meissner Effect


Descartes' magnetic field

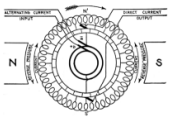
Sec 6: Electric generator


Faraday disk generator


Magnetic effect

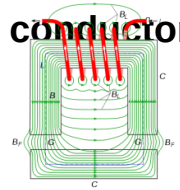

Two phase rotary converter

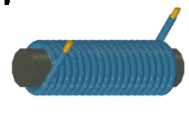

Descartes' magnetic field


Single phase rotary converter

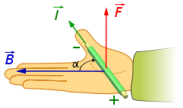
AFTER IMAGE ASSIGNMENT

Sec 2: Magnetic field due to a current carrying conductor


Magnetic field


Simple electromagnet

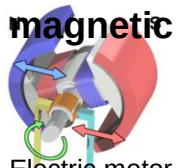
Right hand rule

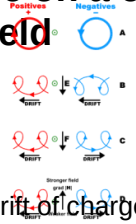


Right hand rule

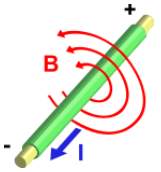

Solenoid

Sec 3: Force on a current carrying conductor in a magnetic field

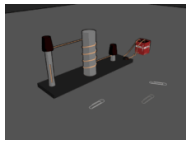

Electric-motor cycle exploits electro magnetism

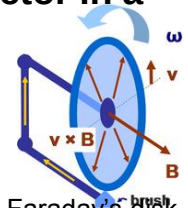


Drift of charged particles



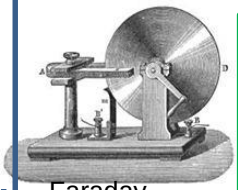
Magnetic field around current

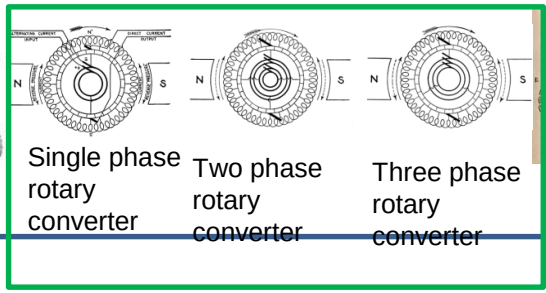

Electromagnets attract paper clips….



Faraday's disk electric generator

Sec 6: Electric generator


Faraday disk generator



Single phase rotary converter

Two phase rotary converter

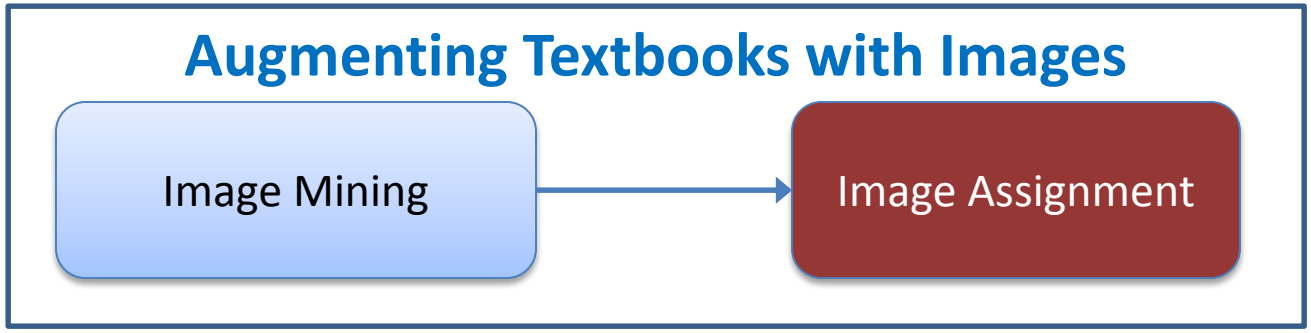
Three phase rotary converter


Descartes' magnetic field

Same image can repeat across sections!

Richer set of images to augment the section

Microsoft Research



MaxRelevantImageAssignment

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

Relevance score of image i to section j

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

s.t.

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

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

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

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

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

Constraint: An image can belong to at most one section

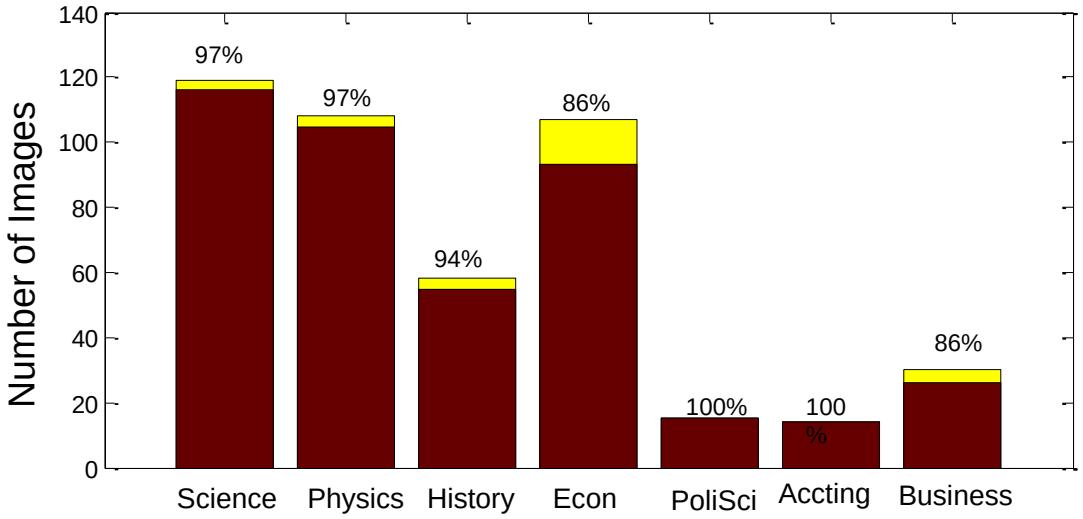
Can be solved optimally in polynomial time

Evaluation on NCERT Textbooks

User-study employing Amazon Mechanical Turk

- HIT: a given image helpful for understanding the section?

The number above a bar indicate helpfulness index for the corresponding subject (% of images found helpful)



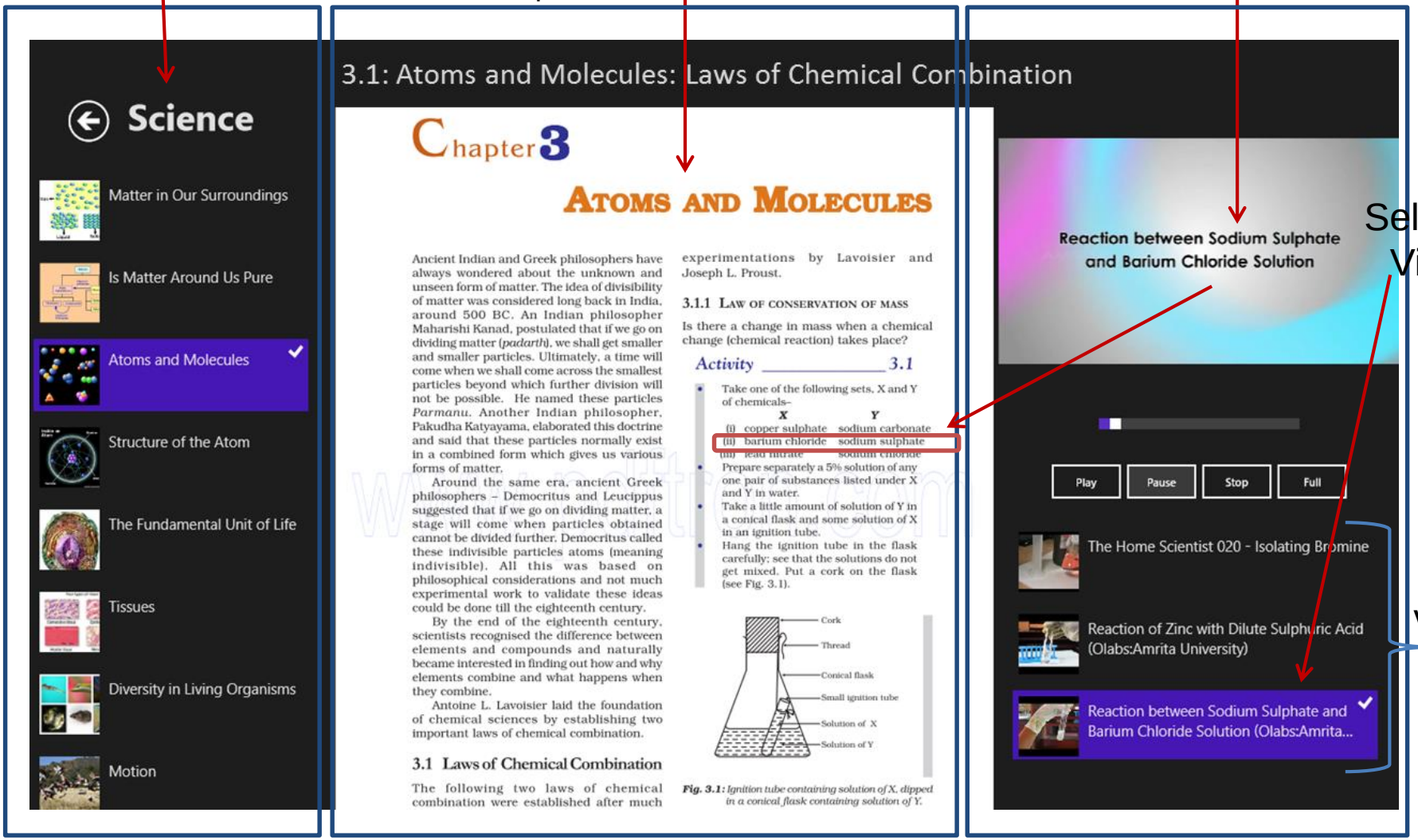
- 94% of images deemed helpful
- Performance maintained across subjects

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



Selected Video

videos

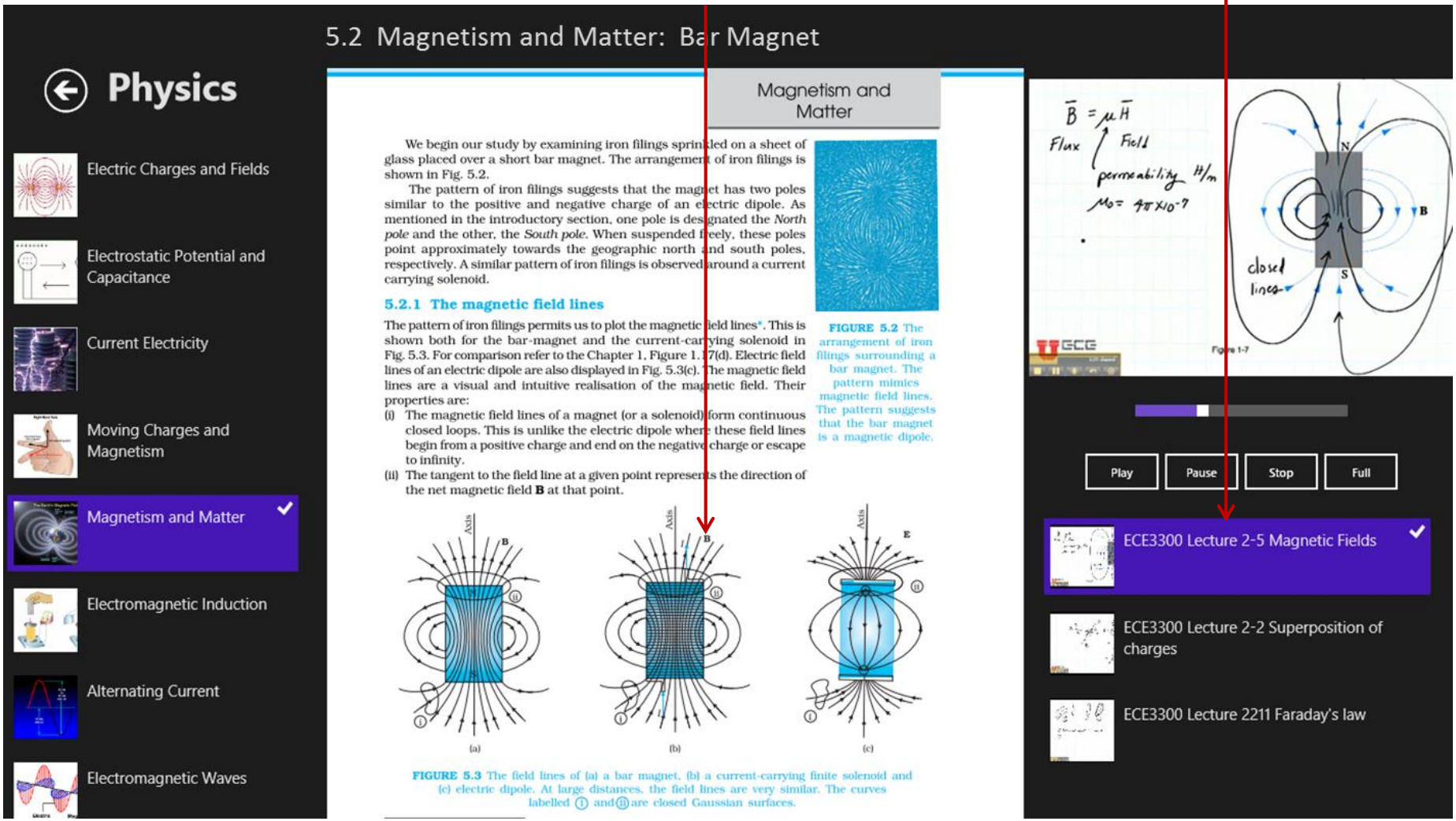
Microsoft Research

Win8 Surface Prototype

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.



Microsoft Research

Win8 Surface Prototype

Data Mining for Enriching 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+13] 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)

Microsoft Research

Joint work with Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi, *et al.*

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.

[ACG+13] Rakesh Agrawal, Maria Christoforaki, Sreenivas Gollapudi, Anitha Kannan, Krishnaram Kenthapadi, Adith Swaminathan. "Augmenting Textbooks with Videos". Under submission, 2013.

[AGT13] Rakesh Agrawal, Behzad Golshan, Evimaria Terzi. "Forming Beneficial Teams of Students in Massive Online Classes", Under submission, 2013.

Outline

- What has changed?
- The central role of technology in rethinking education!
- Whither digital libraries?

The DELOS Manifesto

D-Lib Magazine, 14(3), March/April 2007

Digital Library: Tool at the center of intellectual activity

- From content-centric system supporting the organization and provision of access to particular collections → person-centric system delivering innovative, evolving, and personalized services
- From static storage and retrieval of information → facilitation of communication, collaboration, and other forms of dynamic interaction
- From handling centrally located text → synthesizing distributed multimedia document collections and pervasive computing services

A Sampling of Recent ICADL Papers

- Benjamin Köhncke. Bridging the Gap - Using External Knowledge Bases for Context-Aware Document Retrieval. ICADL 2013.
- Z. Zhou et al. SRec: an Automatic Slide Capturing and Sharing System. ICADL 2013.
- K. Rachert et al. An Architecture for Community-based Curation and Presentation of Complex Digital Objects. ICADL 2013.
- Kanyacome et al. Needs of Collaborative Digital Library for Secondary School Students in Thailand. ICADL 2012.
- A. Motoki et al. The Relation between Comments inserted onto Digital Textbooks by Students and Grades earned in the Course. ICADL 2010.
- A. Gerber et al. A Collaborative Scholarly Annotation System for Dynamic Web Documents - a Literary Case Study. ICDAL 2010.

Need for Refocused Efforts

- Broadly-applicable specialization is valuable
  - Key-word driven document retrieval ≠ Query-by-document ≠ Textbook augmentation
- Transformative changes in underlying assumptions demand rethink of solution approaches


The Oakman Automobile (1899)

- The framework changes with new technology, not just the picture within the frame – Marshall McLuhan

Some Research Ideas

- 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
- Personalized learning plans
- Dynamic formation of classes and study groups
- Performance evaluation methodologies and benchmarks

Magic happens when what is desperately needed meets what is technically feasible

Your Questions and Comments