



Microsoft
Research Silicon Valley

Mining Videos from the Web for Electronic Textbooks

R. Agrawal, M. Christoforaki, S. Gollapudi,
A. Kannan, K. Kenthapadi, & A. Swaminathan

Microsoft Research – Silicon Valley

June 11, 2014

Outline

- Motivation
- Approach
- Experiments
- Concluding Remarks

Outline

- Motivation
- Approach
- Experiments
- Concluding Remarks

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+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]

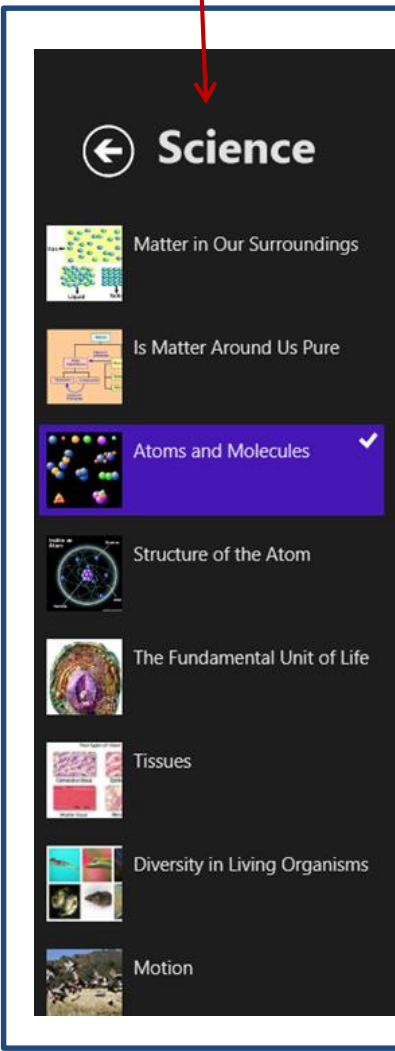
- Validation on textbooks from U.S.A and India, on different subjects, across grades
- Prototypes and research papers (see [References](#))

Video Augmentation: Make inaccessible accessible

Table of contents for navigating the book (automatically extracted)

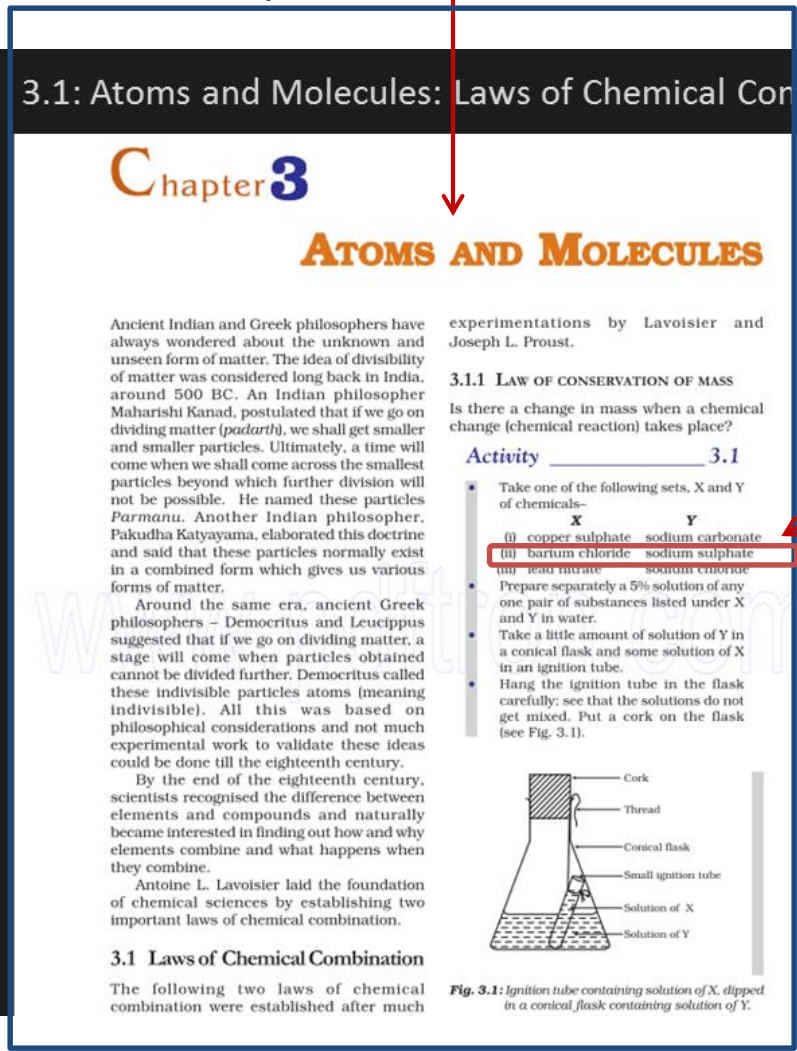
Re-rendered section: This section, about the laws of chemical combination, prescribes an activity for the chemistry lab, but the school might lack the lab to do the experiments

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.

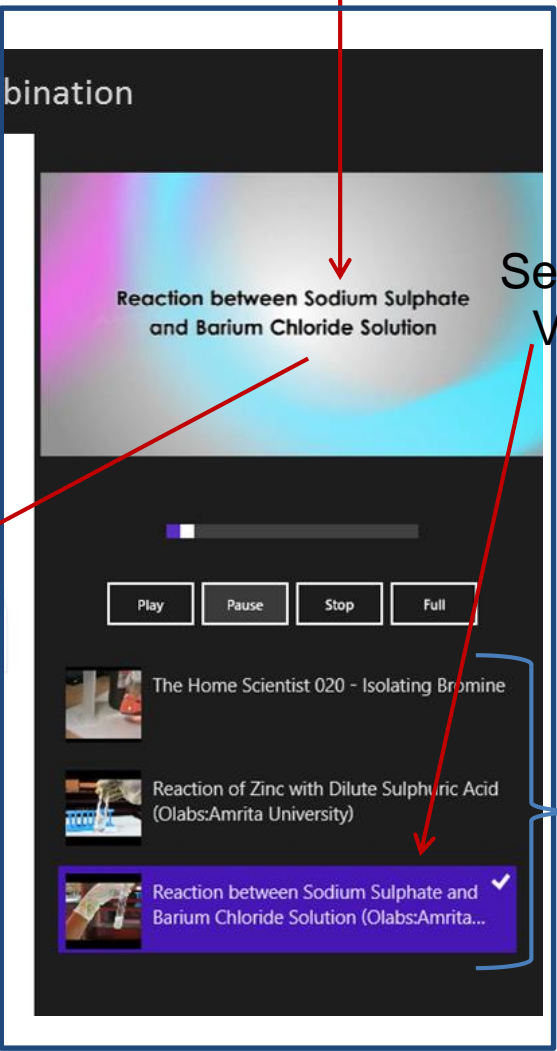
Around the same era, ancient Greek philosophers – Democritus and Leucippus suggested that if we go on dividing matter, a stage will come when particles obtained cannot be divided further. Democritus called these indivisible particles atoms (meaning indivisible). All this was based on philosophical considerations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

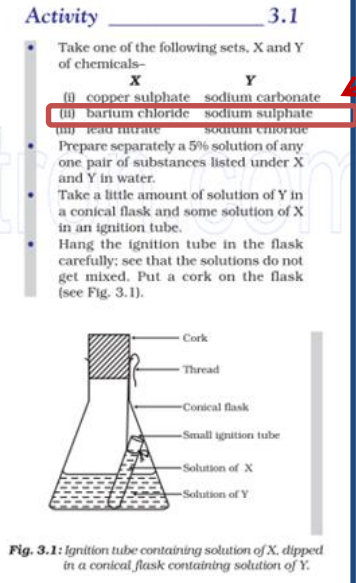


Reaction between Sodium Sulphate and Barium Chloride Solution

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).

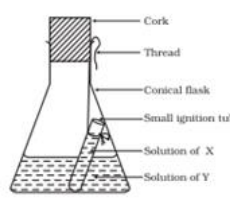


Fig. 3.1: Ignition tube containing solution of X, dipped in a conical flask containing solution of Y.

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 1.17(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.

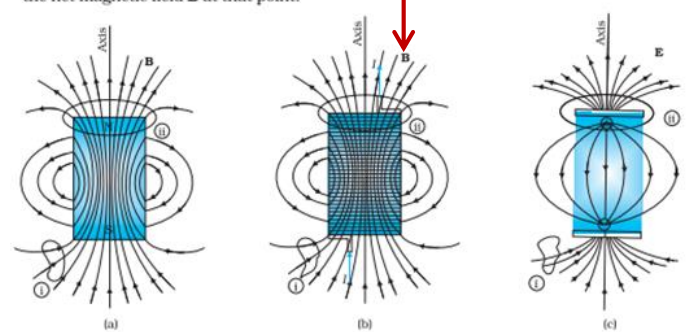


FIGURE 5.3 The field lines of (a) a bar magnet, (b) a current-carrying finite solenoid and (c) electric dipole. At large distances, the field lines are very similar. The curves labelled ⊕ and ⊖ are closed Gaussian surfaces.

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

Flux \nearrow Field

permeability μ_0/m

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

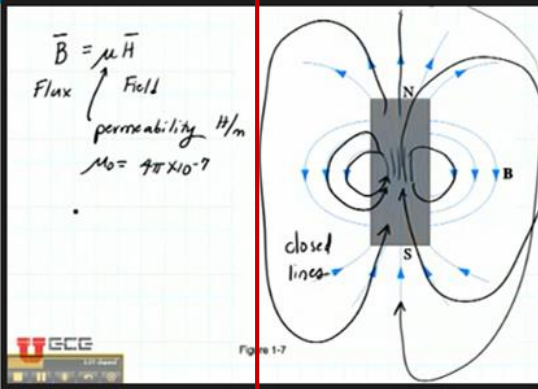


Figure 1-7

ECE330 Lecture 2-5 Magnetic Fields

ECE330 Lecture 2-2 Superposition of charges

ECE330 Lecture 2211 Faraday's law

Outline

- Motivation
- **Approach**
- Experiments
- Concluding Remarks

Well-Written Textbooks

Two organizational principles [GL35, CC98]

[1] Focus:

- Each section explains a few concepts

[2] Unity:

- For each concept, there is a unique section that explains the concept

Focus = Unique contribution of the section

Methodology

- Compute focus of every section of the textbook
- Crawl the Web for educational videos [CBD99, S08]
- Obtain transcripts of the videos [SCF+06]
- Score transcripts with respect to the focus of the section

Methodology

- Compute focus of every section of the textbook
- Crawl the Web for educational videos [CBD99, S08]
- Obtain transcripts of the videos [SCF+06]
- Score transcripts with respect to the focus of the section

Seeking Focus

- Formal concept: Unit of human thought [W05]
- Focus of section $s = \{\text{Formal concept}\}$ in s !
- Formal concepts being abstract, they manifest in the form of concept phrases, *cphrs*
- Extent = {Paragraph}, Intent = {*cphr*} in s
- A formal concept of s : A pair of maximal set of paragraphs P_c and maximal set of *cphrs* C s.t. every *cphr* in C is present in every paragraph in P_c
- $FC1 = \langle \{P1, P2, P3\}, \{x, y\} \rangle$, $FC2 = \langle \{P3\}, \{x, y, z\} \rangle$.
Now what?

Refinement

- Iceberg lattice [S+02]: Robust to small perturbations
- A concept is frequent if at least μ objects individually have every attribute in the intent of the context
- Iceberg concept lattice = partial order over frequent concepts
- Use the leaf nodes of the iceberg concept lattice to represent focus (most specific sets of *cphrs*, or equivalently maximal combinations of *cphrs*)!

Focus of a Section

cphr: A concept phrase present in the text

Indicium of a section s : Set of *cphrs* $C \in s$ s.t.

- C is frequent in s
- No section s' in which C is frequent
- No superset of C is frequent in s

Focus of a section s : Set of indicia present in s

Indicium & Class Hypothesis

- Hypothesis for a class [K04]: A closed itemset occurring in a class but not in other classes
- Minimal hypothesis: Inclusion minimal hypothesis
- Indicium = Maximal frequent (and hence closed) itemset in a class (text section), but infrequent in other classes (sections)
- Indicium = Relaxation of a minimal hypothesis, admitting infrequent presence in other classes

Indicia from Two Physics Sections

(a) Magnetism & Gauss' Laws

field line, magnetic field, monopole

field line, magnet, charged particle

electrostatics, field line, monopole

Section on general magnetism discussing magnetic fields and their effect on moving particles

'Monopole': hypothetical particle that is an isolated magnet with one pole; distinguishes this section on general magnetism from section on earth' magnetism

(b) Earth's Magnetism

field line, magnetic field, earth

equator, meridian, southern hemisphere

earth, solar wind, poles

Discusses how earth acts as magnet

Earth has both poles

'Earth' is generic term but indicium formed by combining it with 'field line' and 'magnetic field' very pertinent for the section

Methodology

- Compute focus of every section of the textbook
- Crawl the Web for educational videos [CBD99, S08]
- Obtain transcripts of the videos [SCF+06]
- **Score transcripts with respect to the focus of the section**

Indicia of Transcripts?

- Videos produced independently and with different viewpoints
- Lack focus and unity properties (Not surprising)

Scoring Transcripts

- Candidate Video V = Transcript of V contains all *cphrs* from at least one indicium
- $\text{Score}(V) = \sum_j \text{significance}(\text{indiciu}_j)$, where summation is over indicia present in V
- Significance of an indicium is a function f of the importance of underlying *cphrs*
 - f must have monotonicity and concentration
 - $f(C) = \sum_{c \in C} \exp(\text{importance}(c))$
- $\text{Importance}(c) =$ probability that c is hyperlinked to the corresponding Wikipedia article [M+07, M+09]

Outline

- Motivation
- Approach
- **Experiments**
- Concluding Remarks

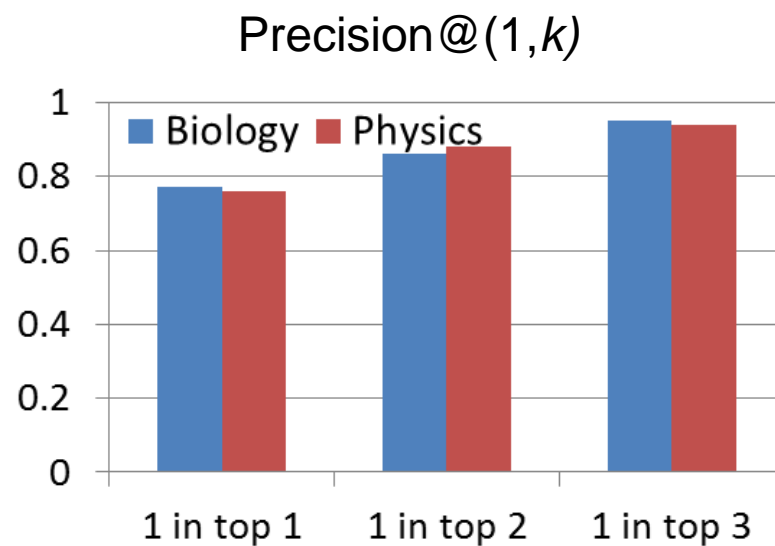
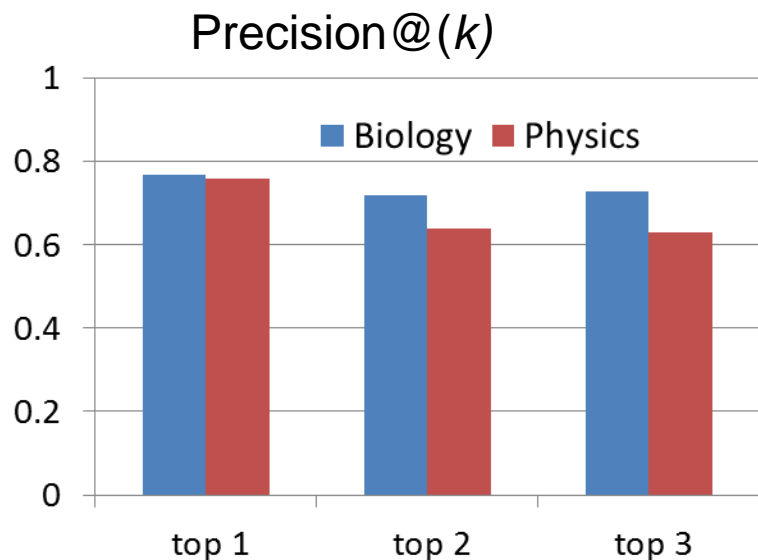
Data Sets

- Text corpus:
 - Books published by ck12.org
 - Books published by NCERT, India
- Video corpus:
 - 50,000 education-related, short videos obtained from focused crawl seeded with educational videos from a few reputed sites
 - Limited to videos that had accompanying user-uploaded transcript of the video content

Precision Study - Setup

- HIT: Read a section, watch a video, judge if the video is relevant (Default: Not Relevant)
- Each HIT judged by seven judges (Judge required to spend minimum 30 mins on a HIT)
- Precision@ k : Fraction of videos in top k positions judged to be relevant
- Precision@(i, k): Whether at least i of the videos in top k positions were found relevant

Precision Study - Results

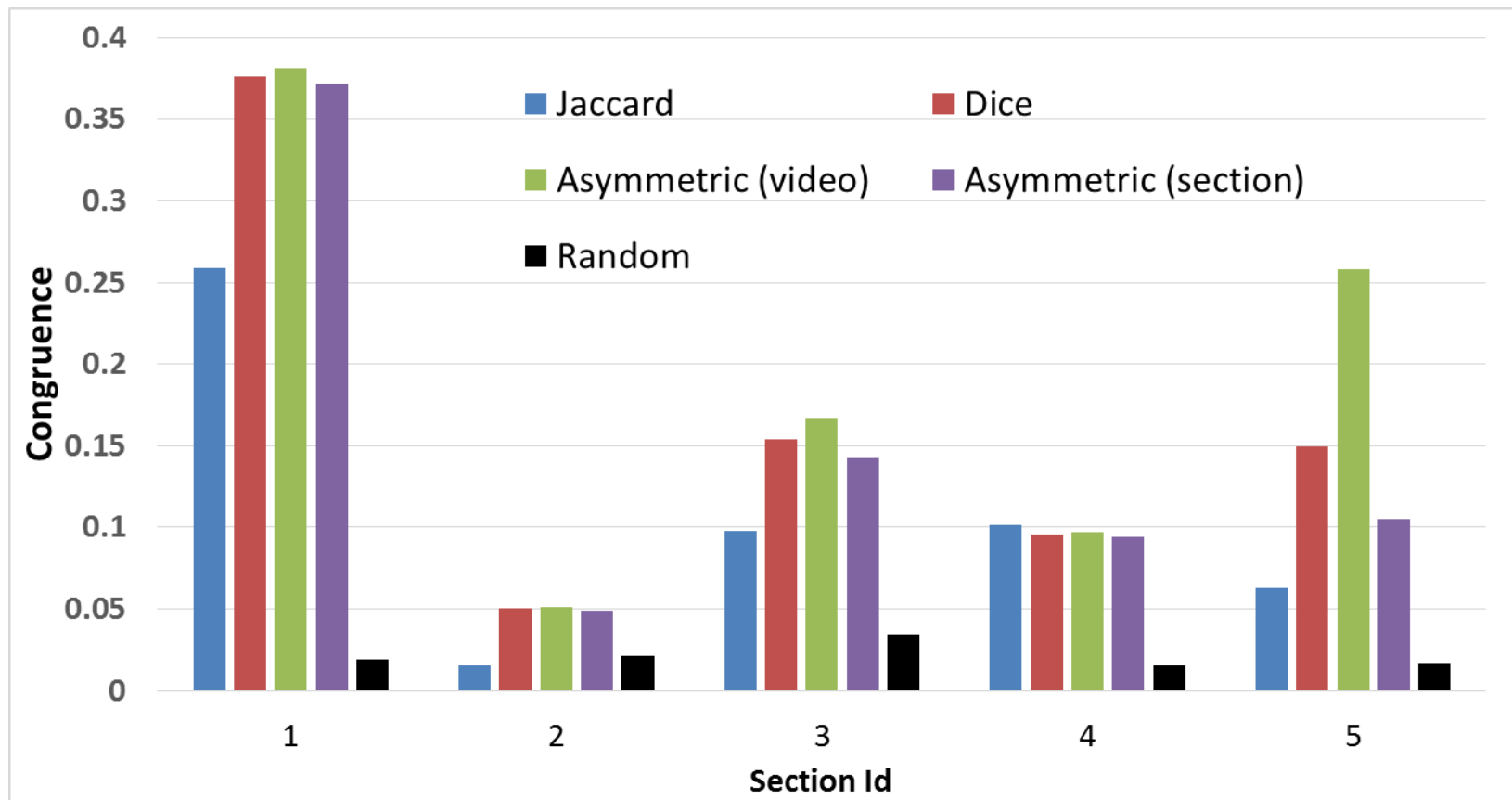


Congruence Study - Setup

- SectionHIT (videoHit): read section, provide 5 phrases that best describe the section (video)
- Each {section, video} pair given to 5 judges (Minimum 30 mins on a HIT)
- Convert phrases into unigrams (remove stop words)
- Performance metric: symmetric as well as asymmetric similarity measures wrt section and video (Jaccard, Dice)

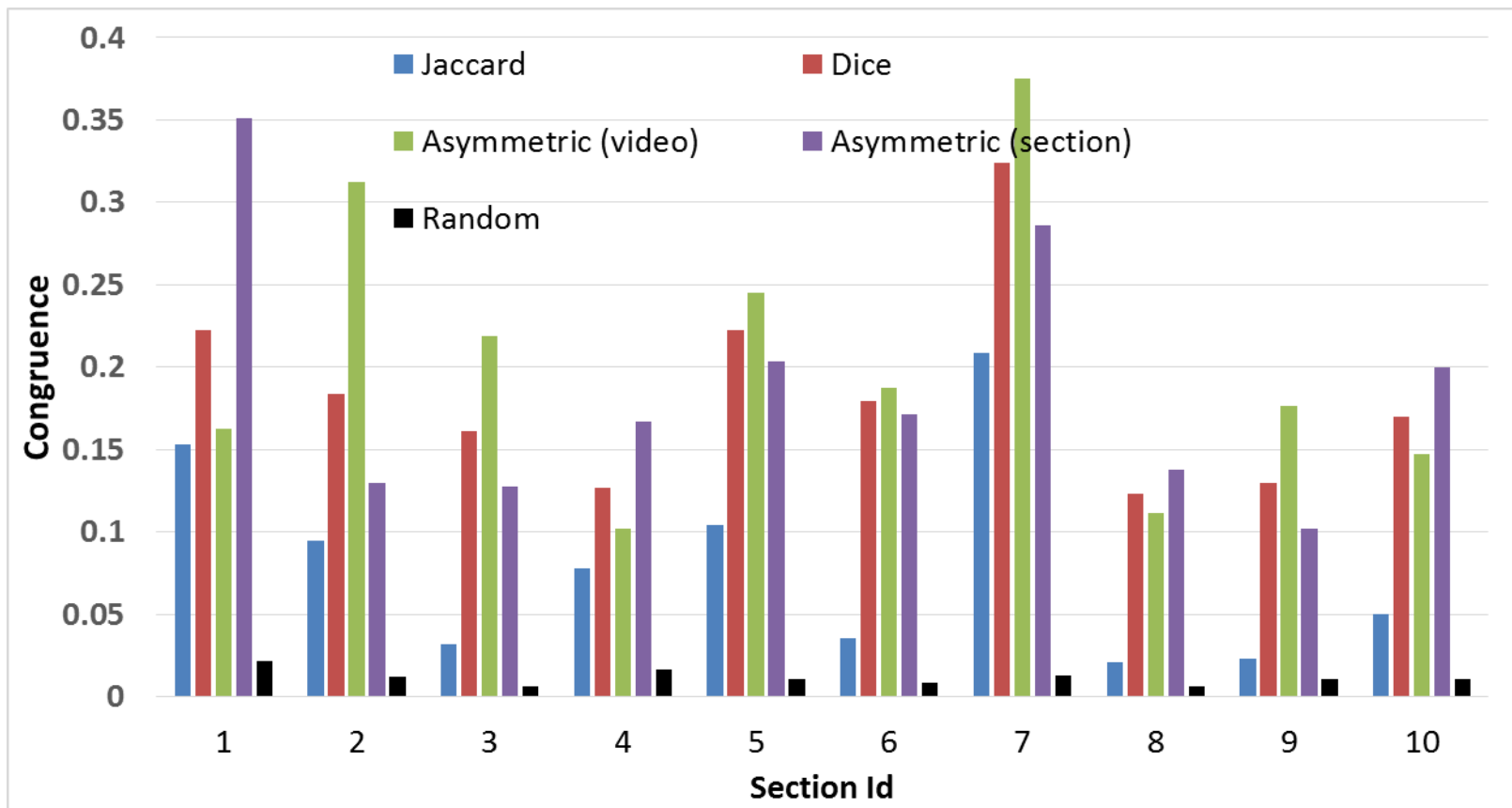
Congruence Study - Results

Physics Book



Congruence Study - Results

Biology Book



Illustrative Results

Magnetism 2



Khan Academy · 4,204 videos

Subscribe

1,781,710

239,035 views

419 3

Like

About

Share

Add to



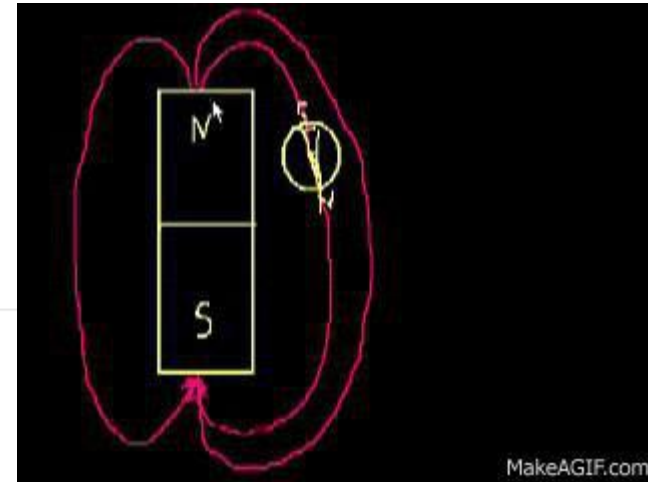
Uploaded on Jul 30, 2008

More free lessons at: <http://www.khanacademy.org/video?v=Nn...>

Magnetic fields and their effects on moving electrical charges

<charged particle, dipole, field line>,

< charged particle, field line, magnetism, monopole >,...



MakeAGIF.com

"Can you feel a solar wind?" (Ask an Astronomer)



NASA Spitzer · 93 videos

Subscribe

6,350

2,823 views

12 1

Like

About

Share

Add to



Uploaded on Jul 18, 2008

Dr. Robert Hurt explains what a solar wind is, and how it affects us here on Earth.

<solar wind, earth, magnetic field>,

<solar wind, earth, poles>,...



MakeAGIF.com

Outline

- Motivation
- Approach
- Experiments
- Concluding Remarks

Technical Report

- Inadequacy of alternate approaches to represent focus
 - TF/IDF
 - LDA
- Properties of scoring function
- More experimental results

Lessons Learnt

- Even widely used books might not follow principles of a well-written book
- An author's sectioning and paragraphing decisions might be arbitrary
- Concept structure in the transcripts can be important
- Evaluation methodology and performance metrics are not easy to come by
- Gap between theory and practice

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](#).