EdRank: ranking webpages by their educational efficacy

Joseph E. Beck

Caution!

- Idea is less fully formed than the others
  - Please pick apart and improve
- No associated papers
- Just some grant proposals

ASSISTments + web pages = WEBsistments

- ASSISTments: web-based ITS
  - Tens of thousands of student users
  - Primarily used for math education, but is independent with respect to curriculum

- Have augmented it with web pages to teach students

How to find effective web pages

1. Do a web search on pages that teach the topic
2. Screen out pages that don't seem useful
3. Students and teachers rate candidate pages
4. Use ratings and data on student performance to find good web pages

How to find effective web pages

1. Do a web search on pages that teach the topic
2. **Screen out pages that don't seem useful**
3. **Students and teachers rate candidate pages**
4. **Use ratings and data on student performance to find good web pages**

- Isn't google good? Why do we need **steps 2-4**?

What web search algorithms do

- Find *popular* content
  - Analyze links that point to a page
  - See how popular they are
  - Pages with important links pointing are important
- Popular and effective are not the same
  - If they were, would have no need for educational research as everyone would be using the right approach

Goal: improve step #1

1. **Do a web search on pages that teach the topic**
2. Screen out pages that don't seem useful
3. Students and teachers rate candidate pages
4. Use ratings and data on student performance to find good web pages

Why do we want to improve search? Simplifies steps 2 - 4

1. Do a web search on pages that teach the topic
2. **Screen out pages that don't seem useful**
3. **Students and teachers rate candidate pages**
4. **Use ratings and data on student performance to find good web pages**

Screening out off-point pages: not time consuming, but hard

1. Do a web search on pages that teach the topic
2. **Screen out pages that don't seem useful**
3. Students and teachers rate candidate pages
4. Use ratings and data on student performance to find good web pages

Don’t want to waste people’s time

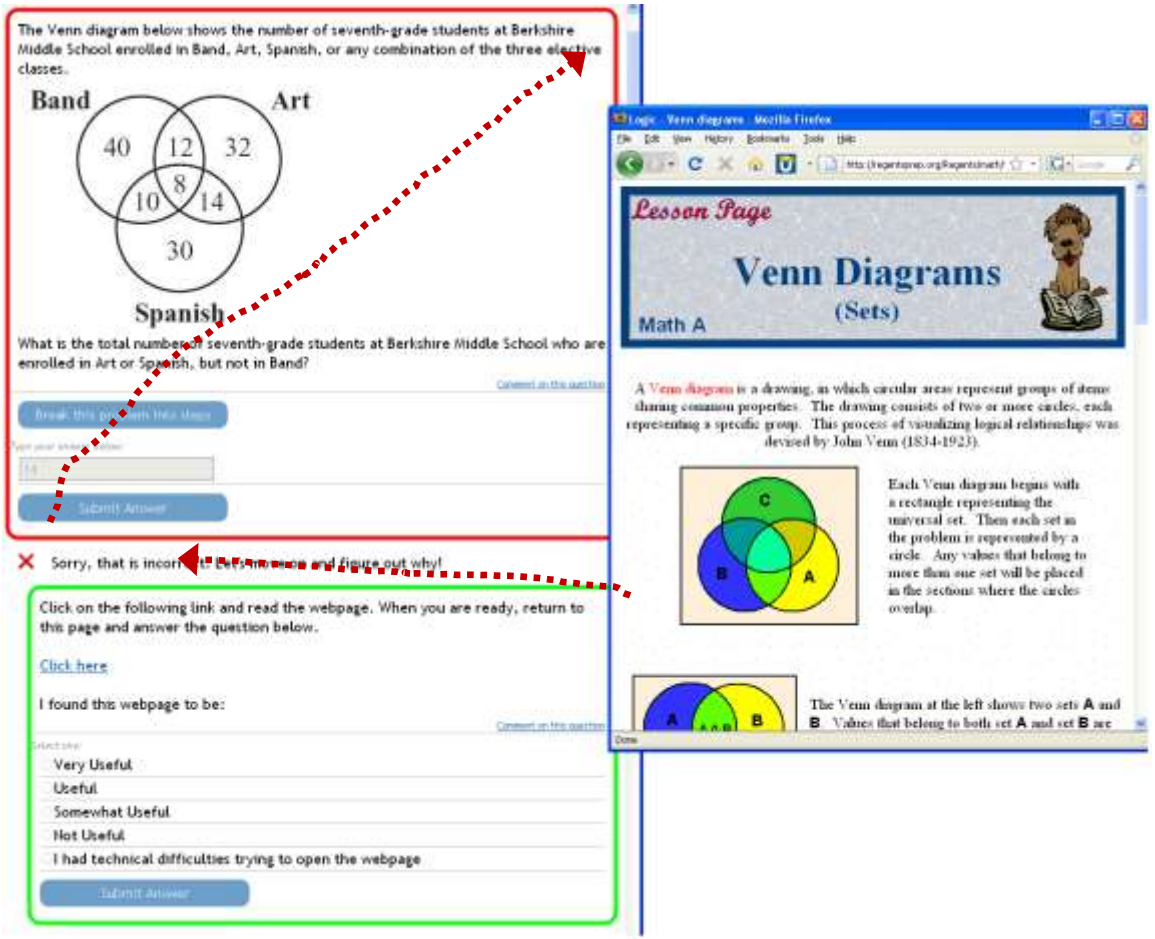
1. Do a web search on pages that teach the topic
2. Screen out pages that don’t seem useful
3. **Students and teachers rate candidate pages**
4. Use ratings and data on student performance to find good web pages

Experimental ethics

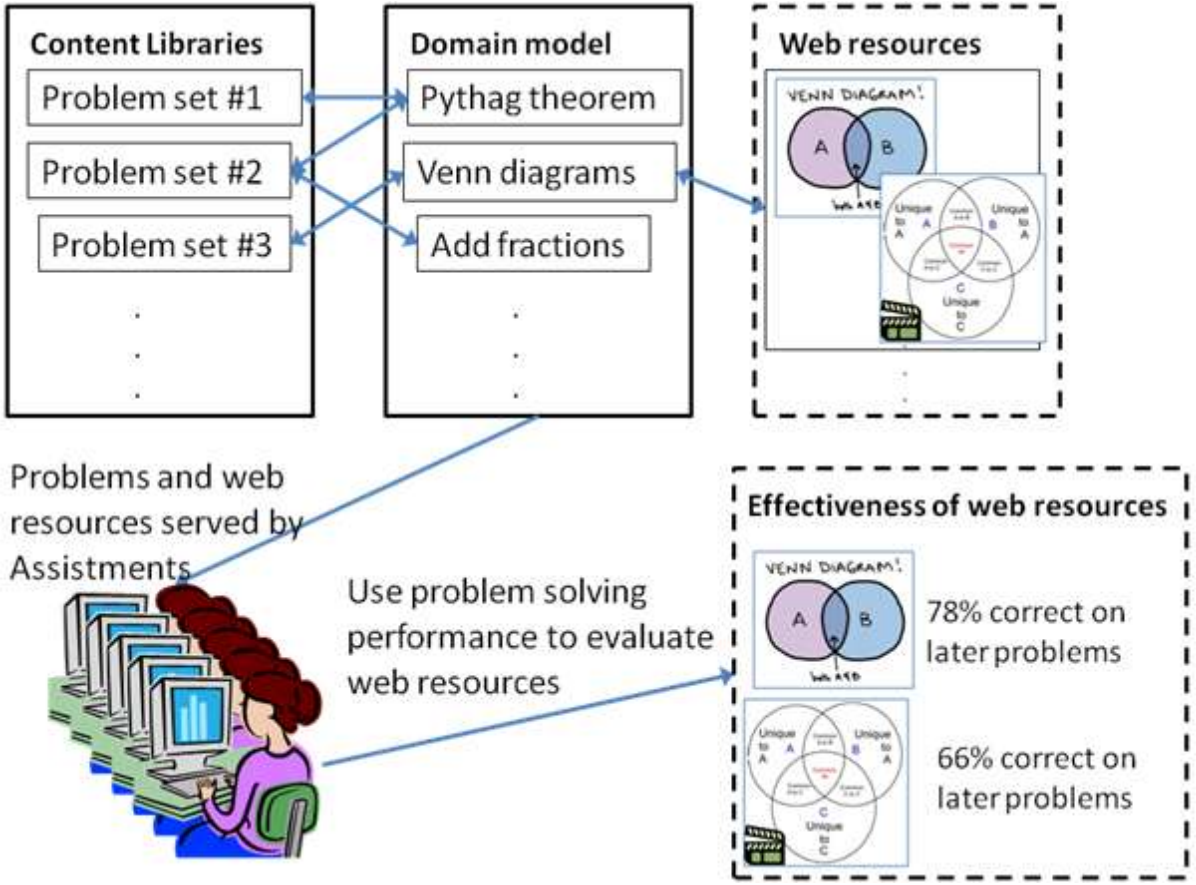
1. Do a web search on pages that teach the topic
2. Screen out pages that don't seem useful
3. Students and teachers rate candidate pages
4. **Use ratings and data on student performance to find good web pages**

Really don't want students spending their time on ineffective web pages

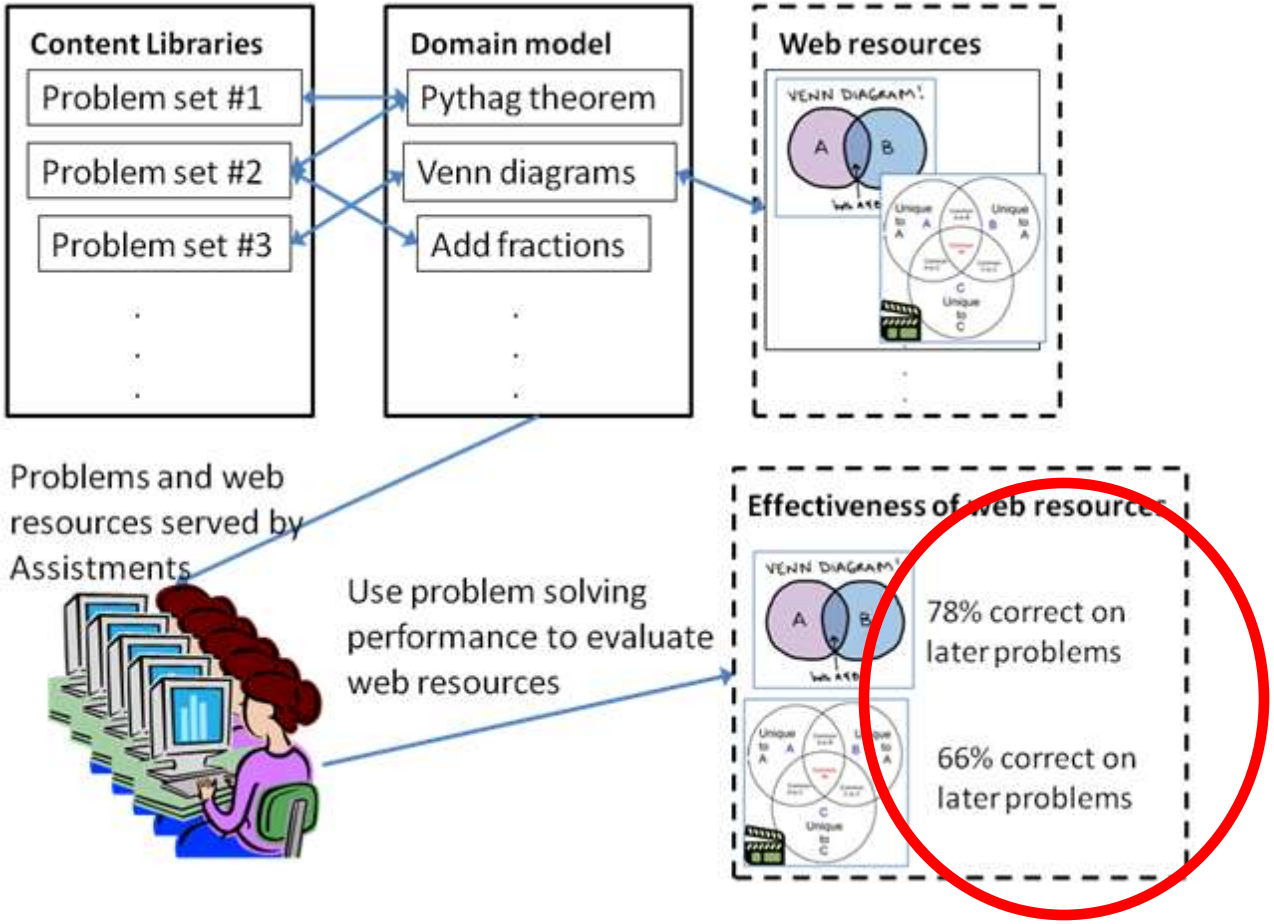
How web pages work within ASSISTments



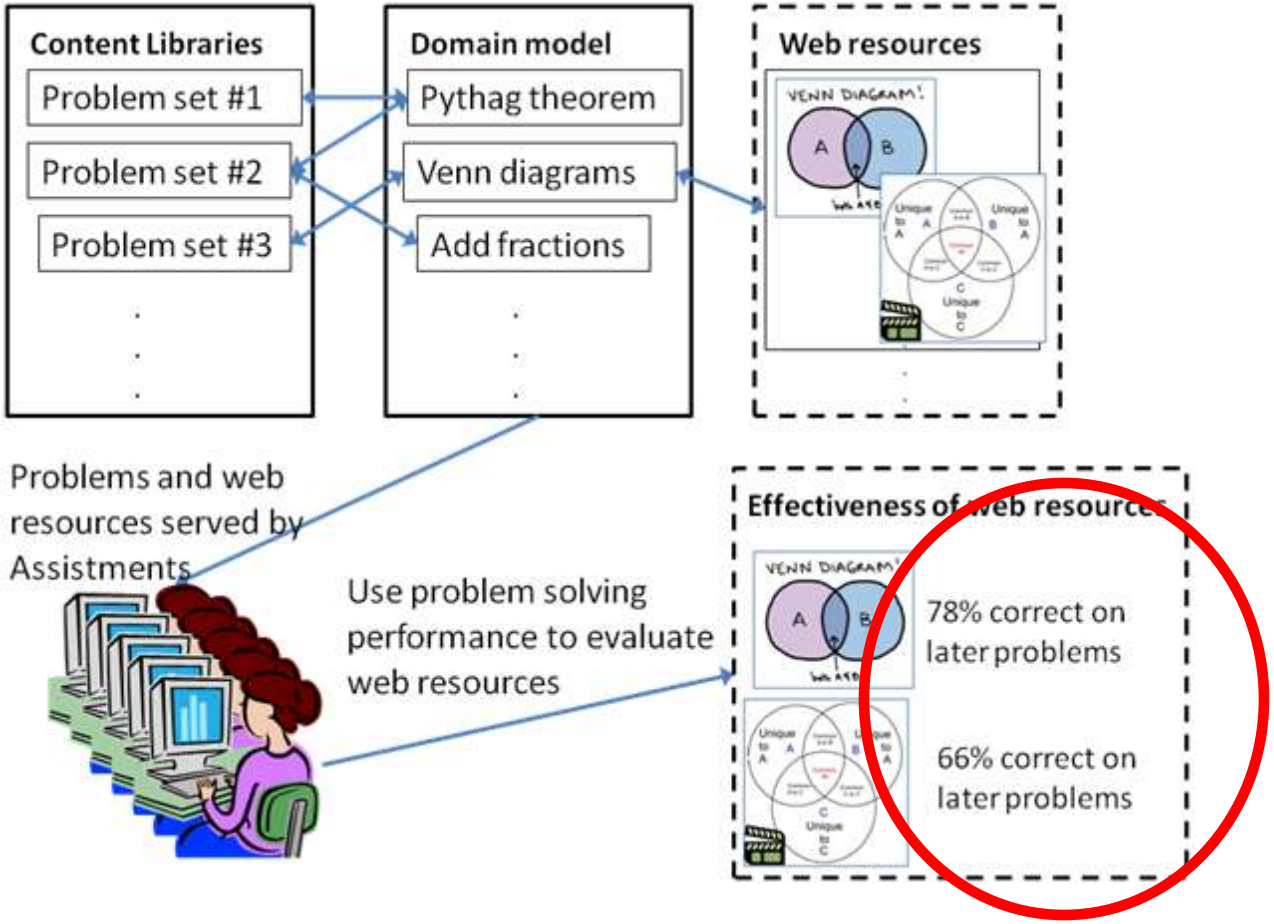
Collect data from a variety of users



Estimate educational efficacy of a web page



Estimate educational efficacy of **a web page** – why not generalize?



Generalization

- “Web page 1431 is ineffective” →
  - “web pages with videos over 3 minutes long are not effective”

- The more general statement is far more useful

Can imagine interaction effects with students

- "Web page 1431 is ineffective" →
  - "web pages with videos over 3 minutes long are not effective"

- "web pages with videos over 3 minutes long are not effective, unless the student is *completely confused* about the topic and has *sufficient self-discipline* to watch it."

Existing infrastructure

- WEBsistments presents web page to student
- Automatically evaluate effectiveness of that page

Collect data about pages we present

- WEBsistments presents web page to student
- Automatically evaluate effectiveness of that page
- **Have meta-data about pages**

What are useful data about pages?

How long is it (# of word, video length)
Vocabulary complexity (too hard to read?)
How frequently it is visited (If the site has a counter?)
Pagerank (popularity of who is pointing)
Alexar rating
When last modified

What will let us generalize about pages?

- URL text
- Incoming links
  - Comments on links
- Comments on the page
- “Length” of page
- Types of media on the page
- Analysis of text on page (difficulty, alignment with concept)
- Any provided meta-data tagging
  - Strongly prefer automated (for us!) features

Collect data about students

- WEBsistments presents web page to student
- Automatically evaluate effectiveness of that page
- **Have meta-data about pages**
- **Have data about students**

Use properties of students for predicting a page's effectiveness

- Proficiency on the current topic
- Proficiency on other math topics
- P(thrash) – will the student master?
- Gender
- Age
- Behavioral detectors (gaming, off task)
- Grit (self-discipline) or other pretests
- Student's ability to learn from other web pages

How to perform the learning?

- Collect data from students learning from web pages
- Automatically evaluate each page's efficacy
- Collect data about each page
  - Easier than what search companies do!
- Collect data about each student

Possible data table

URL	Length	Media	Thrash?	Gender	Grit	Outcome
www.k12.	Short	Text	Yes	Boy	Medium	0.7
Youtube.	Long	Video	No	Girl	High	0.1
Schooltube.xyz….	Medium	Video	Yes	Girl	Low	0.3
…	…	…	…	…	…	…

How to perform the learning?

- Collect data from students learning from web pages
- Automatically evaluate each page's efficacy
- Collect data about each page
  - Easier than what search companies do!
- Collect data about each student

- **Once that is done, it's "easy"**

Learning approach

- Treat as supervised learning problem:
  - F(**page meta data**, **student data**) → page efficacy
- Use function approximator of your choice
- We'll probably use linear regression

Why build EdRank?

- The web is vast
- Want to focus student time productively
- There is life outside of computer tutors
- Provide guidelines for others

The web is vast

- There are a lot of potentially useful web pages out there
- We don’t have time to process them all
  - Let alone have students test their efficacy
- Would be nice if some way to focus project staff and student time on useful aspects

Wait a minute…

- Aren't you trusting EdRank too much?
  - What if ranks the best page as being ineffective?
- That scenario will probably happen
- Far more pages than we can process
  - Better question: Over lifespan of WEBsistments, will we be better off with EdRank?

Want to focus student time productively

- Why have students observe bad web pages to we can measure their effectiveness?
  - Doesn't seem like a good use of *anyone's* time
- Focus student time on pages likely to be beneficial

There is life outside of ITS

- Most students will never use an ITS even once in their lives
- Many will search for a web page to help them understand something
- Long term: develop a browser plugin that can automatically rate a web page's efficacy
  - Or suggest web pages that are better?

Provide guidance to others

- What if we learn that videos with the teacher visible are generally ineffective?
  - Would be great to inform the *broader community* of that fact! (don't mean EDM community)

- Better guidelines for web-based content would help many people
  - But hard to gather for others – one of our methodological "secret weapons"

Discussion

- Open textbooks
  - Would those be good candidates for linking to for good content?

Future work

- All of it :-)

- “We” have WEBsistments infrastructure built
  - Some students have seen web pages
  - Can even tell whether one page looks better

- Lots of work: better outcome measure for pages, automatically collecting features of pages, features about students, …

Conclusions

- EDM provides us the ability to evaluate an "intervention" inside of an ITS
  - Evaluating arbitrary web content is next step forward for the field – bigger than any of us!

- Using the web for learning is a developing field, and we have something to contribute