The Data and Algorithms that Power AskOski

Zachary A. Pardos
Assistant Professor
zpardos

Graduate School of Education
School of Information
UC Berkeley

CAHL Computational Approaches to Human Learning (CAHL) research lab

GRADUATE SCHOOL OF EDUCATION

UC Berkeley School of Information

What is AskOski today?

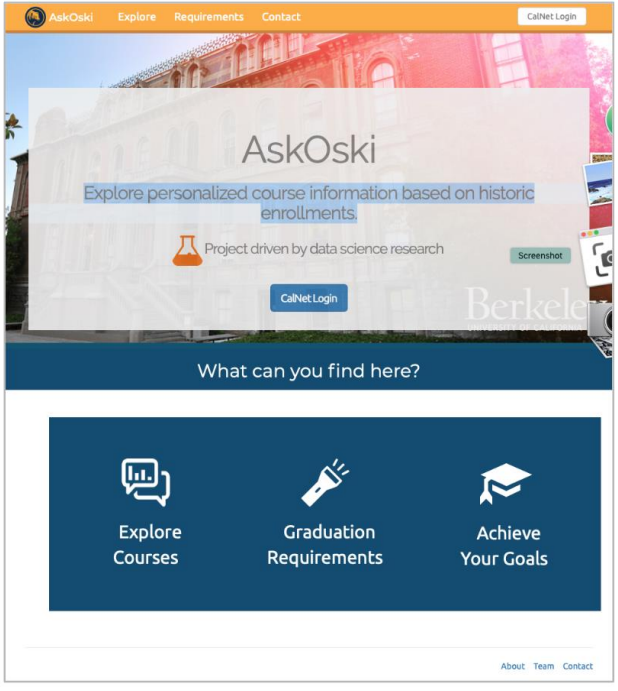
AskOski

AskOski provides personalized course recommendations to students based on:

- The user's attributes and filters
- Historic enrollments at Berkeley

AskOski uses **machine learning** to allow students to:

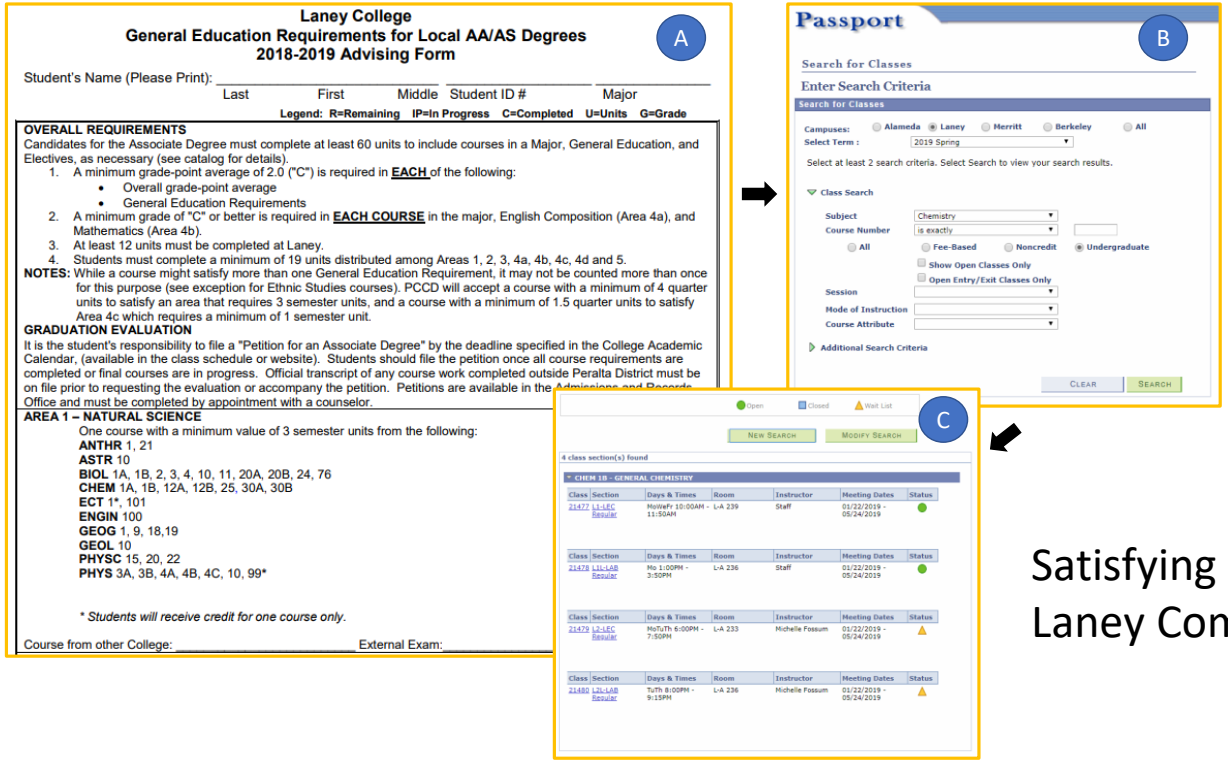
- **Explore Courses** based on a favorite course
- Find courses that satisfy personal **graduation requirements**



Description from UCB Usability Workshop Feedback

Why are recommender systems needed in higher education?

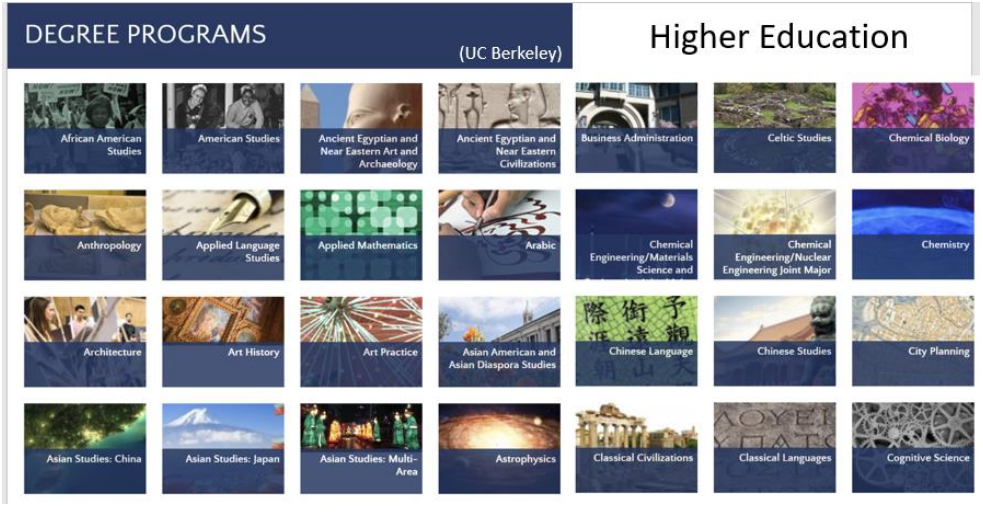
Motivation: Graduation planning is hard



Satisfying a degree requirement at Laney Community College

Motivation: Exploration is hard

- Where do I start?
- How are courses conceptually related?
- Am I ready to take a course in a completely different field?
- Will this course count for credit at the school I want to transfer to?

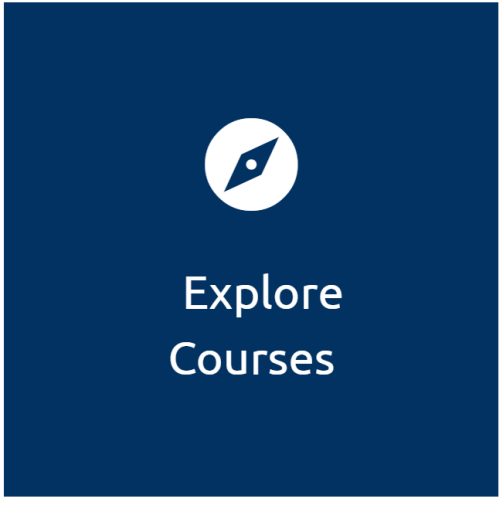


40% 4-year graduation rate (USA)

1:400 or 1:1000 student to adviser ratio

Timeline of AskOski development

- 2015: $600,000 in NSF grants awarded (IIS BIGDATA)
- 2016: First prototype of course recommender released at Berkeley
- 2017: Small scale user study of satisfaction and desired functionality
- 2018: Beginning of research and development of course exploration feature
- 2019: Beginning of research and development of requirement satisfaction feature
- 2019: User study on relevance of search “inferred keywords”
- **2020**: Explore, Requirements, Search, PLAN features live (askoski.berkeley.edu)




Graduation
Requirements

Timeline

- 2015: $600,000 in NSF grants awarded
- 2016: First prototype of AskOski released
- **2017**: Small scale user study of satisfaction and desired functionality

Measure of acceptance	Construct	Average rating
If available, I would use the AskOski service in the future	Use intentions	**4.25**
I would recommend AskOski to a peer	Use intention	3.85
The AskOski interface is intuitive to use	Perceived ease of use	3.80
The AskOski service provided information I couldn't get elsewhere	Perceived usefulness	3.55
The AskOski service improved my comfort with enrollment planning	Overall satisfaction	3.55

Ratings are from 1 (strongly disagree) to 5 (strongly agree)

Pardos, Fan, & Jiang (2019)

Timeline

- 2015: $600,000 in NSF grants awarded
- 2016: First prototype of AskOski released
- **2017**: Small scale user study of satisfaction and desired functionality

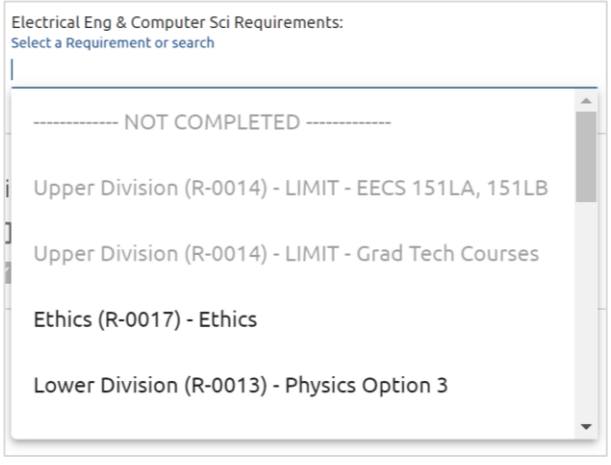
Student course selection priorities	Average rating
Satisfying major requirements	**4.85**
Graduating on time	4.25
Maintaining GPA	4.15
Intellectual enrichment	4.10
Career goals	4.05
Satisfying course prerequisites	4.05
Satisfying breadth requirements	3.80

Average ratings are on a scale from 1 (not important) to 5 (very important)

Retrieving per-student degree audit info from SIS

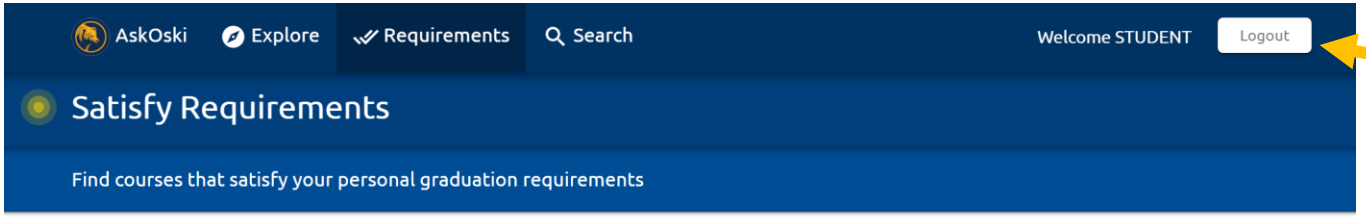
Undergraduate Degrees

ACADEMIC_PLAN_NM	list_coverage
Industrial Eng & Ops Rsch BS	100
Education UG	100
Environmental Eng Science BS	100
BioE/MSE Joint Major BS	100
Disability Studies UG	100
EECS/MSE Joint Major BS	100
Chem Eng/MSE Joint Major BS	100
EECS/NE Joint Major BS	100
Jewish Studies UG	100
ME/NE Joint Major BS	100
Environmental Sciences BS	97.22
Chemical Engineering BS	96.67
Chemistry BS	95
Chemical Biology BS	95
250AMU	94.87
MCB-Immunology BA	94.44
Development Studies BA	94.44
Slavic Lang & Lit BA	94.44
Cognitive Science BA	94.12



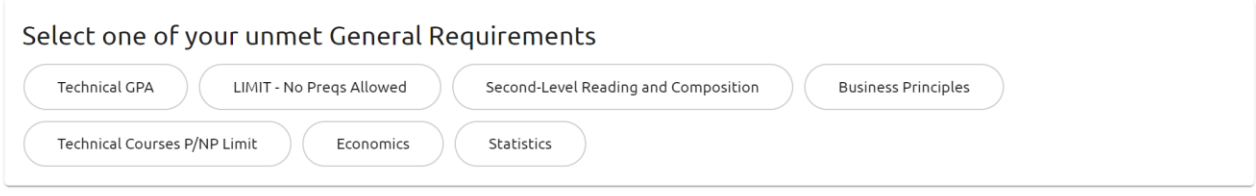
Percentage of searchable (not grey) requirements in AskOski

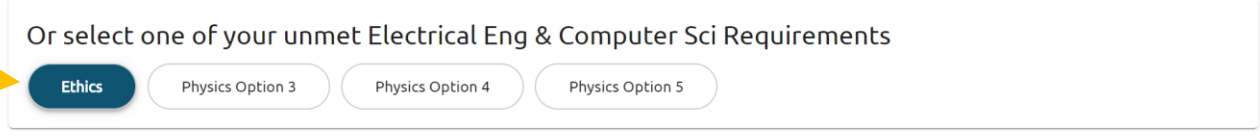
Queries the Academic Plan Review (**APR**) Module of People Soft



1

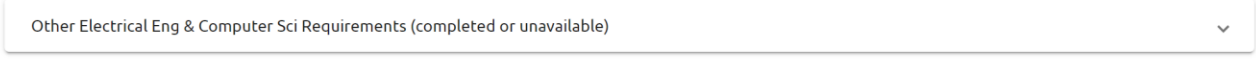
Student logs-in

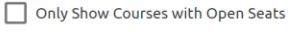




2

Selects one of her remaining requirements



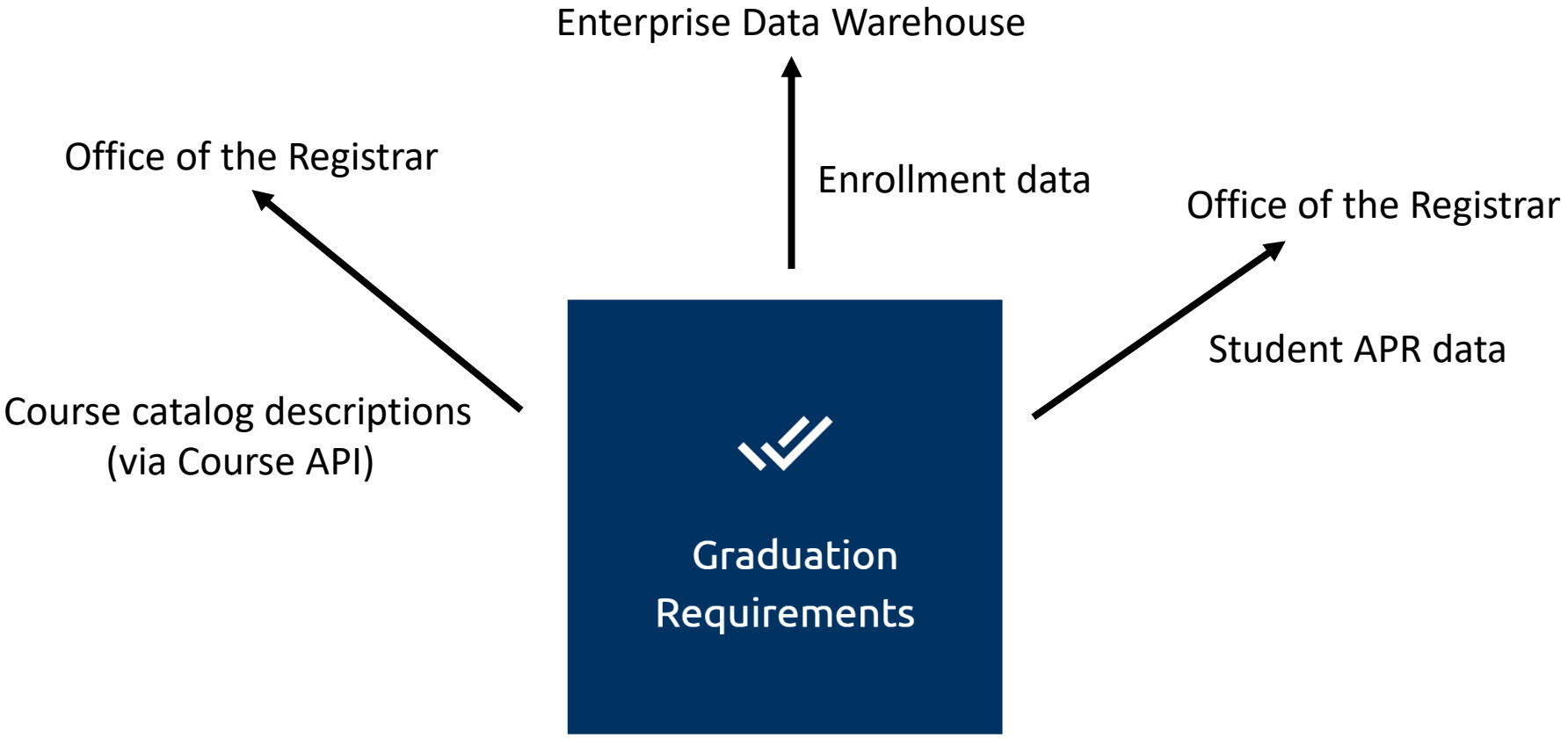


3

Is shown courses satisfying that requirement available in the next semester

Considerations for Fall 19:

CCN	Class	Subject
19319	Social Implications of Computer Technology (195)	Computer Science
29166	Energy and Society (C100)	Energy and Resources
27356	Ethics in Science and Engineering (100)	Bioengineering


Enterprise Data Warehouse
Enrollment data
Office of the Registrar
Course catalog descriptions
(via Course API)
Office of the Registrar
Student APR data
Graduation
Requirements


Explore
Courses

“We want students to find their own path, not follow others’”

- UC Berkeley Staff

“Will a recommender system only show the popular courses?”

- UC Berkeley Student

Generate serendipitous course suggestions

[1] Student selects a favorite course (relevance signal)

Choose a Favorite Course

Subject: Sociology ×

Course: Evaluation of Evidence (5) ×

Search

Considerations across campus

[2] Generate results based on similarity to the favorite course in our model. Diversify by Department

#1 Data Science Connector

Letters & Science (88)

Connector courses are intended to connect the Foundations of Data Science (COMPSCI C8/INFO C8/STAT C8) course with particular fields of study. They will apply the concepts and techniques of the foundation course to topics of interest in a particular discipline in order for students to develop critical thinking in data in subject areas that most interest them; these courses also provide a more nuanced understanding of the context in which the data comes into existence.

#2 Introduction to Urban Data Analytics

City & Regional Planning (101)

This course (1) provides a basic intro to census and economic data collection, processing, and analysis; (2) surveys forecasting and modeling techniques in planning; (3) demonstrates the uses of real-time urban data and analytics; and (4) provides a socio-economic-political context for the smart cities movement, focusing on data ethics and governance.

#3 Introduction to Ecological Data Analysis

Env Sci, Policy, & Mgmt (173)

Introduces concepts and methods for practical analysis of data from ecology and related disciplines. Topics include data summaries, distributions, and probability; comparison of data groups using t-tests and analysis of variance; comparison of multi-factor groups using analysis of variance; evaluation of continuous relationships between variables using regression and correlation; and a glimpse at more advanced topics. In computer laboratories, students put concepts into practice and interpret results.

#4 Cartographic Representation

Geography (183)

Problems in the representation of quantitative and qualitative data on thematic maps.

How is ***Serendipity*** defined?

Unexpectedness + ***Successfulness***

(Shani & Gunawardana, 2011)

Diversity

Mechanism frequently used to produce unexpectedness

- Choose items from across several appropriate categories (domain specific)
- Often used in combination with a signal of user relevance (content/collaborative-based)

- Has been the primary optimized outcome in recommender systems (e.g., CTR, consuming, buying, following)

Filter Bubble

Over personalized/Overfit/Narrowing model of user interest

(Kay, 2000; Nguyen, 2014)

Online Validation: User Study

- **70** UC Berkeley students from across campus participated
- Each student specified **a favorite course** they had taken at the university
- 10 course recommendations were generated from our candidate serendipity-designed ML models and from the session-based RNN

Online Validation: User Study

- For each course recommended, they rated their level of agreement with the following **3** statements:
  - "This course was unexpected" [**unexpectedness**]
  - "I am interested in taking this course" **[successfulness]**
  - "I did not know about this course before" **[novelty]**
- Students also rated the 10 recommendations from an algorithm as a whole, with agreement to the following:
  - "Overall, the course results were diverse" [**diversity**]
  - "The course results shared something in common with my favorite course" [**commonality**]

Online Validation: User Study Results

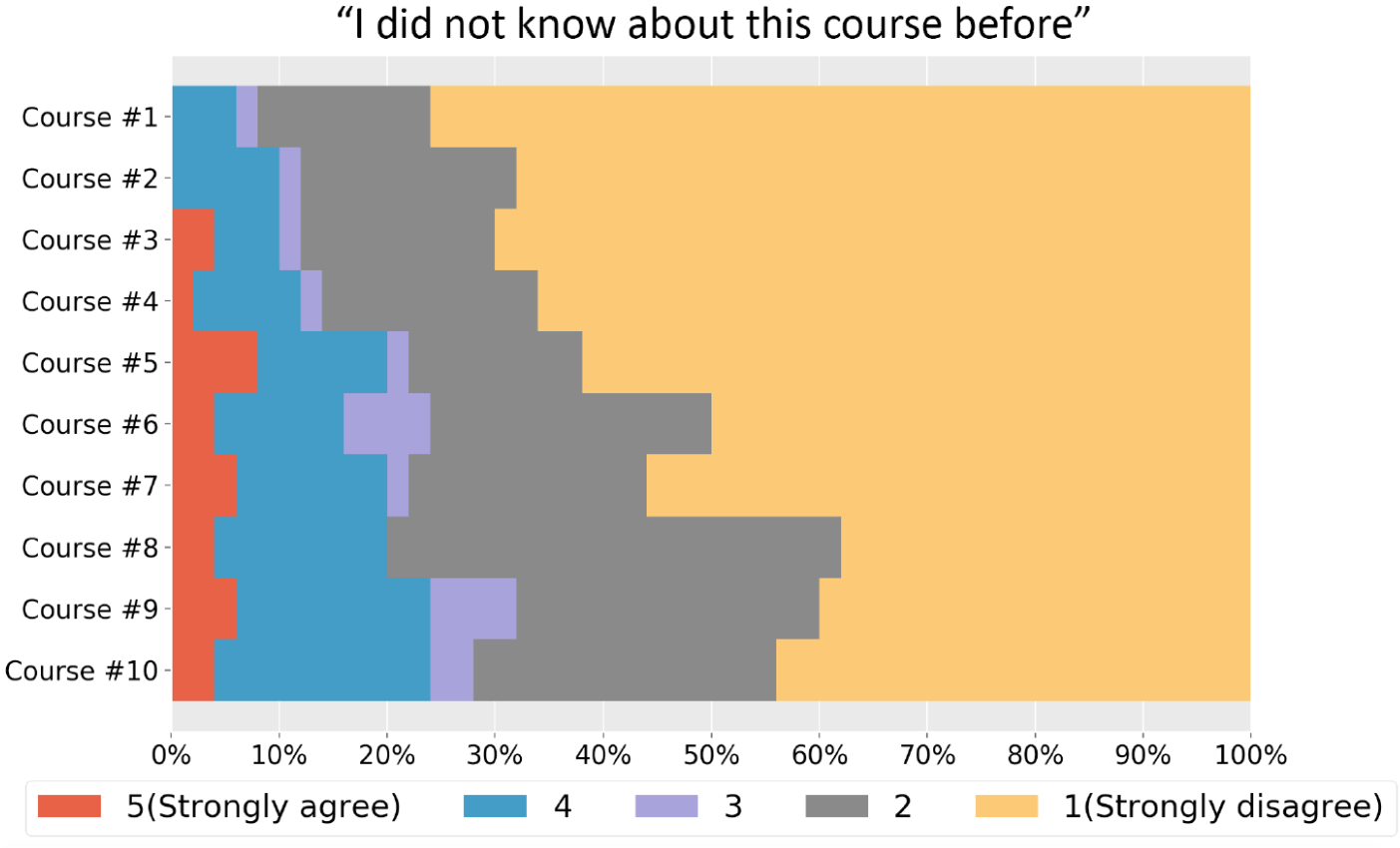
Average user ratings of courses based on five measures

representation	unexpected	successful	serendipity	novelty	diversity	commonality
Catalog Descrip.	**3.550**	2.904	**3.227**	**3.896**	4.229	3.229
tfidf & multi-c2v (div)	3.473	2.851	3.162	3.310	**4.286**	2.986
bin. &multi-c2v (div)	3.297	2.999	3.148	3.323	4.214	3.257
bin. &multi-c2v (non-div)	2.091	**3.619**	2.855	2.559	2.457	**4.500**
Collaborative	2.184	3.566	2.875	**1.824**	3.160	4.140

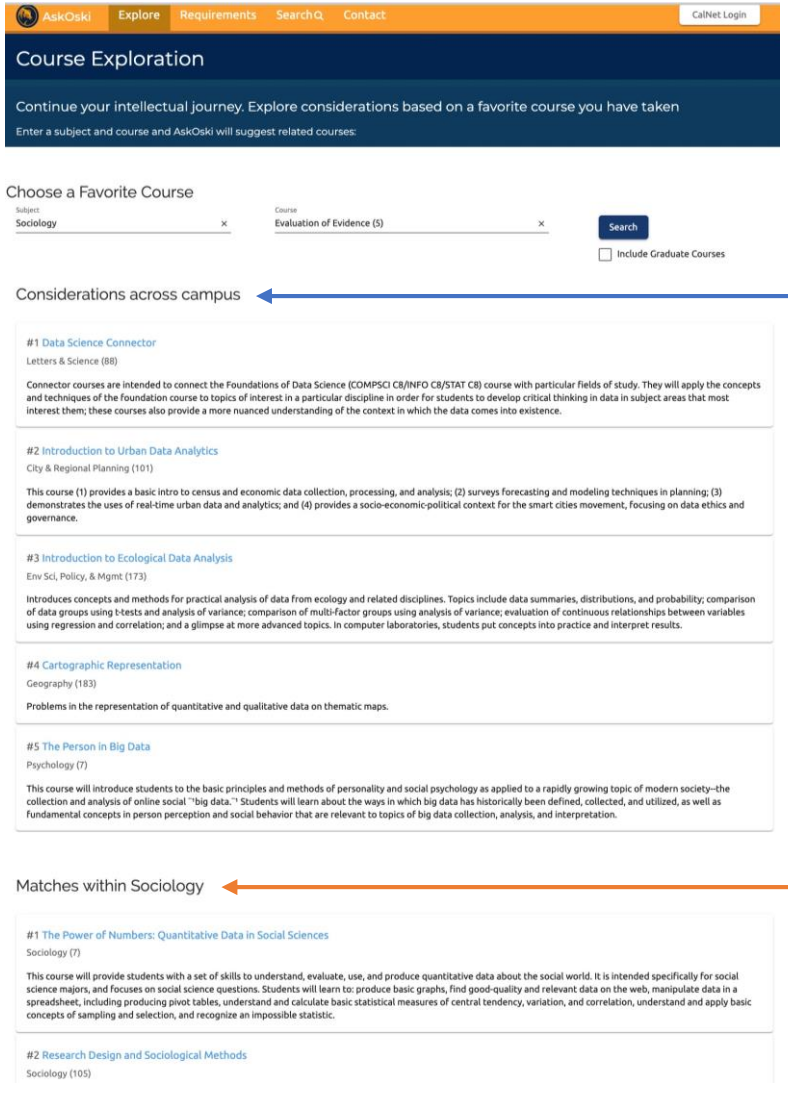
(serendipity is the average of unexpected and successful)

Pardos, Z.A., Jiang, W. (In press) Designing for Serendipity in a University Course Recommendation System. In K. Verbert, M. Scheffel, N. Pinkwart, & V. Kovanovic (Eds.) *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020). ACM. Frankfurt, Germany.

Lack of novelty in “people like you took X” style recommendation in user test



System Implementation

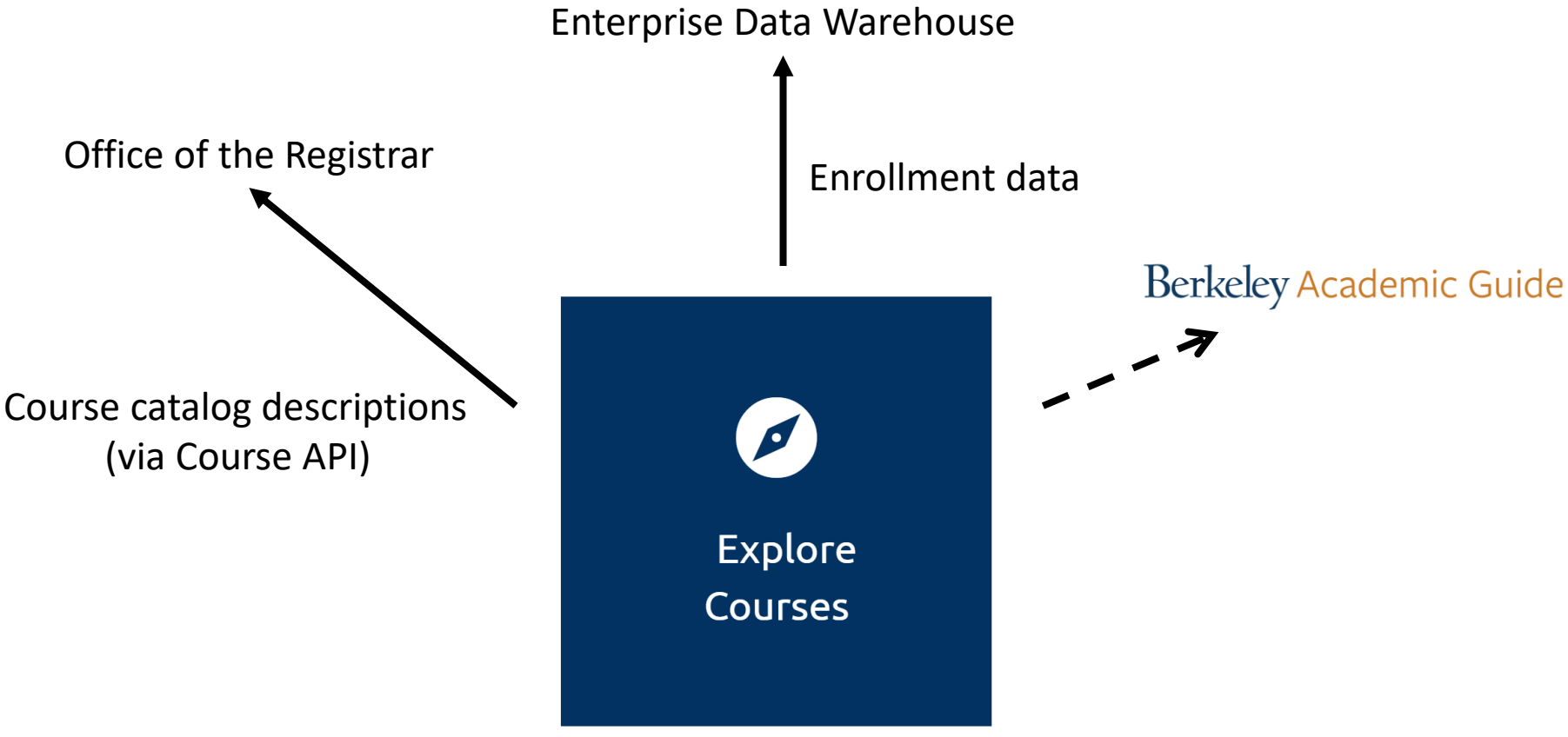


BOW (div) now generates the “across campus” results

- best serendipitous rated model

Binary BOW + multi-c2v now generates the within department results

- best equivalency model


Enterprise Data Warehouse
Office of the Registrar
Enrollment data
Berkeley Academic Guide
Course catalog descriptions
(via Course API)
Explore
Courses


Search

Infer latent topics of courses

Course description may not be informative

2019 Fall **LATIN 1 001 - LEC 001** offered through **Classics**

Elementary Latin

Emily Mullin
M, TU, W, TH | 10:00 am - 10:59 am | Dwinelle 106
Class #: 22195 | Units: 4

Open Seats
9 Seats Reserved for Students with Enrollment Permission

Beginners' course.

Course description may contain unfamiliar jargon

2019 Fall **EDUC C260F 001 - LEC 001** offered through **Graduate School of Education**

Machine Learning in Education

Class #: 33193 | Units: 3

Open Seats
6 Unreserved Seats

This course covers computational approaches to the task of modeling learning and improving outcomes in Intelligent Tutoring Systems (ITS) and Massive Open Online Courses (MOOCs). We will cover theories and methodologies underpinning current approaches to knowledge discovery and data mining in education and survey the latest developments in the broad field of human learning research....

Students are not experts. Their queries will reflect basic pre-conceptions

Infer latent topics of courses

Course description may not be informative

2019 Fall **LATIN 1 001 - LEC 001** offered through **Classics**

Elementary Latin

Emily Mullin
M, TU, W, TH 10:00 am - 10:59 am Dwinelle 106
Class #: 22195 Units: 4

Open Seats
9 Seats Reserved for Students with Enrollment Permission

Beginners' course.

Inferred Keywords: prose, poetry, grammar, medieval, greek

Course description may contain unfamiliar jargon

2019 Fall **EDUC C260F 001 - LEC 001** offered through **Graduate School of Education**

Machine Learning in Education

Class #: 33193 Units: 3

Open Seats
6 Unreserved Seats

This course covers computational approaches to the task of modeling learning and improving outcomes in Intelligent Tutoring Systems (ITS) and Massive Open Online Courses (MOOCs). We will cover theories and methodologies underpinning current approaches to knowledge discovery and data mining in education and survey the latest developments in the broad field of human learning research....

Inferred Keywords: algorithms, covers, computational, cover, modeling, discussions, communication, implementation

Dong, M., Yu, R., Pardos, Z.A. (2019) Design and Deployment of a Better Course Search Tool: Inferring latent keywords from enrollment networks. In M. Scheffel & J. Broisin (Eds.) *Proceedings of the 14th European Conference on Technology Enhanced Learning* (EC-TEL). Delft, The Netherlands. Springer. Pages 480-494.

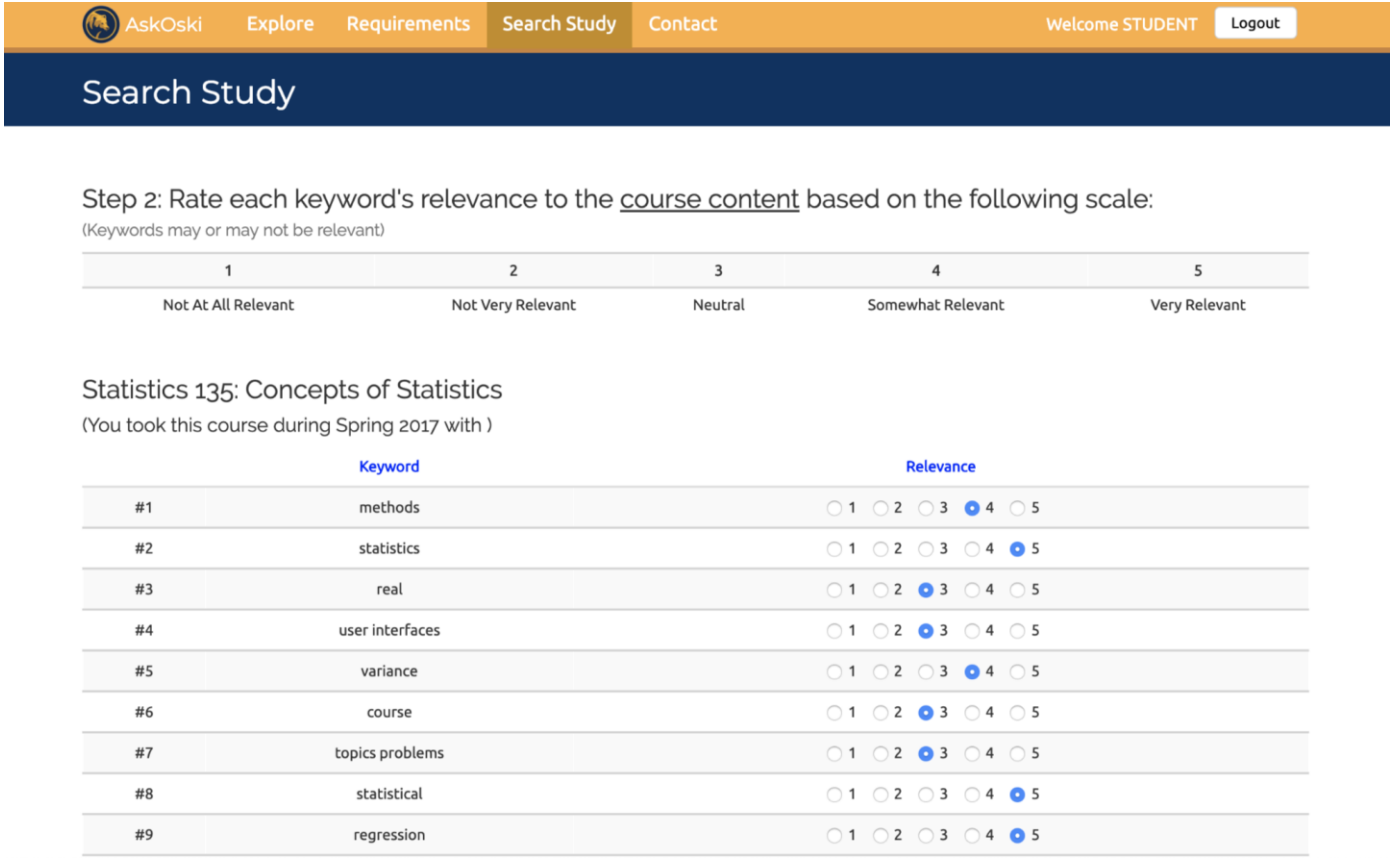
Online Validation: User Study

- **75** UC Berkeley students from across campus participated
- Each student specified **6 favorite courses** they had taken at the university
- For each course, keywords from the translation model and other baselines were displayed in random order
- Students rated each keyword on a five point Likert scale

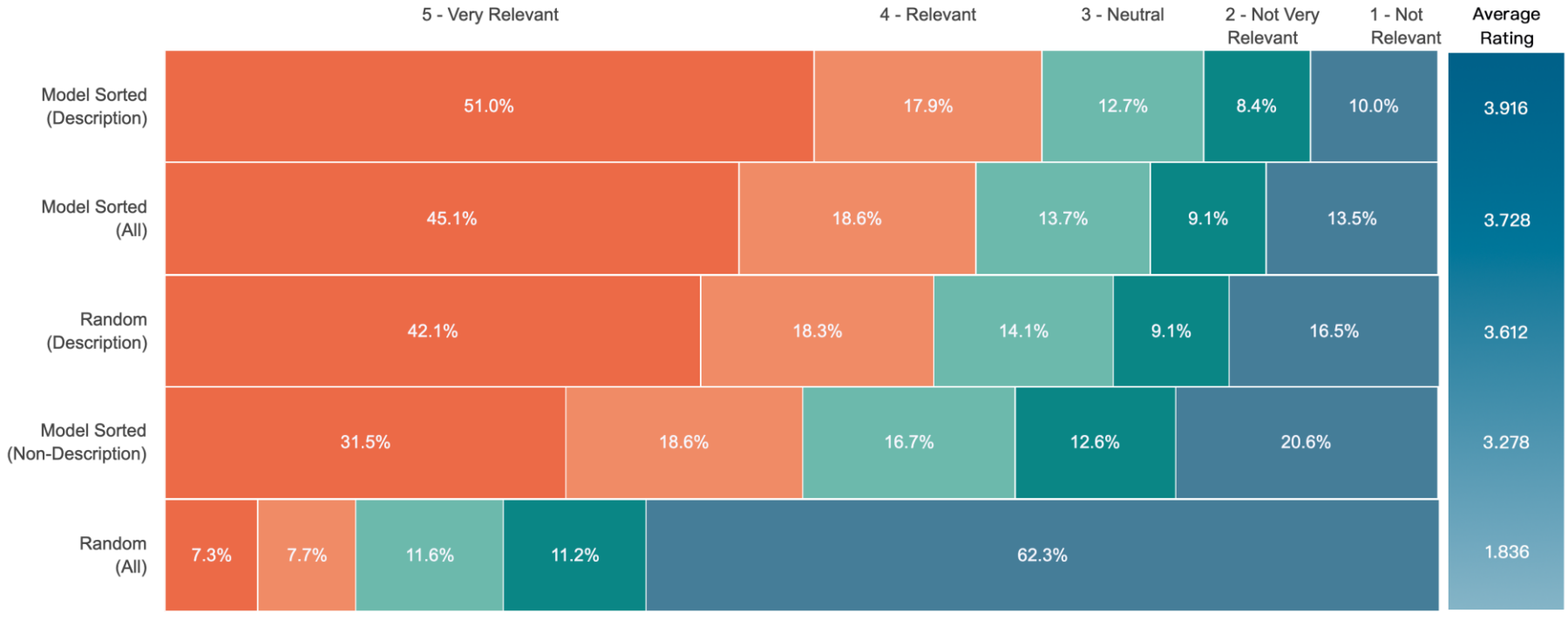
Which method produces the most relevant course keywords as judged by students who took the course?

How relevant are the keywords to the course as compared to words from the actual description?

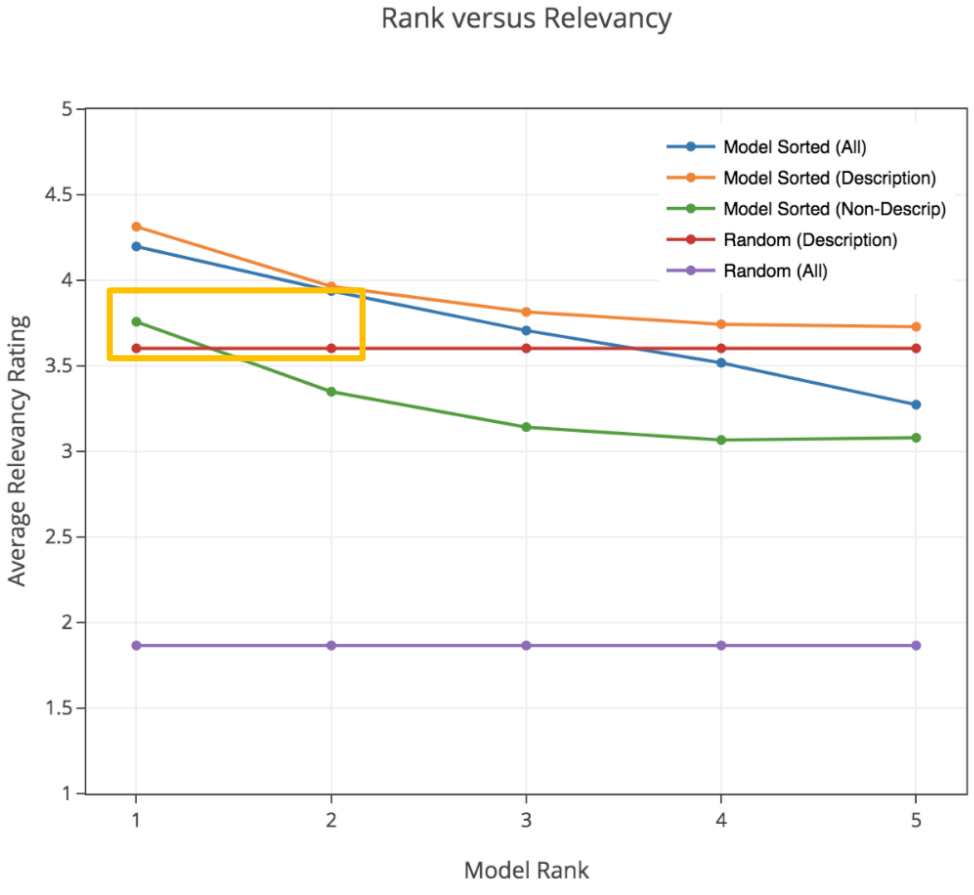
Online Validation: User Study



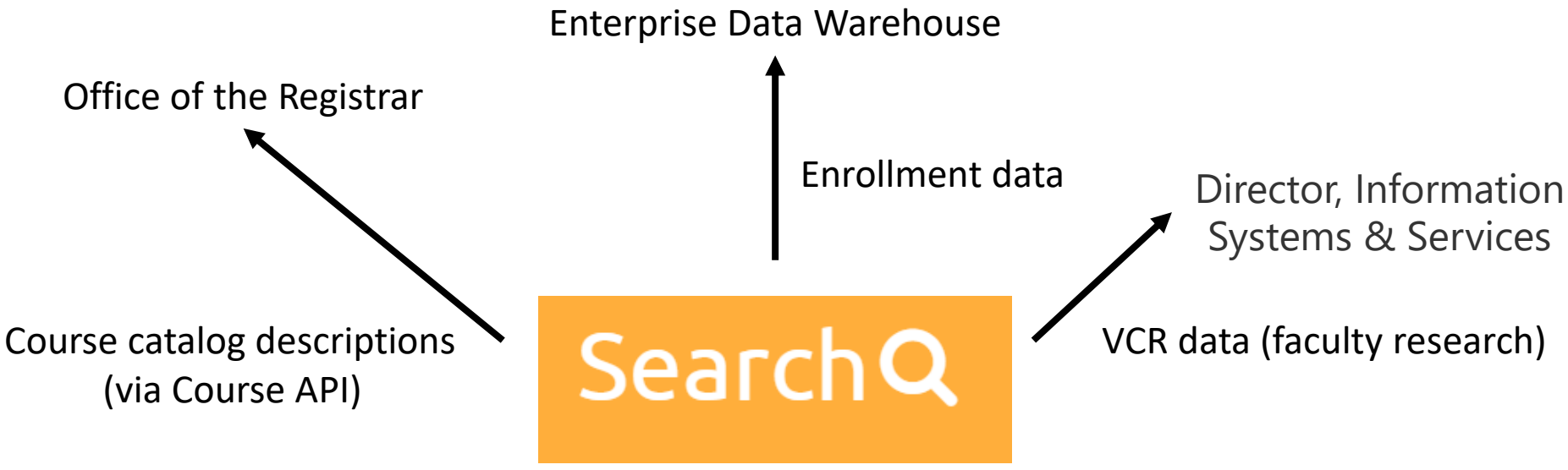
Online Validation Results



Online Validation Results



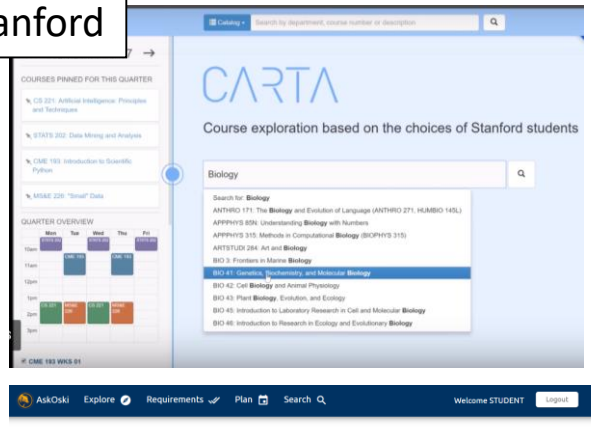
Correlation between probability and rank was used to select a threshold of probability from Model Sorted (All) group, such that all keywords above that probability can be expected to be more relevant, on average, than a word chosen at random from the description.



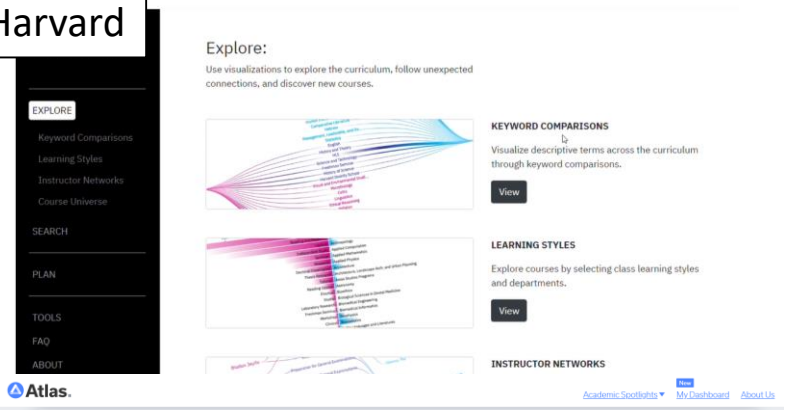
(Instructor results preview)

Nascent Course Information Platforms

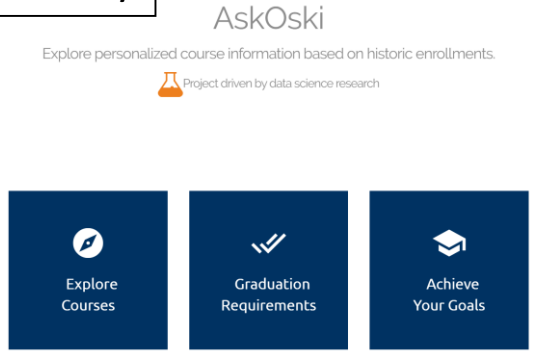
Stanford



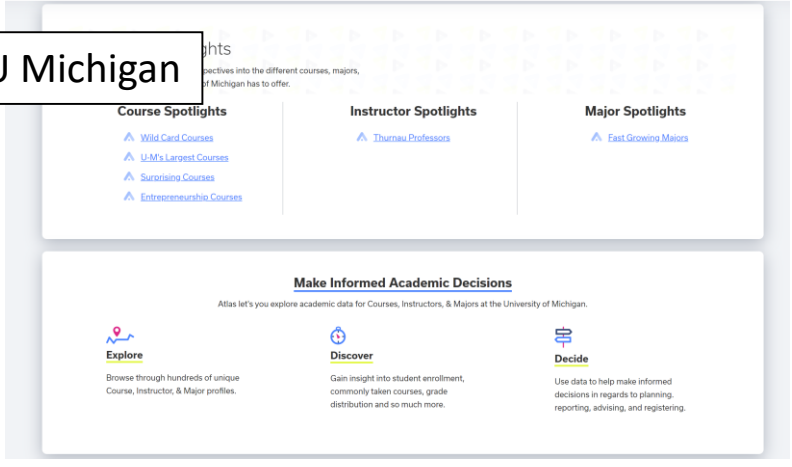
Harvard



UC Berkeley



U Michigan



DISCUSSION

- Give students the data you want them to make decisions based on
- More Information is not always better

(Do not show students grade distributions)

Chaturapruek, S., Dee, T. S., Johari, R., Kizilcec, R. F., & Stevens, M. L. (2018, June). **How a data-driven course planning tool affects college students' GPA: evidence from two field experiments**. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale* (pp. 1-10).

My Future Directions

- Within UC Berkeley
- Outside of UC Berkeley

References

- Pardos, Z.A., Jiang, W. (In press) Designing for Serendipity in a University Course Recommendation System. In K. Verbert, M. Scheffel, N. Pinkwart, & V. Kovanovic (Eds.) *Proceedings of the 10th International Conference on Learning Analytics and Knowledge* (LAK 2020). ACM. Frankfurt, Germany.
- Dong, M., Yu, R., Pardos, Z.A. (2019) Design and Deployment of a Better Course Search Tool: Inferring latent keywords from enrollment networks. In M. Scheffel & J. Broisin (Eds.) *Proceedings of the 14th European Conference on Technology Enhanced Learning* (EC-TEL). Delft, The Netherlands. Springer. Pages 480-494.
- Pardos, Z. A., Fan, Z., Jiang, W. (2019) Connectionist Recommendation in the Wild: On the utility and scrutability of neural networks for personalized course guidance. *User Modeling and User-Adapted Interaction,* 29(2), 487–525. https://doi.org/10.1007/s11257-019-09218-7
- Jiang, W., Pardos, Z.A., Wei, Q. (2019) Goal-based Course Recommendation. In C. Brooks, R. Ferguson & U. Hoppe (Eds.) *Proceedings of the 9th International Conference on Learning Analytics and Knowledge* (LAK). ACM. Tempe, Arizona. Pages 36-45. https://doi.org/10.1145/3303772.3303814
- Jiang, W., Pardos, Z.A. (2019) Time Slice Imputation for Personalized Goal-based Recommendation in Higher Education. In D. Tikk & P. Brusilovsky (Eds.) Proceedings of the 13th ACM Conference on Recommender Systems. Copenhagen, Denmark. ACM. Pages 506-510.
- Pardos, Z.A., Chau, H., Zhao, H. (2019) Data-Assistive Course-to-Course Articulation Using Machine Translation. In J. C. Mitchell & K. Porayska-Pomsta (Eds.) *Proceedings of the 6th ACM Conference on Learning @ Scale* (L@S). Chicago, IL. ACM.
- Pardos, Z.A. (2017) Big Data in Education and the Models that Love Them. *Current Opinion in Behavioral Sciences*. Vol 18, 107-113. [html]

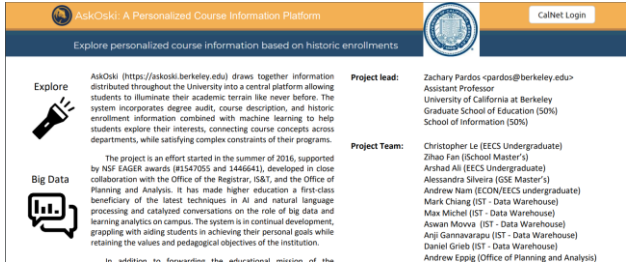
The Data and Algorithms that Power AskOski
Thank You!

Thank You!

Questions?

Zachary A. Pardos
University of California at Berkeley

@zpardos



One-page recommender system synopsis: tiny.cc/askoski

Acknowledgement: This work was funded by the National Science Foundation (Awards #1547055, #1446641)