



EXTENDING ENSEMBLE – AN EDUCATION DIGITAL LIBRARY FOR COMPUTER SCIENCE EDUCATION

COMPUTER SCIENCE, VIRGINIA TECH
YINLIN CHEN AND EDWARD A. FOX

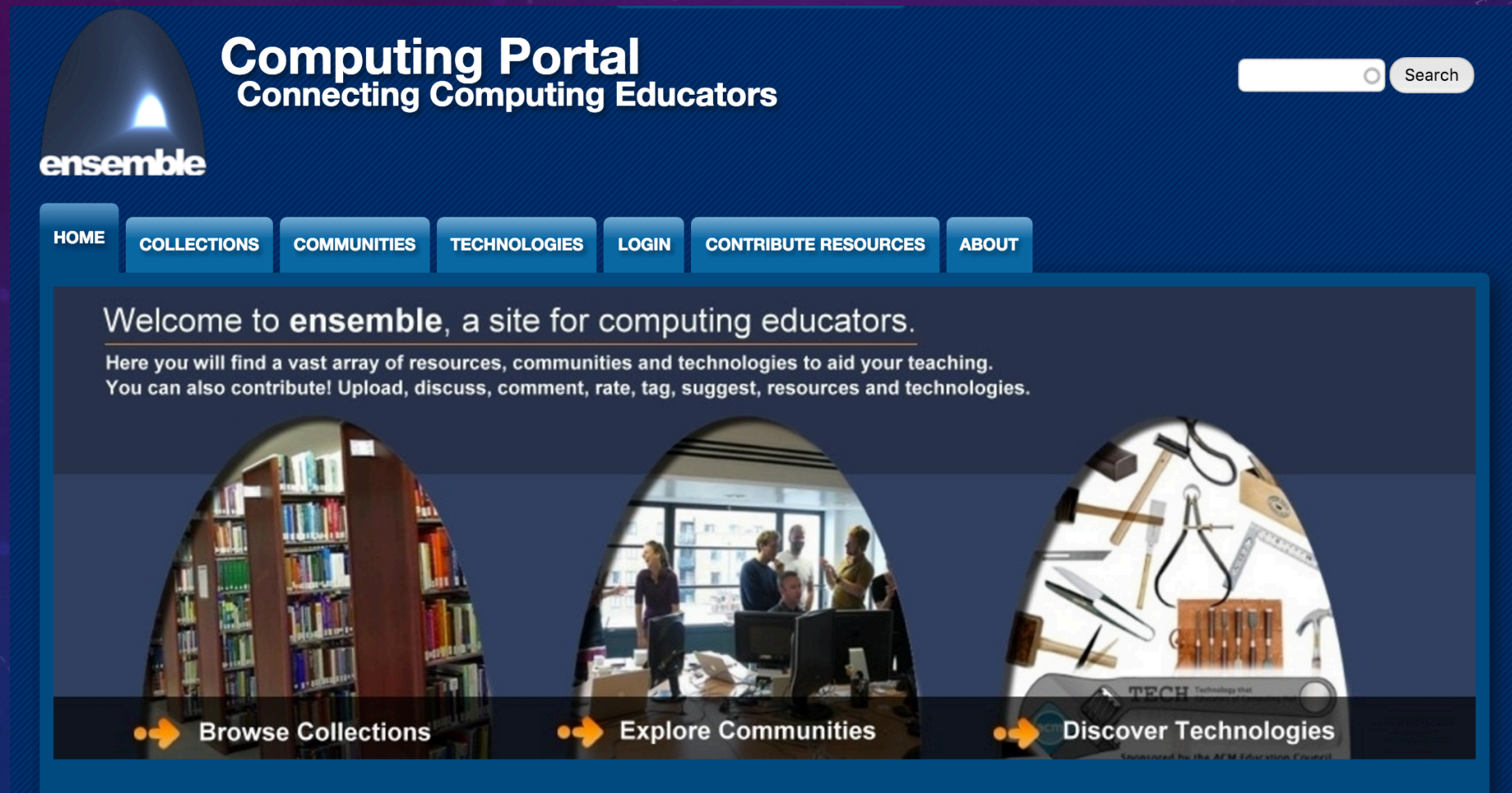




INTRODUCTION

- Ensemble – an educational digital library for computing education.

<http://www.computingportal.org>



The screenshot displays the homepage of the Computing Portal. At the top left is the 'ensemble' logo, featuring a stylized light source within a dark, dome-like shape. To its right, the text 'Computing Portal' is prominently displayed in a large, bold, white font, with the tagline 'Connecting Computing Educators' in a smaller font directly beneath it. In the top right corner, there is a white search bar with a magnifying glass icon and the word 'Search' to its right. Below the header, a horizontal navigation bar contains seven blue buttons with white text: 'HOME', 'COLLECTIONS', 'COMMUNITIES', 'TECHNOLOGIES', 'LOGIN', 'CONTRIBUTE RESOURCES', and 'ABOUT'. The main content area has a dark blue background. It begins with a white text block that reads: 'Welcome to **ensemble**, a site for computing educators. Here you will find a vast array of resources, communities and technologies to aid your teaching. You can also contribute! Upload, discuss, comment, rate, tag, suggest, resources and technologies.' Below this text are three large, arched image windows. The first window shows a library with bookshelves and is labeled 'Browse Collections' with an orange double-arrow icon. The second window shows a group of people in a modern office or classroom setting and is labeled 'Explore Communities' with the same icon. The third window shows various tools and equipment, including a hammer, pliers, and a wrench, and is labeled 'Discover Technologies' with the icon. At the bottom of the third window, there is a small logo for 'TECH' and the text 'Technology that' and 'Sponsored by the ACM Education Council'.

ENSEMBLE CHALLENGES

1. Find and select good educational resources to include in the collection.
2. Ensure that the collection matches users' needs.
3. Describe these educational records completely and correctly.
4. Correctly categorize these resources in support of browsing/searching by category.

EDUCATIONAL RESOURCE



EDUCATIONAL RESOURCE



EDUCATIONAL RESOURCE

67% of millennials agree that they can find a YouTube video on anything they want to learn.

Ref: <https://www.thinkwithgoogle.com/articles/i-want-to-do-micro-moments.html>

METHODOLOGY

- Use Machine **learning** approaches to:
 - **Identify computing educational resources**
 - **Categorize new harvested resources**
- Use **Mturk** to:
 - **Evaluate resources**
- Build a series of **process** to curate new computing educational collections

METHODOLOGY

- Build classifiers using ACM Digital Library (DL) metadata records
 - Computing education classifier:
 - Dataset: ACM papers in Computing Education proceedings
 - Binary classification: Naive Bayes
 - ACM CCS classifier:
 - Dataset: ACM papers in each ACM CCS category
 - Multiclass classification: SVM

ACM DIGITAL LIBRARY (DL) METADATA RECORDS



```
{
  "id": "1352233",
  "timestamp": "2008-12-03T05:00:00Z",
  "title": "A model for high school computer science education",
  "doi": "10.1145/1352135.1352233",
  "url": "http://dl.acm.org/citation.cfm?id=1352233",
  "abstract": "This paper presents a model program for high school computer science education. It is based on an analysis of the structure of the Israeli high school computer science curriculum considered to be one of the leading curricula worldwide. The model consists of four key elements as well as interconnections between these elements. It is proposed that such a model be considered and/or adapted when a country wishes to implement a nation-wide program for high school computer science education.",
  "primary_cat": "K.3.2:Computer science education",
  "concept_desc": [
    "CCS->Social and professional topics->Professional topics->Computing education->Computing education programs->Computer science education"
  ],
  "concept_significance": [
    "0.10003456.10003457.10003527.10003531.10003533:500"
  ],
  "acronym": "SIGCSE '08",
  "proc_class": "symposium",
  "proc_title": "Computer science education",
  "conference_number": "39",
  "series_title": "Technical Symposium on Computer Science Education",
  "series_id": "SERIES307",
  "proc_desc": "Proceedings of the 39th SIGCSE technical symposium",
  "journal_code": "SGCS",
  "journal_abbr": "SIGCSE Bull.",
  "journal_name": "ACM SIGCSE Bulletin"
}
```



BUILDING COMPUTING EDUCATION CLASSIFIER

- Papers in Computer **Science** proceedings (e.g. SIGCSE)
- Years from 2000 to 2011
- **User** paper title and abstract as features, **removed** stop **words** and build the classifier
- Select the **top signification** words to generate search **terms** sets

Table 1: Top 10 lists used in computer science education papers

Human Factors	pedagogy	Computer science education
Design	computer science education	Model curricula
Experimentation	education	Computer-assisted instruction
Languages	curriculum	Information systems education
Management	CS1	Programming teams
Algorithms	visualization	HCI design and evaluation methods
Measurement	active learning	Computer-managed instruction
Performance	software engineering	Software creation and management
Theory	assessment	History of hardware

THE 2012 ACM COMPUTING CLASSIFICATION SYSTEM



The ACM Computing Classification System (CCS)				Switch to Flat View	Generate CCS Codes
General and reference	Hardware	Computer systems organization	Networks		
Software and its engineering	Theory of computation	Mathematics of computing	Information systems		
Security and privacy	Human-centered computing	Computing methodologies	Applied computing		
Social and professional topics	What is the CCS?				



THE 2012 ACM CCS - PAPERS

Networks	
Network architectures	4013
Network protocols	10503
Network components	555
Network performance evaluation	2821
Network properties	3507
Network services	4542
Network types	10451

Security and Privacy	
Cryptography	7244
Security services	2010
Intrusion/anomaly detection and malware mitigation	739
Security in hardware	25
Systems security	3090
Network security	212
Database and storage security	918
Software and application security	673
Human and societal aspects of security and privacy	981

Applied computing	
Electronic commerce	2387
Enterprise computing	8561
Physical sciences and engineering	26128
Life and medical sciences	11887
Law, social and behavioral sciences	3674
Arts and humanities	4545
Computers in other domains	11026
Operations research	8052
Education	10411
Document management and text processing	5589

BUILDING ACM CCS CLASSIFIER

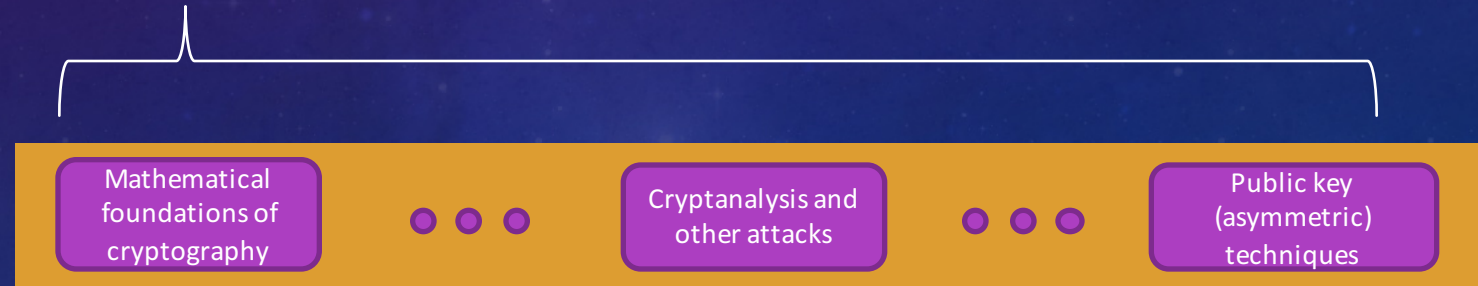
ACM Level 1 ccs



ACM Level 2 ccs



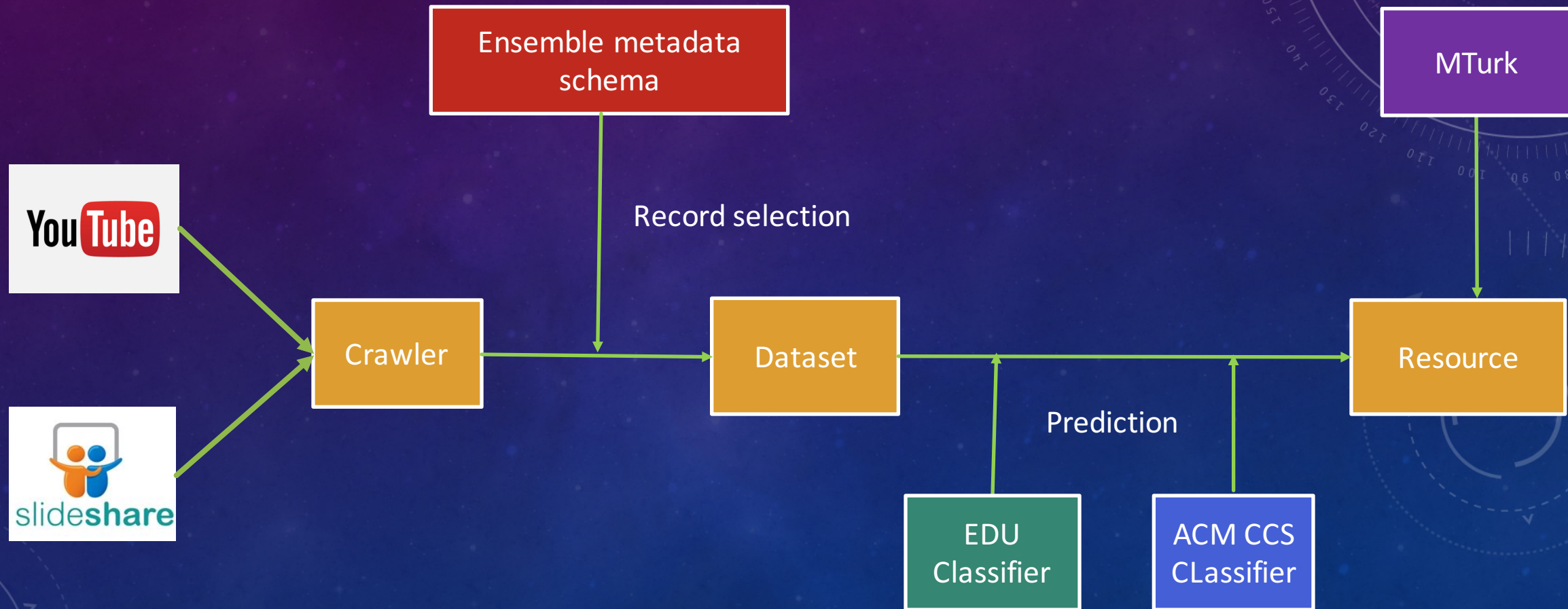
ACM Level 3 ccs



ENSEMBLE METADATA SCHEMA

- Extended version of the Dublin **Core**
- Combined with **NSDL** metadata guidelines, with associated controlled vocabularies.
 - https://wiki.ucar.edu/display/nsdl/docs/nsdl_dc
- We use this document to aid quality, since records were removed that lacked content in required or suggested metadata fields.
- <http://computingportal.org/docs/metadata-ensemble-recommendation.pdf>

OVERALL PROCESS





YOUTUBE & SLIDESHARE RECORDS

```
{
  "username": "mycodeschool",
  "rating": 4.92727279663,
  "description": "See complete series on data structures
here:\nhttp://www.youtube.com/playlist?
list=PL2_aWCzGMAwI3W_JlcBbtYTwiQSsOTa6P\n\nIn this lesson, we have
discussed binary tree in detail. We have talked about different types
of binary tree like \"complete binary tree\", \"perfect binary tree\"
and \"balanced binary tree\" and their properties.",
  "title": "Data structures: Binary Tree",
  "url": "https://www.youtube.com/watch?v=H5JubkIy_p8",
  "views": 121262,
  "author": "mycodeschool",
  "dislikes": 10,
  "comments": 46,
  "thumbnail": "http://i.ytimg.com/vi/H5JubkIy_p8/sddefault.jpg",
  "keywords": [
    "course",
    "job",
    "programming",
    "coding",
    "software",
    "Data Structure (Literature Subject)",
    "computer",
    "science",
    "computer science",
    "yt:cc=on",
    "programming interview"
  ],
  "likes": 540,
  "published": "2014-01-18 06:08:28",
  "duration": "00:16:17",
  "embedurl": "https://www.youtube.com/embed/H5JubkIy_p8",
  "category": "Education",
  "id": "H5JubkIy_p8"
}
```

```
<Slideshow>
  <ID>6824159</ID>
  <Title>Applied Computer Science Concepts in Android</Title>
  <Description>In this presentation, the unaware or indirect applications
of essential computer science concepts are dicussed as showcase. Jim
Huang presented in Department of Computer Science and Engineering,
National Taiwan University.</Description>
  <Status>2</Status>
  <Username>jserv</Username>
  <URL>http://www.slideshare.net/jserv/applied-computer-science-concepts-
in-android</URL>
  <Created>2011-02-05 19:00:05 UTC</Created>
  <Updated>2011-02-05 19:02:51 UTC</Updated>
  <Language>en</Language>
  <Format>pdf</Format>
  <DownloadUrl>http://s3.amazonaws.com/ppt-download/applied-cs-in-android-
110205130013-phpapp02.pdf?response-content-disposition=attachment&
Signature=2qDgIWfaZBpvus7Cls3oVUquSD8%3D&Expires=1412279914&
AWSAccessKeyId=AKIAI6DXMWX6TBWAHQCQ</DownloadUrl>
  <UserID>5962453</UserID>
  <PPTLocation>applied-cs-in-android-110205130013-phpapp02</PPTLocation>
  <Tags>
    <Tag Count="1" Owner="1">vm</Tag>
    <Tag Count="1" Owner="1">database</Tag>
    <Tag Count="1" Owner="0">kernel</Tag>
  </Tags>
  <NumDownloads>762</NumDownloads>
  <NumViews>9038</NumViews>
  <NumComments>0</NumComments>
  <NumFavorites>17</NumFavorites>
  <NumSlides>66</NumSlides>
</Slideshow>
```



MTURK YOUTUBE RECORD

Algorithms Lesson 1: Bubblesort



Description: For this lesson, we demonstrate graphically how to perform the bubblesort algorithm and analyze its time complexity. For C++ bubble sort code, go to our website xoax.net.

Category: Data Structure and analysis

Description: For this lesson, we demonstrate graphically how to perform the bubblesort algorithm and analyze its time complexity. For C++ bubble sort code, go to our website xoax.net.

Category: Data structures design and analysis

Section 1:

A) On a scale of 0 - 4, 0 being "Strongly Disagree" and 4 being "Strongly Agree", is this record of educational value in computing education?

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

On a scale of 0 - 4, 0 being "No confidence" and 4 being "Complete confidence", indicate the degree of confidence you have in your ability to successfully perform this task using the following scale:

No confidence	Less Confidence	Moderate Confidence	Strong Confidence	Complete confidence
<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

NEW RESOURCE TO ENSEMBLE



EXPERIMENT RESULTS

- A computing educational **classifier** is built
- Multiple classifiers are built to classify records in 2012 ACM CCS
- 2000 records from YouTube and SlideShare
- Topics in Security and privacy, Theory of computation, Human-centered computing, and Software and its engineering

CONCLUSIONS

- Explains how large numbers of additional high-quality resources will be gathered from YouTube and SlideShare and then validated, through machine learning and crowdsourcing methods, and thus help us continue gathering CS education resources into Ensemble.
- Addresses the challenges of building an educational digital library and extending Ensemble's value and level of usage for computing education.

Q & A

This research is supported by NSF Grants DUE-0840713, 0840715, 0840719, 0840721, 0840668, 0840597, 0836940, and 0937863. Our thanks go to ACM for providing the ACM DL metadata to use in this research.

