

Designing Curated Conversation-Driven Explanations for Communicating Complex Technical Concepts

April Wang | University of Michigan

aprilww@umich.edu

aprilwang.me

Parmit Chilana | Simon Fraser University

pchilana@sfu.ca

hci.cs.sfu.ca



The need for learning about emerging technologies and technical concepts among the general population has increased in recent years.

Chilana et al. CHI 16



Non-technical users in professional settings who participate in technical conversations



UX designer



Business Analyst



Human Resources



Product Manager



Alex business intern

Learning big data concepts to follow along technical conversation with his seniors

Stack Overflow Products Customers Use cases Search...

Home PUBLIC Stack Overflow Tags Users Jobs

What is Big Data & What classifies as Big data?

Asked 3 years, 7 months ago Active 2 years ago Viewed 1k times

I have went through a lot of articles but I dont seem to get a perfectly clear answ BIG DATA is. In one page I saw "any data which is bigger for your usage, is big considered big data for your mailbox but not your hard disc". Whereas another s be usually more than 1 TB with different volume / variety / velocity and couldn't i system". Also that data should be stored in a NOSQL db with Hadoop used to tr

Further, I have been working on a solution and was wondering if I could classify Snippets on the solution below,

- Millions of raw data records and usually 500 plus GB of data.
- SQL database as back-end and SSIS / SQL queries to cleanse/process the a meaningful form.
- Visualization using Spotfire

Kenneth Cukier: Big da 347,350 views • Sep 23, 2014

WIKIPEDIA The Free Encyclopedia Article Talk Read Edit View

Big data

From Wikipedia, the free encyclopedia

coursera Explore What do you want to learn? For Enterprise

Global Info

Information from, or otherwise

nal data-processing power, while data with discovery rate.[2] Big data ring, transfer, data was originally we handle big data, we g data often includes data within an acceptable

alytics, user behavior ue from data, and seldom ta now available are ecosystem.[10] Analysis

Offered By **Google Cloud**

Data Engineering, Big Data, and Machine Learning on GCP Specialization

Data Engineering on Google Cloud Platform. Launch your career in Data Engineering. Deliver business value with big data and machine learning.

Enroll for Free Starts Oct 01

Try for Free: Enroll to start your 7-day full access free trial

Financial aid available

50,192 already enrolled!

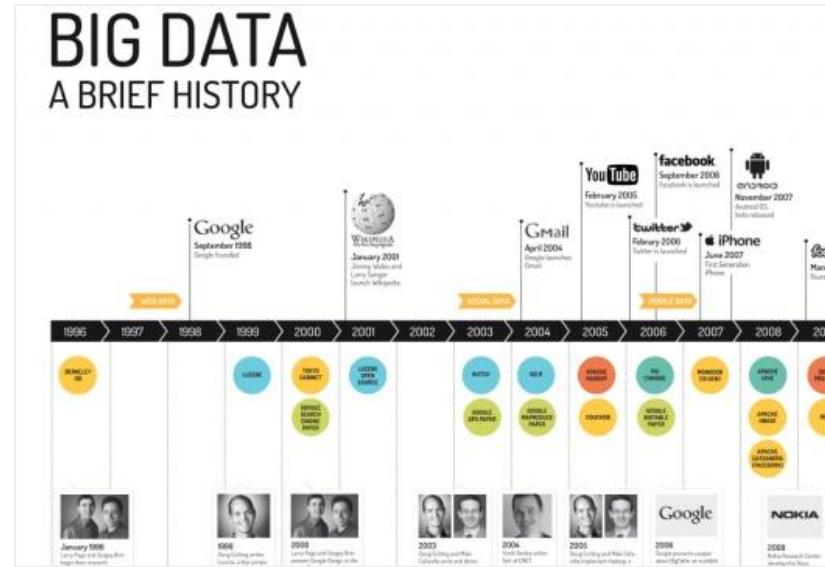
About How It Works Courses Instructors Enrollment Options FAQ



Morden learning resources **fail** non-technical learners!

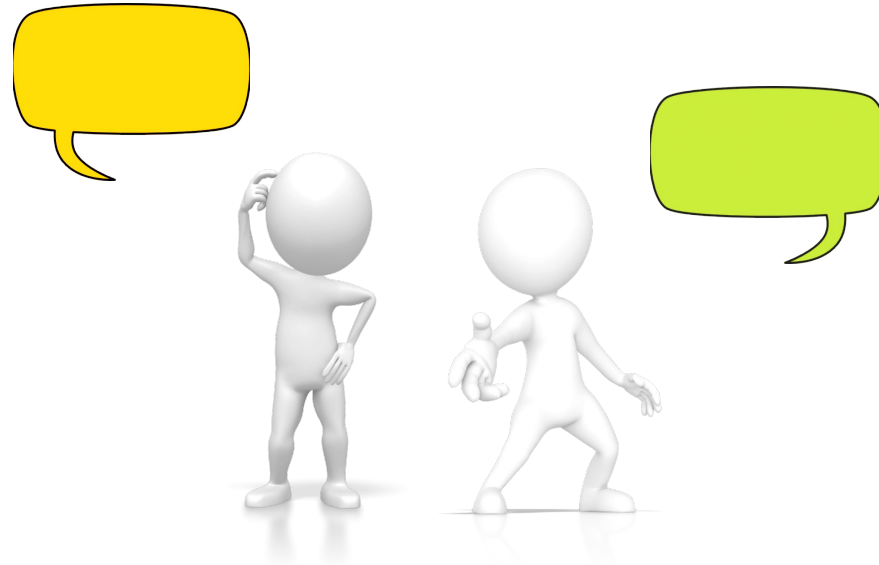
Wang et al. CHI18

- Takes too much time
- Too much focus on syntax and logic
- Explanations are not relevant









What if we could explain a complex technical concept by providing an **example** of how it can be used in a **conversation**?

JargonLite is a **conversation-driven** dictionary for technical concepts.

JargonLite Richard Log Out

Classification Analysis

Classification is a field of research to classify things/objects/images/sound/text etc etc using machine learning/Statistical Learning techniques. For example, consider problem of spam detection for an email. In such case, an email can be either a spam or not a spam, so there are two classes in this problem and classifying an email to spam or non-spam is a Classification Problem.

4
▲
▼
0

👤 Jennifer, HR Coordinator@RightBrain Tech user-generated 📧 0

Subject how/when the technology is used in real development scenarios

Character

- 👤 **Bob** Senior Developer
- 👤 **Alice** Business Intern

👤 What are the practical application areas of this technique?

❤️ 0 ◀ ▶ 📧 0

Step 1/3

The screenshot displays the JargonLite interface. At the top, there is a search bar and user information for 'Richard' with a 'Log Out' link. The main heading is 'Classification Analysis'. Below this, a post by Jennifer, HR Coordinator@RightBrain Tech, explains classification. The post includes a scatter plot with axes labeled 'Yellow Apples' and 'Red Apples', and illustrations of various fruits. A 'user-generated' tag is present. Below the post is a dialogue-style interactive component with a 'Subject' field containing 'how/when the technology is used in real development scenarios'. It features two characters: Bob, a Senior Developer, and Alice, a Business Intern. Alice has sent a message: 'What are the practical application areas of this technique?'. The interface includes a 'Step 1/3' indicator and navigation arrows.

concise and easily “skimmable” explanations added by other users using minimal jargon

a dialogue-style interactive component that explains how the technical concept can be used in a real-world conversation

Spark

Explanation

Example in a Conversation

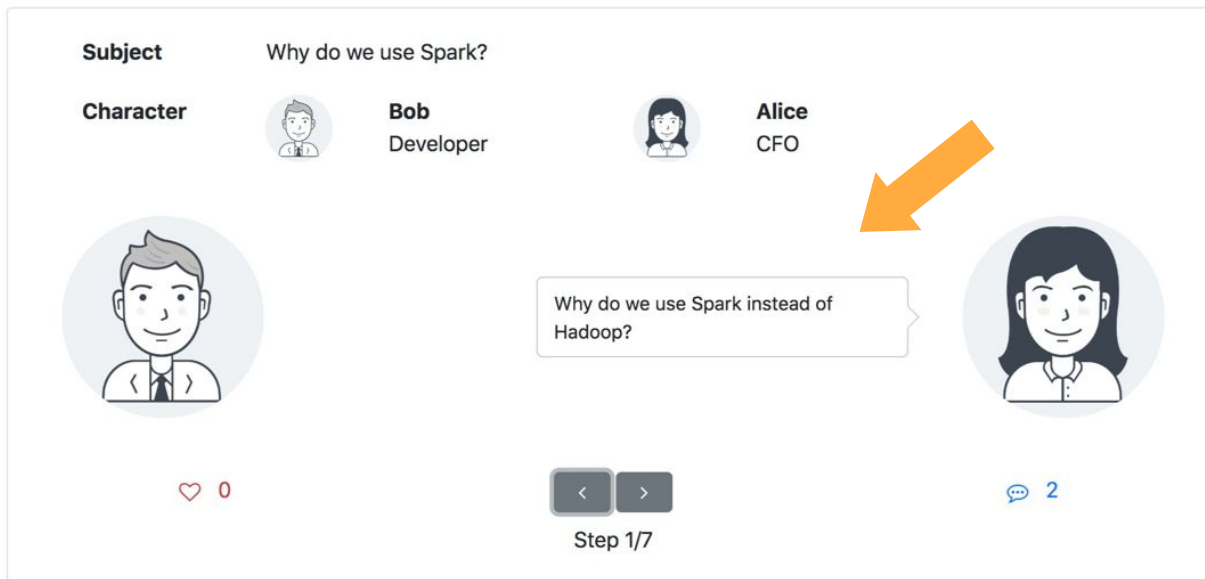
Add a New Example

Subject Why do we use Spark?

Character

Bob Developer

Alice CFO



Why do we use Spark instead of Hadoop?

Step 1/7

Spark

Explanation

Example in a Conversation

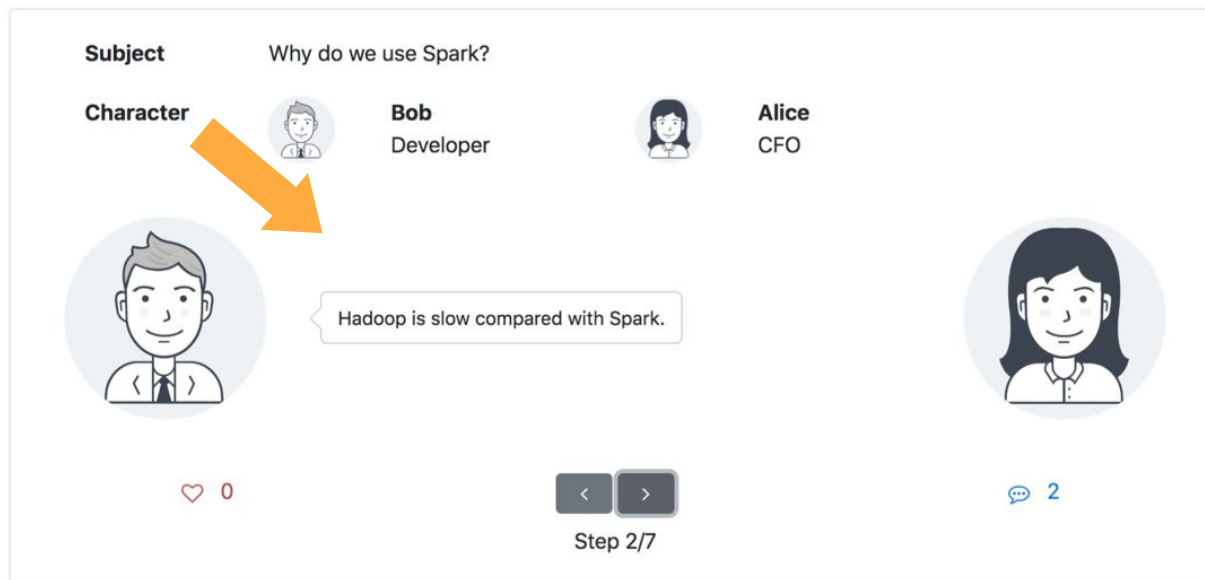
Add a New Example

Subject Why do we use Spark?

Character

Bob Developer

Alice CFO



Hadoop is slow compared with Spark.

0

2

Step 2/7

Spark

Explanation

Example in a Conversation

Add a New Example

Subject Why do we use Spark?

Character

Bob Developer

Alice CFO

Why is it slow?

Step 3/7

Spark


ExplanationExample in a Conversation

Add a New Example


Subject

Character


Why do we use Spark?




Bob
Developer



Alice
CFO


♥ 0

Data processing algorithms are iterative, MapReduce writes/reads data to/from disk at each iteration, which is costly.


💬 2

< >
Step 4/7

JargonLite is a **conversation-driven** dictionary for technical concepts.

JargonLite [Richard](#) [Log Out](#)

Classification Analysis

Classification is a field of research to classify things/objects/images/sound/text etc etc using machine learning/Statistical Learning techniques. For example, consider problem of spam detection for an email. In such case, an email can be either a spam or not a spam, so there are two classes in this problem and classifying an email to spam or non-spam is a Classification Problem.

4
▲
▼
0

👤 Jennifer, HR Coordinator@RightBrain Tech [user-generated](#) [🔗](#) [💬 0](#)

Subject how/when the technology is used in real development scenarios

Character

- Bob**
Senior Developer
- Alice**
Business Intern

What are the practical application areas of this technique?

👍 0 [⏪](#) [⏩](#) [💬 0](#)

Step 1/3

Evaluation

Seeding Community-Curated

Content

STEP 1

In-Lab Usability Study

STEP 2

Seeding Community-Curated Content

- Picking Big Data Related Concepts
- Simulating Community-Curated Content
 - 6 graduate students from a data science program as contributors

Term (picked for the study)	Number	Term	Number
Data Mining	6	Clustering Analysis	4
Data Visualization	6	Distributed File System	4
Database	6	IoT	4
Hadoop	6	Load Balancing	4
Machine Learning	6	R	4
MapReduce	6	Structured Data	4
Neural Network	6	Data Aggregation	3
NoSQL	6	ETL	3
Spark	6	Metadata	3
SQL	6	Petabyte	3
Classification Analysis	5	Database Administrator	2
Cloud Computing	4	Predictive Analytics	2

In-Lab Usability Study

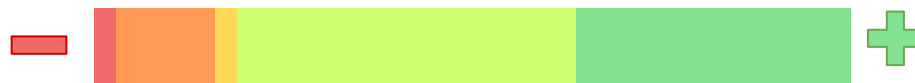
- 12 participant with little or no technical training
- Each participant was given 6 concepts
- For each concept
 - 3 minutes to check the explanation using the JargonLite dictionary
 - As if they were going to need to use these concepts in a conversation
 - Questionnaire asking perceptions and subjective feelings after each concept
- Post-test interviews

Key Findings

- **Ease-of-Understanding**
- **Maintaining Focus**
- **Trusting the Explanations**
- **Perceptions of Conversation-Driven Explanations**
- **Contributors' Perspective on the Content Creation Process**

- **Ease-of-Understanding**
- Maintaining Focus
- Trusting the Explanations

avg = 3.97



We are trying to communicate [technical] stuff in a simpler way, in a way that general people will understand...they [Wikipedia contributors] seem [to be] already educated in CS for so many years... they may not even realize it themselves, but they are using so many words that people who are not from tech background would hardly understand. (U3)

- Ease-of-Understanding
- **Maintaining Focus**
- Trusting the Explanations

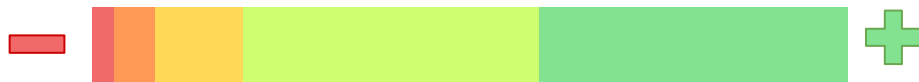
avg = 3.78



I lost focus after reading the first few words from textbook explanations and had to restart from the beginning... [the conversation-driven explanation] ... it really kept my interest to read and think. (U1)

- Ease-of-Understanding
- Maintaining Focus
- **Trusting the Explanations**

avg = 4.11



Could be more trustworthy if more information about the contributors could be included!

I wish this system could connect to websites like LinkedIn...to provide me more background of the other users who write the examples. (U6)

Key Findings

- Ease-of-Understanding
- Maintaining Focus
- Trusting the Explanations
- **Perceptions of Conversation-Driven Explanations**
- Contributors' Perspective on the Content Creation Process

Perceptions of Conversation-Driven Examples

- Participants appreciated the conversations because they felt like they could relate to them as personal stories they had already experienced.

[Here] it's about other people's experience... It's different from a standard explanation by Wiki [Wikipedia]. (U4)

Perceptions of Conversation-Driven Examples

- Different opinions on the style of conversation-driven explanations
 - the intuitive metaphors and long stories that encouraged them to think
 - the shorter, one-sentence “what/why” conversation

Key Findings

- Ease-of-Understanding
- Maintaining Focus
- Trusting the Explanations
- Perceptions of Conversation-Driven Explanations
- **Contributors' Perspective on the Content Creation Process**

Contributors' Perspective on the Content Creation Process

- The process of adding example conversations is easy (avg = 4.17)
- Willing to contribute towards such a system in real life

I [build] online reputation when I am answering domain-specific questions on Quora. So, I would totally be willing to contribute to this tool (JargonLite) to enhance my online reputation with other business people, who might be my potential customer or collaborator. (C3)

Contributors' Perspective on the Content Creation Process

- Unsolved problem -- generating an appropriate example conversation can be challenging and open-ended

For some concepts, it is easy to make comparisons, for example, Hadoop and Spark. Some concepts (e.g., data mining) are very general. It is not clear what the pros, and the cons should be. (C6)

Key Findings from Evaluation

- **Ease-of-Understanding**
- **Maintaining Focus**
- **Trusting the Explanations**
- **Perceptions of Conversation-Driven Explanations**
- **Contributors' Perspective on the Content Creation Process**

Future Work

- Improving Conversation-Driven Explanations
 - Different **formats** of instructions or tutorials for adding high-quality explanations
 - Different **levels of explanations** for different categories of technical concepts?

Future Work

- Conversation-Driven Explanations in Other Domains



Future Work

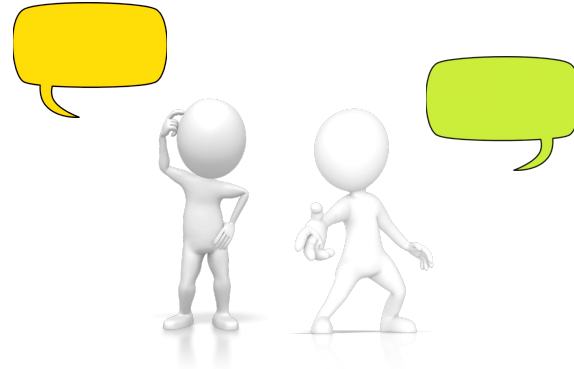
- Conversation-Driven Explanations in Other Domains



Future Work

- Conversation-Driven Explanations in Other Domains





conversation-driven mechanism



Designing Curated Conversation-Driven Explanations for Communicating Complex Technical Concepts

- present a novel approach for explaining technical concepts to non-technical users through the design of conversation-driven explanations
- provide initial insights into how non-technical users can benefit from conversation-driven explanations



April Wang | University of Michigan

aprilww@umich.edu
aprilwang.me

Parmit Chilana | Simon Fraser University

pchilana@sfu.ca
hci.cs.sfu.ca