

Topic

talk about annotation/comment strategies in big data technologies like Spark, Hive, etc

Introduction

Goal: tackle the 5 W's: Who what when where why

- What are annotations (**what**)
 - What are comments
- Why are they used (**why**)
- Brief history on annotations & comments detailing **who** gave the sensation a name, **where** and **when** they were first used

Body 1

Goal: Now that we know what Annotations & Comments are and the history behind them, now we go into the actual strategies

General Spark annotation & Commenting strategies

General Hive annotation & strategies

~~Interesting Big data tech strategies on commenting and annotations~~ (Moving the second section)

Body 2

Goal: Answer the questions I posed in the proposal, but gear them towards big data tech

- Is there a universal annotating/commenting rubric for SPARK/HIVE or is it on an employer by employer basis
 -
- Are there some unspoken rules that data scientists follow when annotating/commenting
- Are there differences based on region?
- What is the consensus on when & where to annotate/comment in the world of data science?
- What all information should your comment contain
 - On that note, how much is too much?

~~Body 3~~ (unneeded)

Conclusion

Restate all topic sentences

~~Contain a call to action~~

Something cheeky

Under the hood:

AVG 6 min medium article stats

WC of 1000 - 1500 words (9 - 14 pgs)

6th grade reading level is considered optimal

The formula for calculating the automated readability index is given below:

$$4.71 \left(\frac{\text{characters}}{\text{words}} \right) + 0.5 \left(\frac{\text{words}}{\text{sentences}} \right) - 21.43$$

Score	Age	Grade Level
7	11-12	Sixth Grade

1. Senter, R.J.; Smith, E.A. (November 1967). "[Automated Readability Index](#)". *Amrl-Tr. Aerospace Medical Research Laboratories (U.s.). Wright-Patterson Air Force Base*: 1–14. [PMID 5302480](#). AMRL-TR-6620. Retrieved March 18, 2012.

24-25 sentences

12 - 15 words per sentence

With occasional **3-5 word sentences** separating paragraphs

Use only **one** power word per title

Font Size **15**

Double spaced

Georgia or **Arial** is relatively common

<https://medium.com/the-mission/after-10-000-data-points-we-figured-out-how-to-write-a-perfect-medium-post-58c41c314f6a>