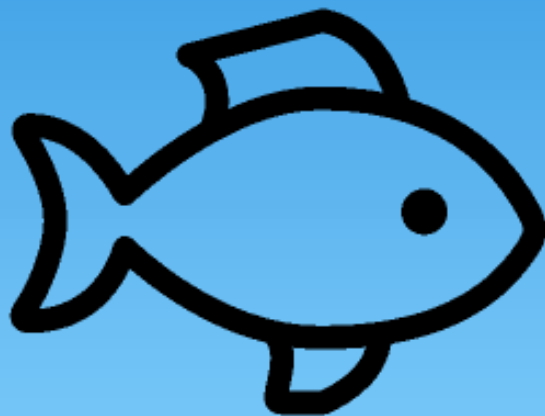


燒酒螺

英文文章閱讀平台



第八組

鄭以琳、張譯文、林修安



緣起

燒酒螺 各類魚貨 精選漁貨 魚譜

註冊 登入

那些石沈大海的單字，
你還記得多少？

成為漁夫

願者上鉤



本日頭條



A wave of hope as South Korea
mulls reopening schools in
MERS outbreak



Interpol suspends \$22.5m anti
match-fixing partnership with
FIFA



5 things to know for your New
Day -- Friday, June 12

(CNN)Dogs have joined the search for



Kurds Close in on IS-held Syrian
Town

Kurdish militias are closing in on an




網站功能

閱讀文章
單字查詢

燒酒螺 各類魚貨 精選漁貨 魚譜 cloud

Andrey Pavlov hits nightmare score of 17 on first hole at Lyonesse Open



(CNN)There are bad days at the office, and then there's the one Andrey Pavlov endured at the Lyonesse Open.

The Russian will be keen to forget about his Friday at the Diamond Country Club in Atzenbrugg, Austria, after writing himself into the record books for the wrong reason.

Pavlov, who shot an impressive 71 in the opening round Thursday after receiving a tournament invitation, registered a nightmare score of 17 on the par-five first hole -- the joint-second worst in European Tour history.

What's more, if that was not frustrating enough for the 1,598th-ranked Pavlov, he proceeded to find the water on a whopping six different occasions in that time.

Pavlov started the day six shots off the lead but ended it languishing down in the lower echelons of the tournament leaderboard and failed to make the cut.

He can at least take consolation from the fact that his **score** on the first hole was not as poor as Philippe Porquier's effort of 20 on a single hole at the 1978 French Open.

score

n. 分數；二十；記錄；刻痕

vt. 獲得；評價；劃線，刻劃；把...記下

vi. 得分；記分；刻痕

n. (Score)人名：(英)斯科爾

加入魚譜



網站功能

燒酒螺 各類魚貨 ▾ 精選漁貨 魚譜

單字	音標	解釋
science	'saɪəns	n. 科學；技術；學科；理科 n. (Science)人名；(英)臺恩斯
comments		n. 評論（comment的複數）；注解 v. 評論（comment的第三人稱單數形式）；解釋
physiology	'fɪzɪ'ɒlədʒi	n. 生理學；生理機能

單字庫

燒酒螺 各類魚貨 ▾ 精選漁貨 魚譜

標題	簡述	狀態
Dallas police headquarters shootin...	(CNN)Shortly after midnight on Sat...	閱讀中
'Jurassic World': Should you see it?	(CNN)Well, at least the dinosaurs a...	完成
Andrey Pavlov hits nightmare score...	(CNN)There are bad days at the off...	完成

閱讀紀錄



網站功能

閱讀心得

文章推薦

燒酒螺

各類魚貨 ▾

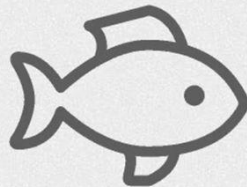
精選漁貨

魚譜

lala ▾

讀後心得：

送出



目前還沒有任何留言呦！

試著挑戰



Lifetime unveils first look at tell-all 'Full House' TV movie

(CNN)The Tanners are back on the small screen but there's a twist! Before Netflix's 2016 "Full House" revival, Lifetime has unveiled the first look at its upcoming "The Unauthorized Full House Story." The cable network has also announced the cast for upcoming behind-the-scenes movie that follows last year's "





EC2

RDS

SQS

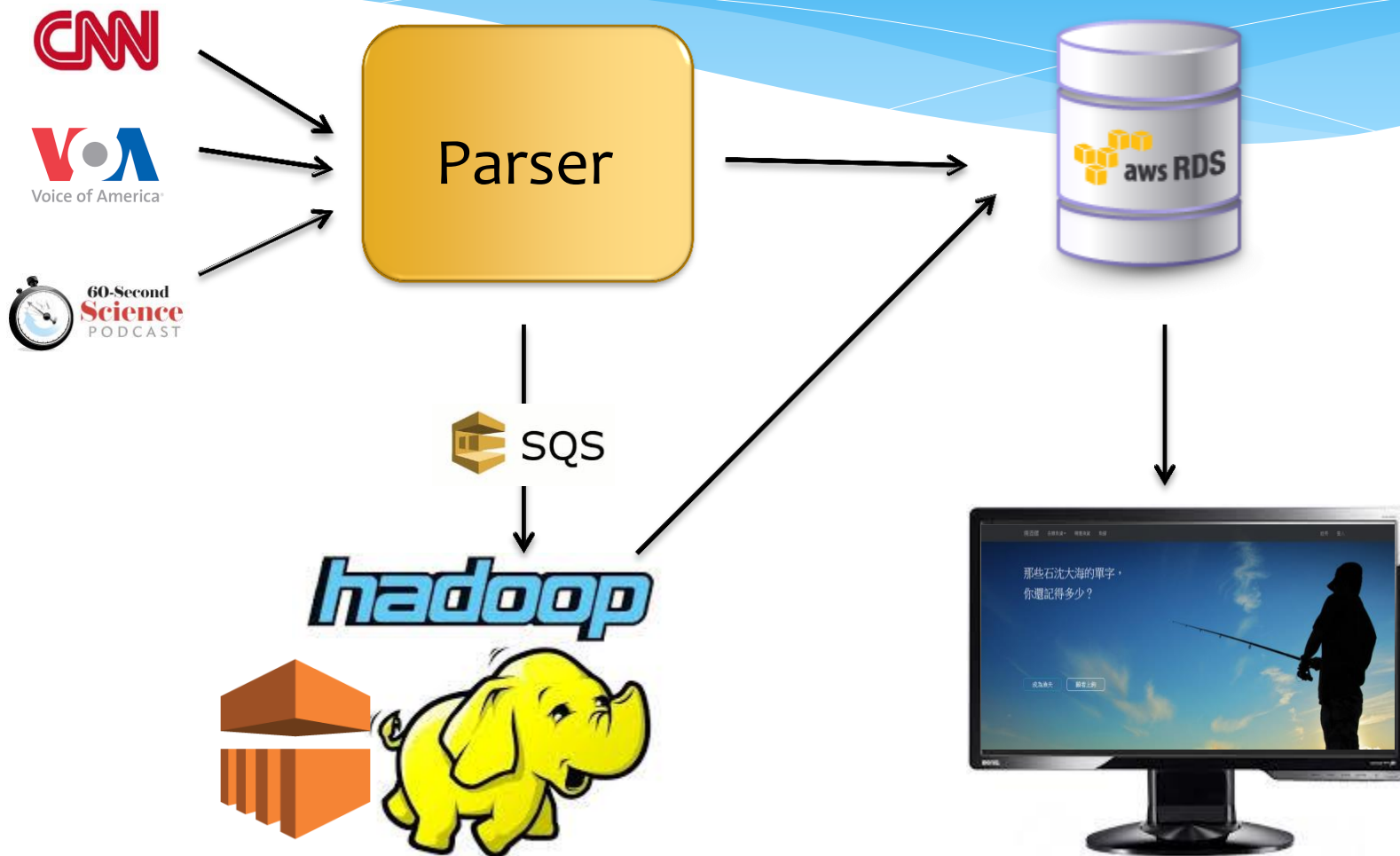
S3

EMR



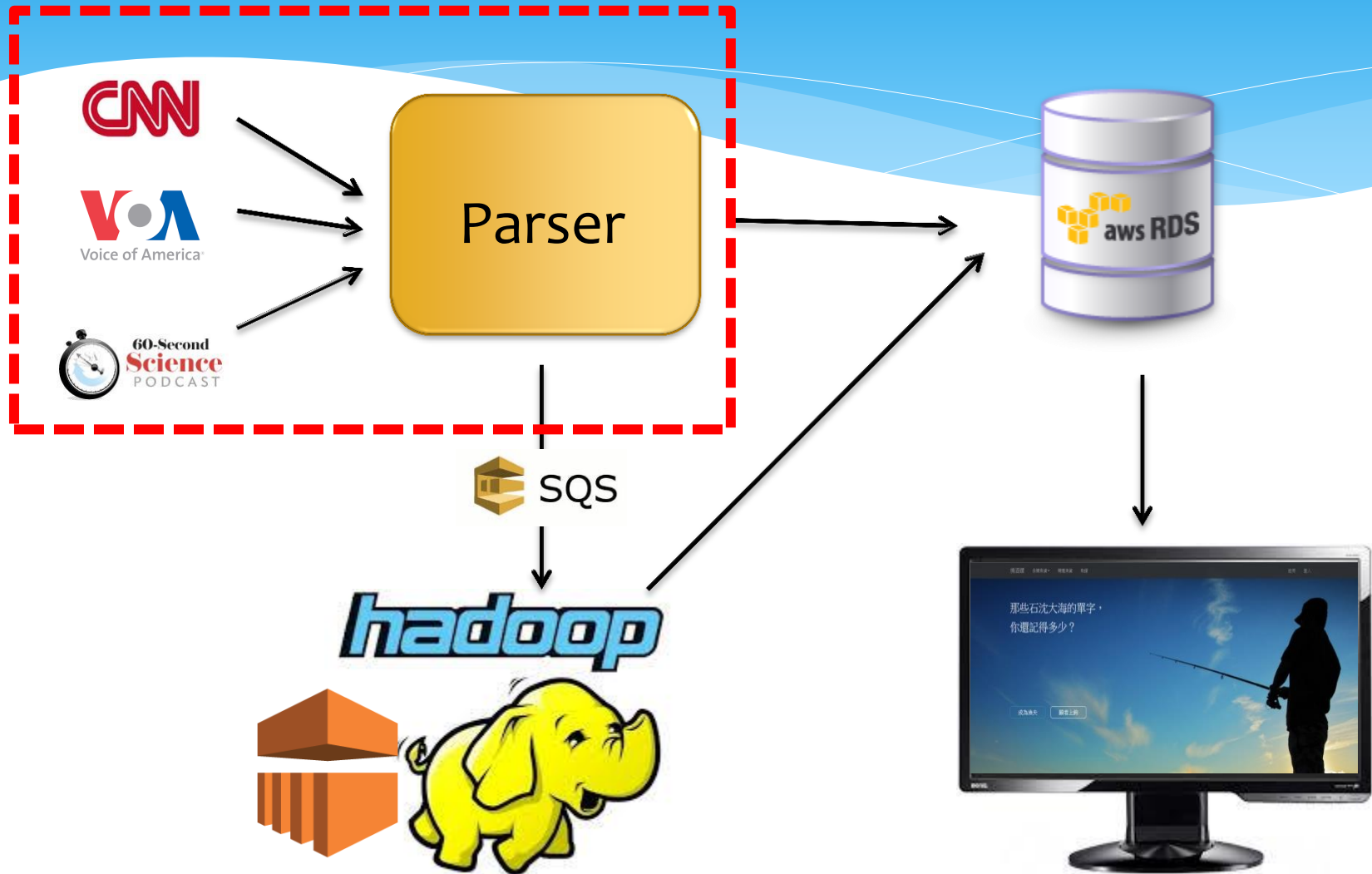


網站架構





Parser





Parser

常用字彙頻率表

根據Wikipedia提供的頻率表，將頻率轉換成難度(0-13)後存入HBase

英漢字典

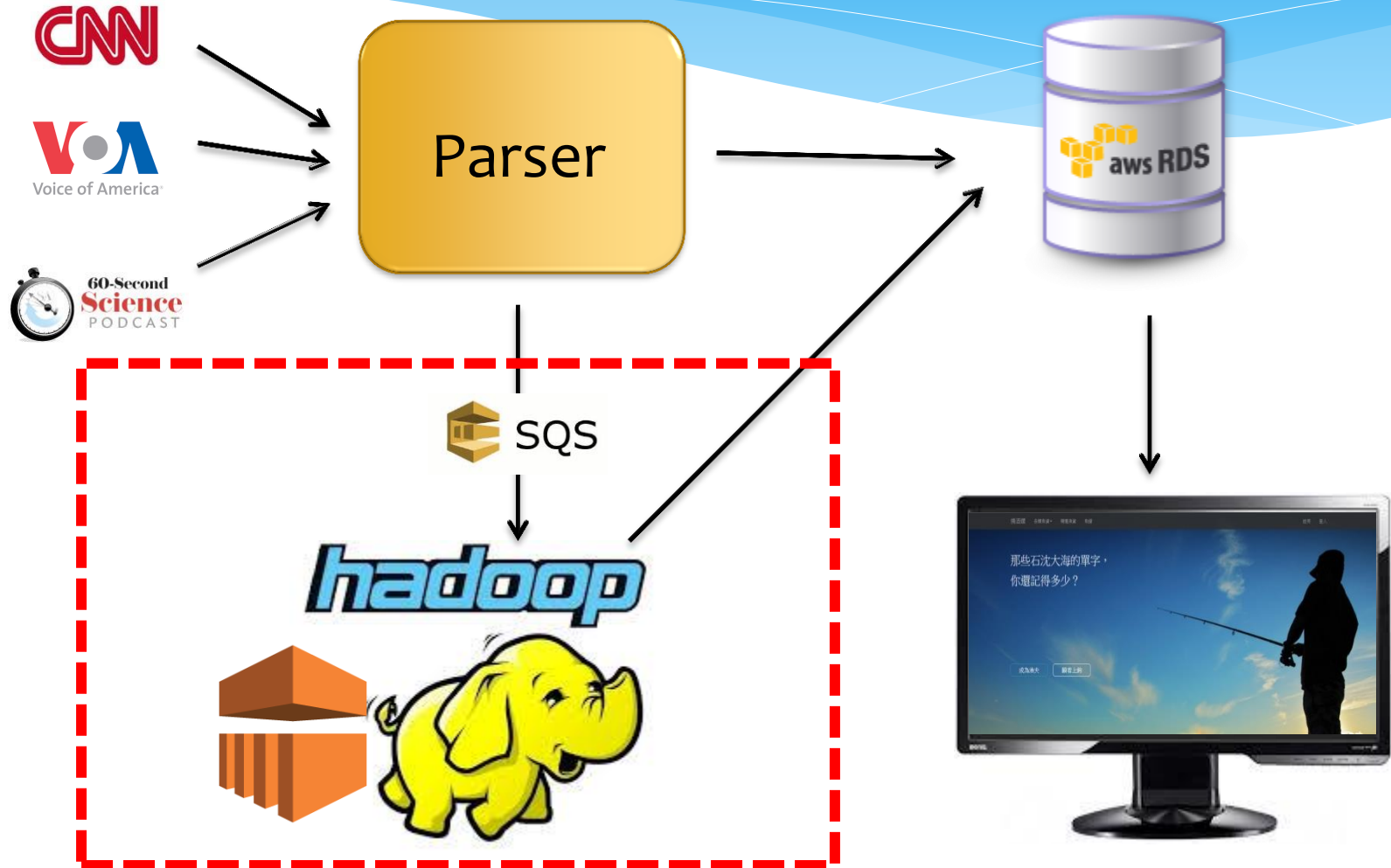
申請「有道詞典」的API，預先查詢常用的三萬~四萬字放入資料庫

最新文章

利用RSS訂閱，每小時至CNN、VOA、Scientific American 收集文章。
把內容輸出到S3，再將S3的連結送入SQS



EMR (Hadoop)





EMR (Hadoop)



啟動EMR

後端用Python做background SQS polling，當文章數量超過十篇即啟動EMR

計算文章難度

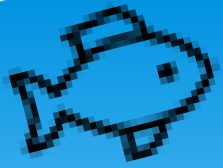
用MapReduce找出文章中不同的單字，從Hbase找出單字的對應難度，將文章依難度分為十級

計算單字難度分布

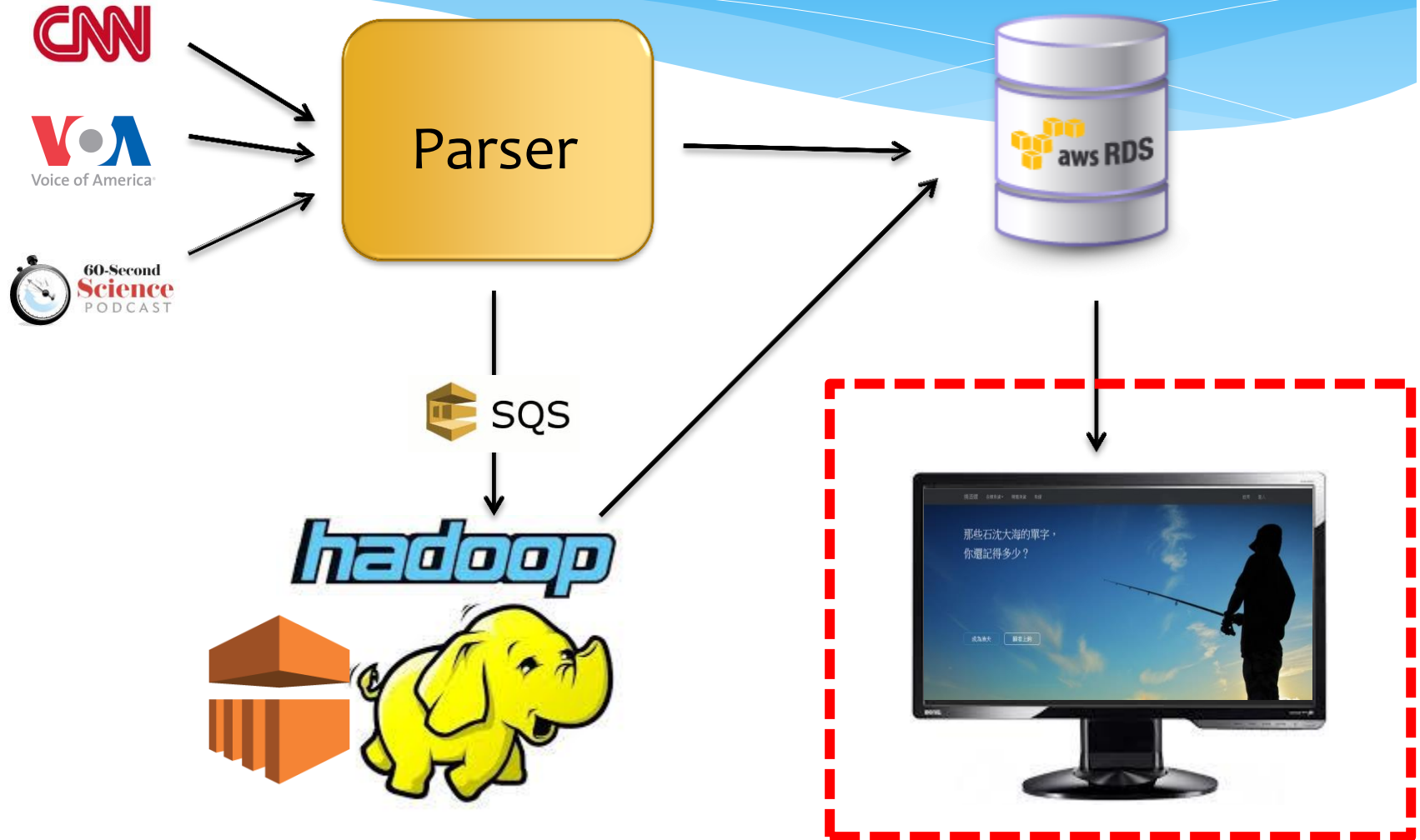
將單字依照難度分成難、中、易三類，並用MapReduce算出各類別在文章中的數量，用來判斷使用者的程度

連接資料庫

將算完的難度存入RDS



Web





Web

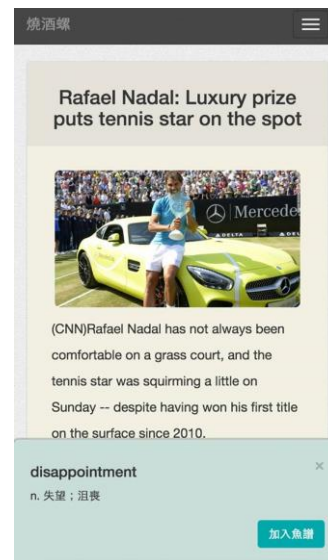
單字查詢

若使用者查詢的單字不在資料庫中，就透過API取得單字解釋，並加入資料庫

文章推薦

根據使用者查詢的單字量及單字難度，推薦適合使用者程度的文章

手機版網頁





實作細節

單字難度計算

文章難度計算

文章推薦



單字難度計算

Wikipedia提供的字彙頻率表其範圍為6~1,200,000，約四萬多筆單字。

將單字做以下轉換：

$$L(w) = \ln \frac{MAX_{wf}}{freq(w)}$$

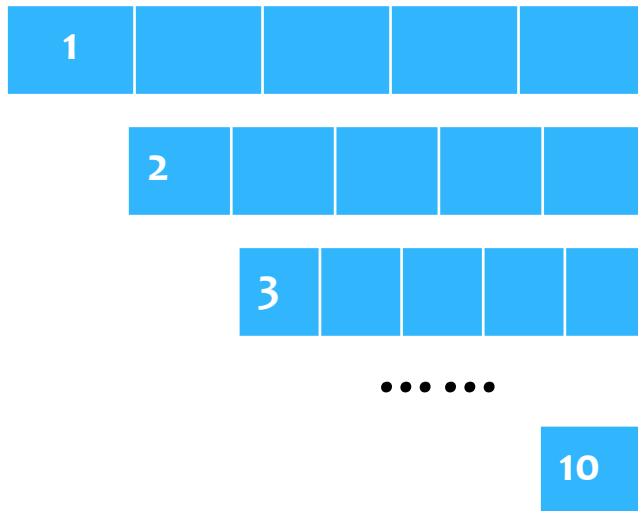
會得到難度 $0 \leq L(w) < 13$

另外收集了stop word單字表，將所有stop word 的難度設為0



文章難度計算

計算每篇文章中有幾種不同的單字，從HBase查詢其難度並加總，得到的分數範圍約在0~3000之間，依照分數做以下的分級：



簡單的文章區間較大
越難的文章則區間越小

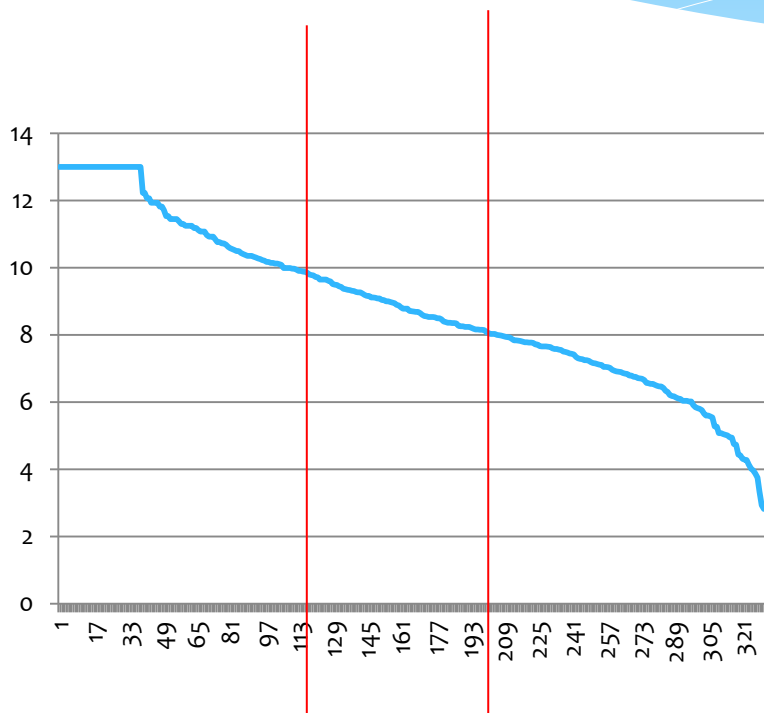


文章推薦

1. 記錄使用者查詢的單字數量及難度
2. 計算使用者查詢的單字中，困難單字佔的比例：
 - (1) 比例高於60% → 文章偏易
則計算使用者查詢了文章中多少比例的困難單字
 - (2) 比例低於40% → 文章偏難
則計算使用者查詢了文章中多少比例的簡易單字



文章推薦





DEMO

<http://54.172.200.50:49152/>