

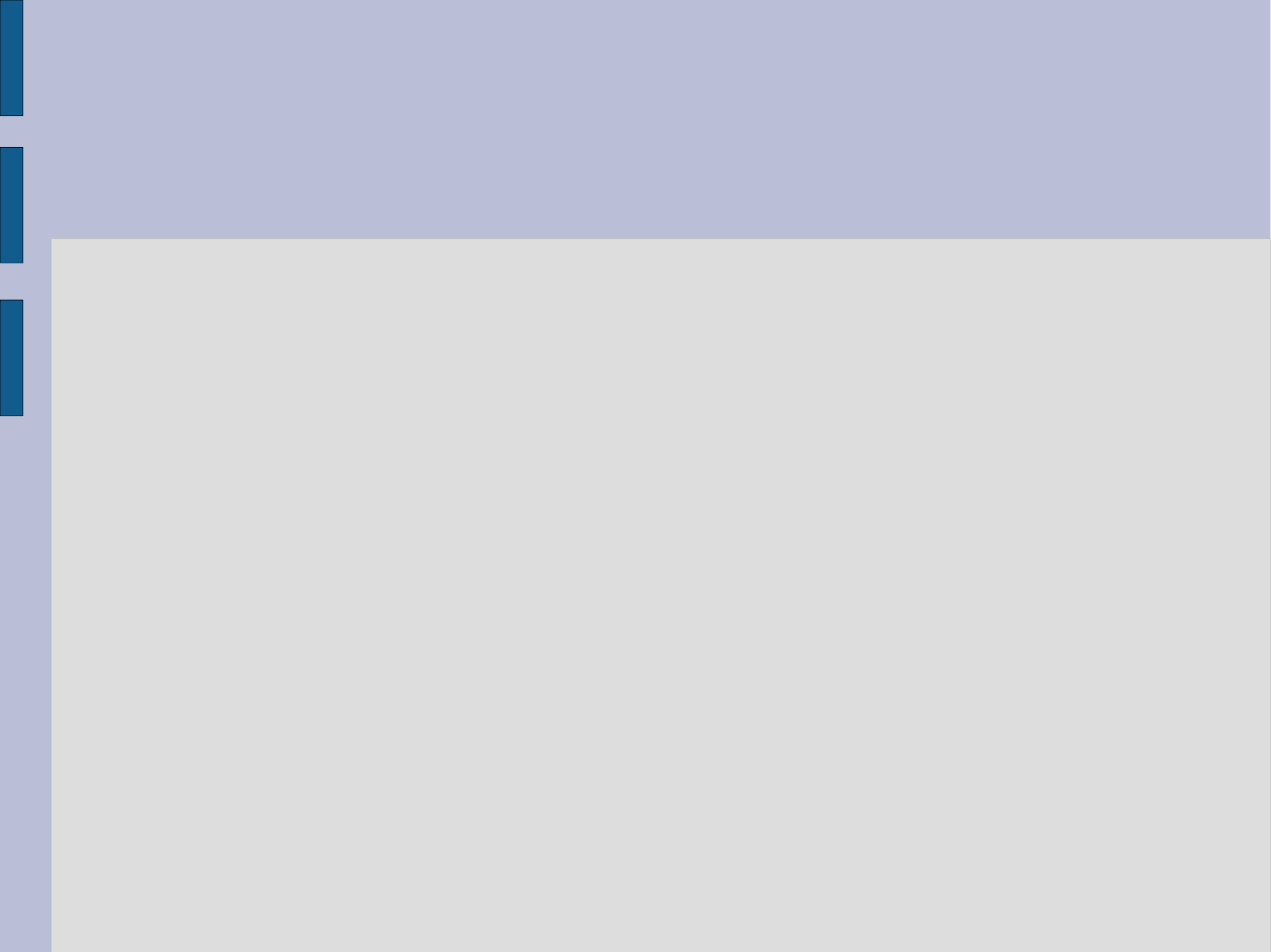
你好，
我是 seamxh

今天要來講

新酷音輸入法

詞庫

詞頻



詞庫

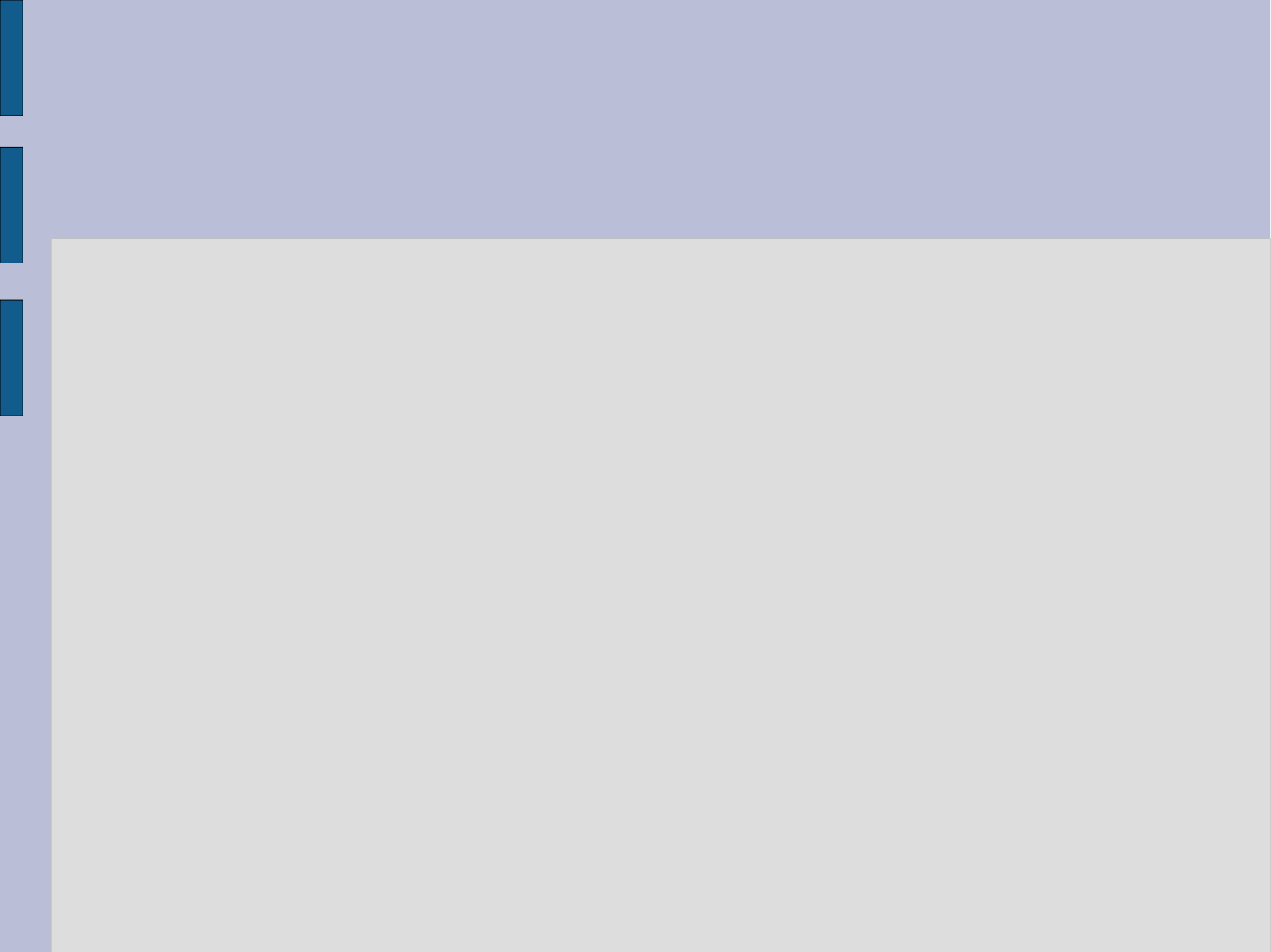
到底 螫 舌 麼 東西

—————

到底螫[☆]舌麼東西

到底是[☆]什麼東西



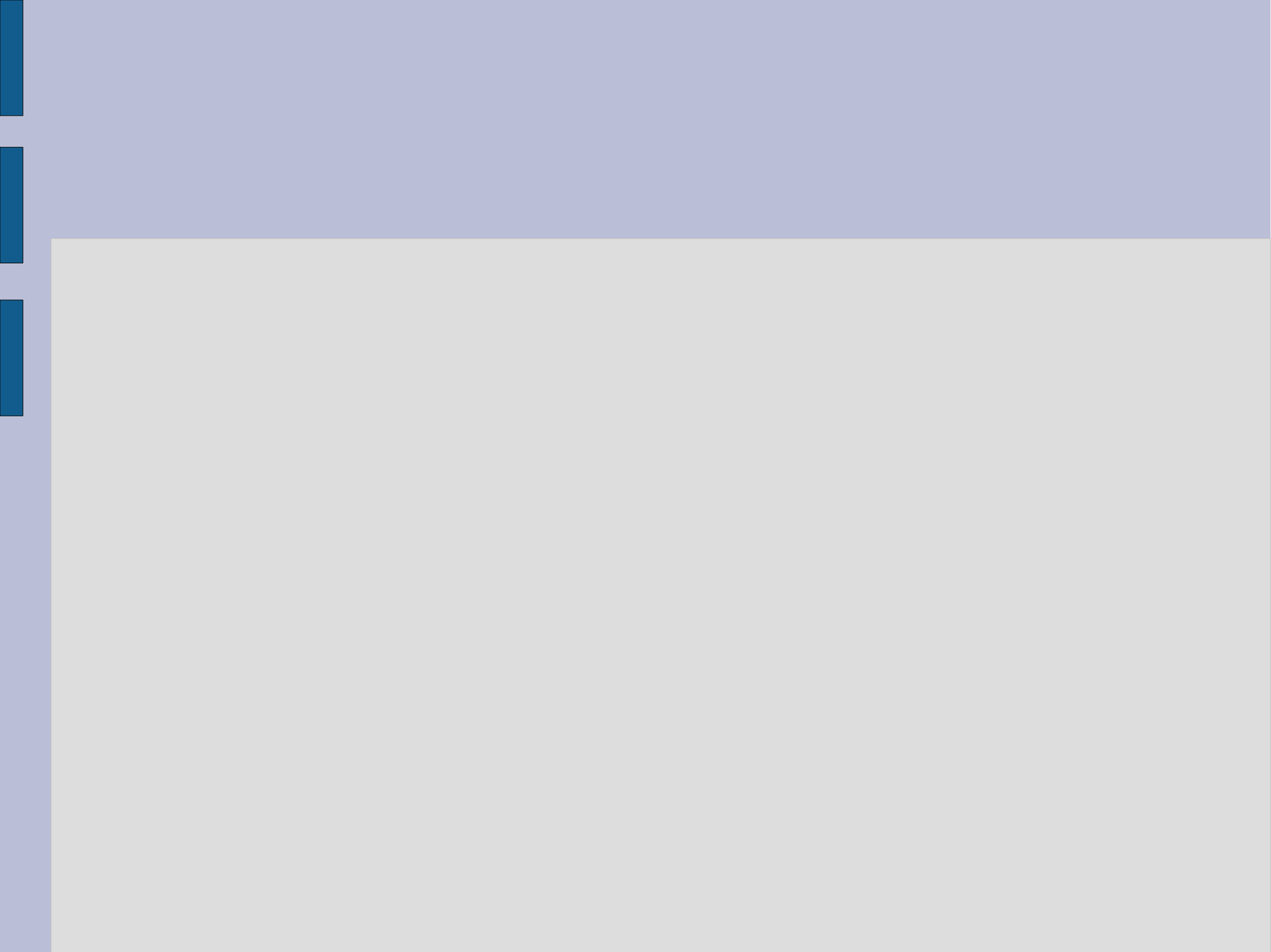


詞頻

讓你省市

讓你省市

讓你省事

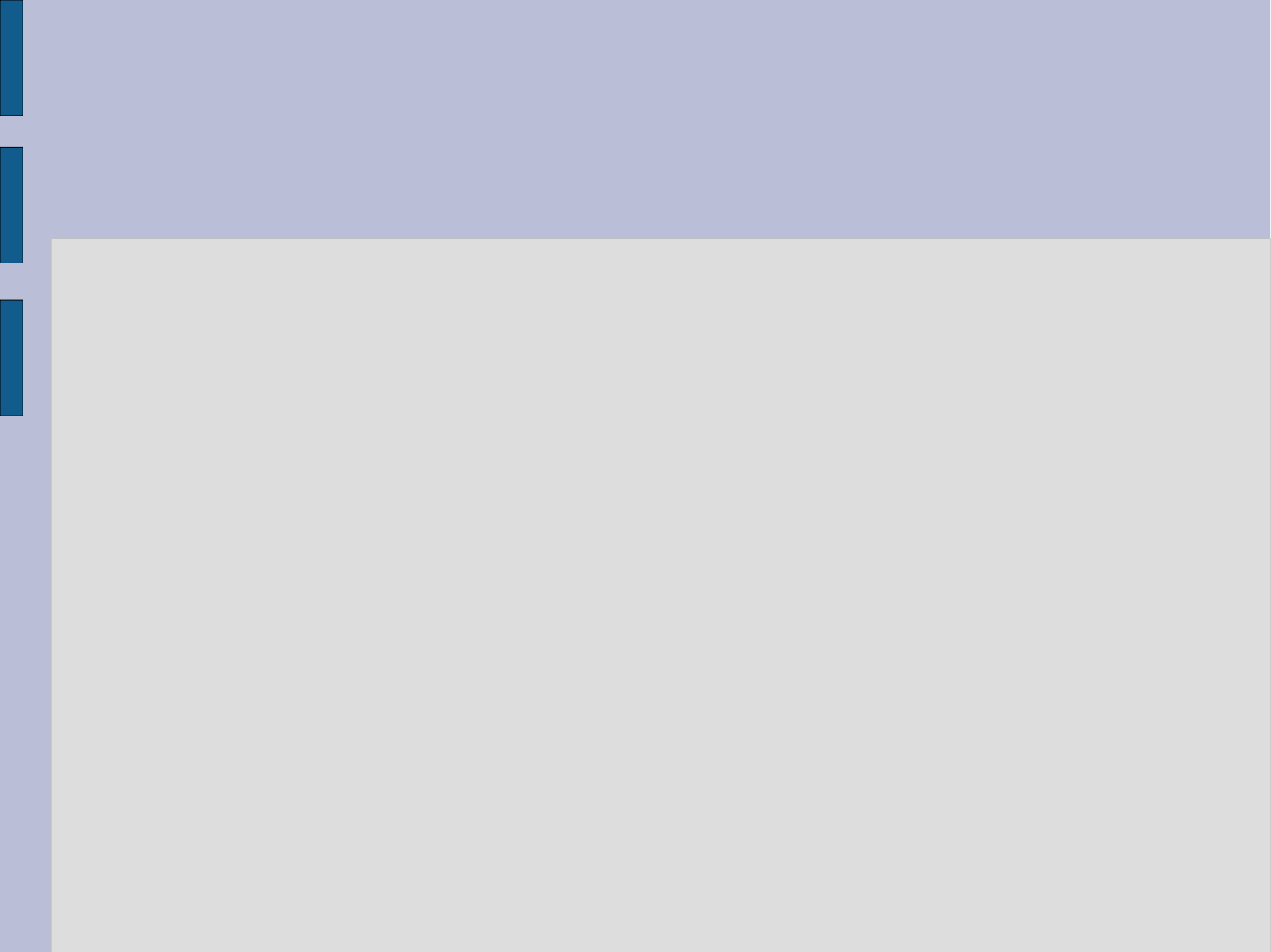


好吧

來講點政經的

真的要開始

講正經的



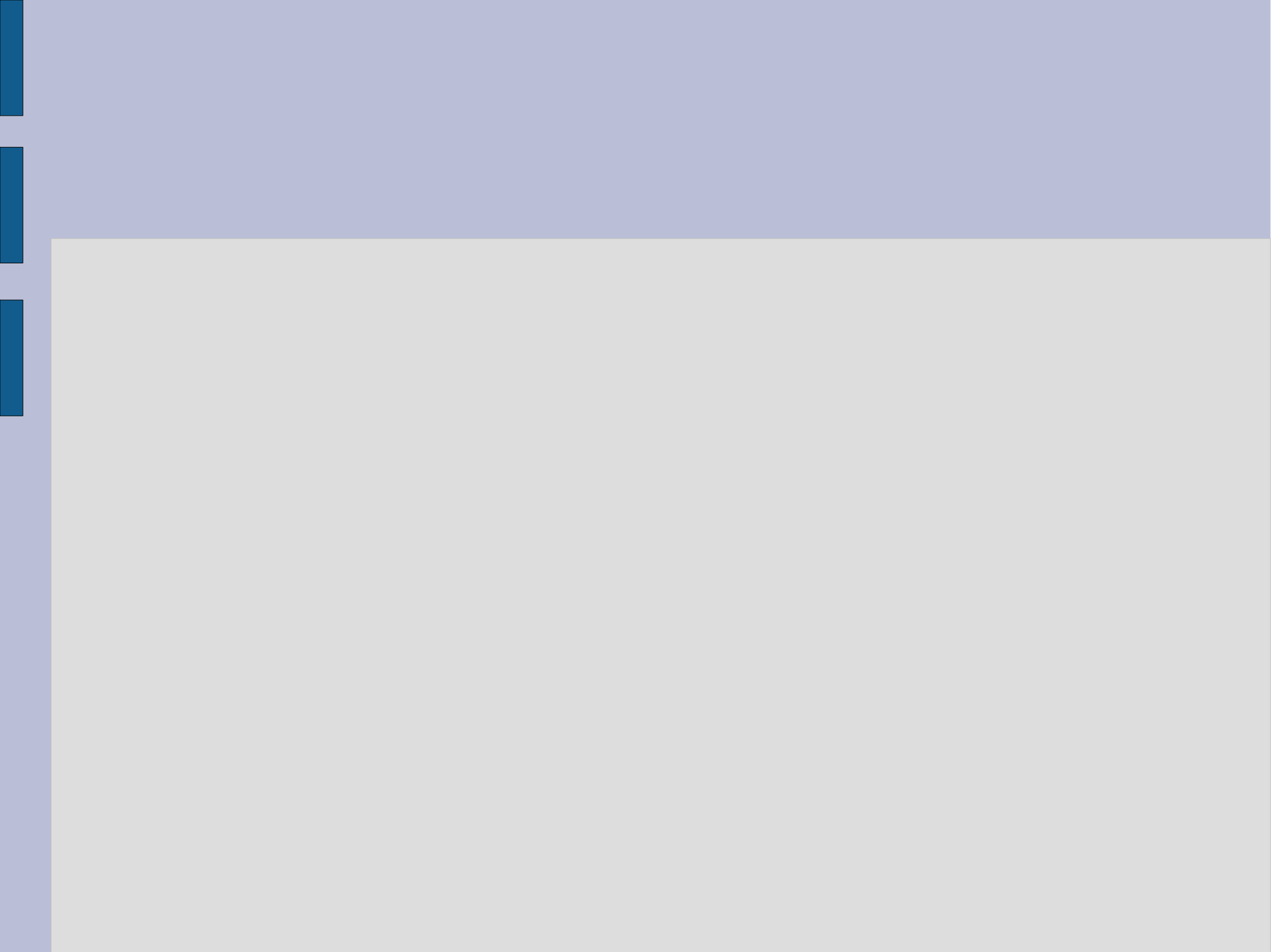
詞庫

詞庫 = 詞典

新酷音使用者

好幸福～

每人都有
兩個辭庫



靜態詞庫

不隨使用者習慣改變

也稱：系統詞庫

160,000

什麼東西

337 (freq)

尸古 ✓ 冂丕 • 勿乂厶 丁一

什麼東西

是個詞，沒錯

甚麼、東西

兩個

什、麼、東、西

算是四個... 詞

麼東西？

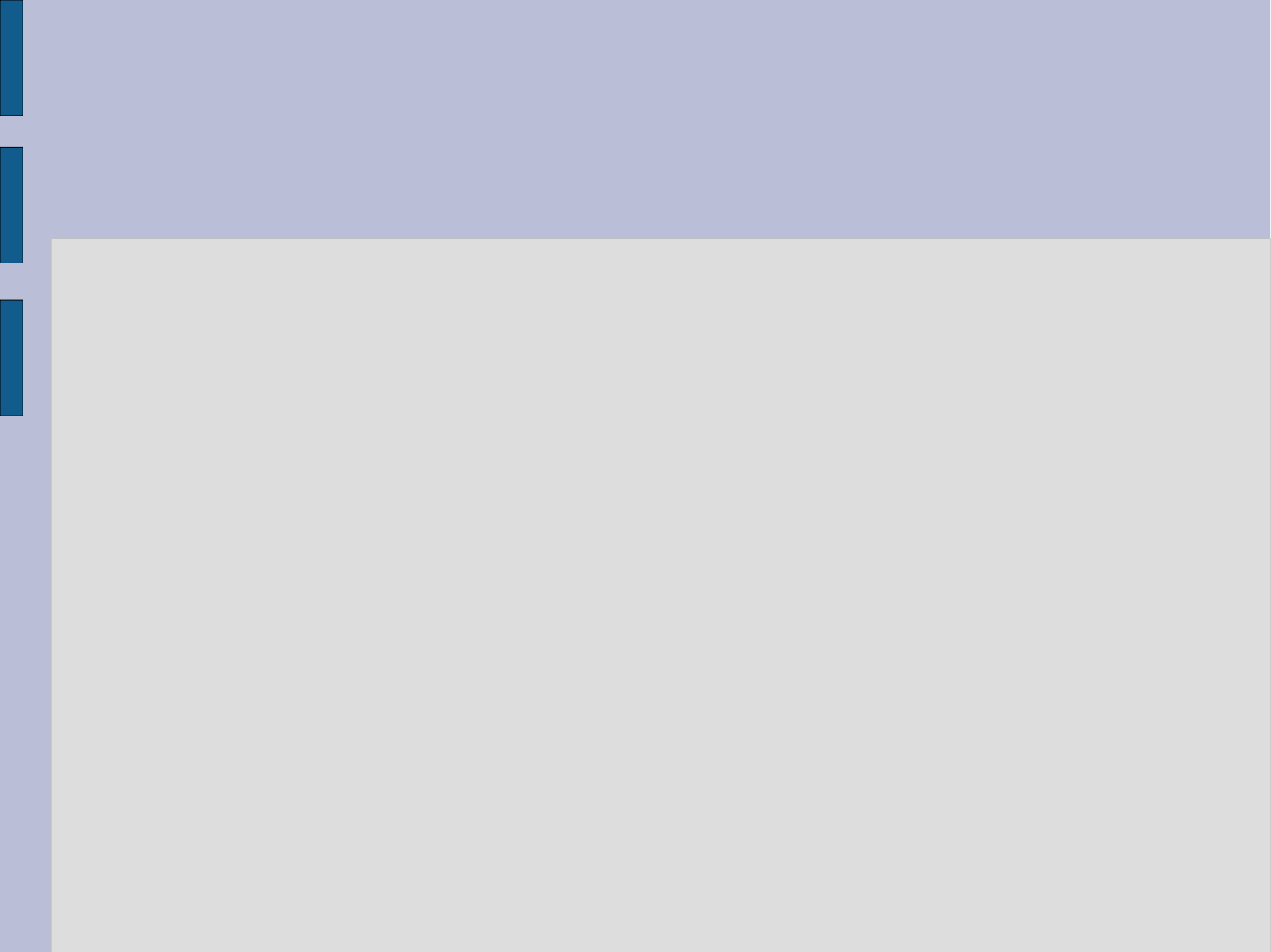
不可否認，詞庫，偶爾也會有 bug

所以



**I WANT YOU
FOR U.S. ARMY**

NEAREST RECRUITING STATION



... 離題一下

使用者詞庫裡有一缸子
永遠刪不掉的單字，

～很正常

什、麼、東、西

也算 是 詞



有一靜，必有一動

動態詞庫

a.k.a. 使用者詞庫

記錄使用者輸入模式

系統詞庫 — 使用 → 動態詞庫

儲存使用者自訂詞彙

ctrl+[1-9]

什麼東西

尸古 ✓ 冂丕 • 勿メ厶 丁一

501 (*freq*)

8873 (*recent time*)

337 (*orig freq*)

2359 (*max freq*)

江湖一點訣

說破才有價值

什麼東西

尸ㄛ ㄛ ㄇㄛ • ㄉㄨㄥ ㄊ一

501 (*freq*, 詞頻)

8873 (*recent time*, 最近一次輸入)

337 (*orig freq*, 原始詞頻)

2359 (*max freq*, 同音詞中最大詞頻)



詞頻

很重要

讓你省事

給你幸福



滿城近視黃金甲

- a) 近世
- b) 進士
- c) 盡是
- d) 近視



盡是

詞頻變化：

近世 —

進士 —

盡是 ↑

近視 —

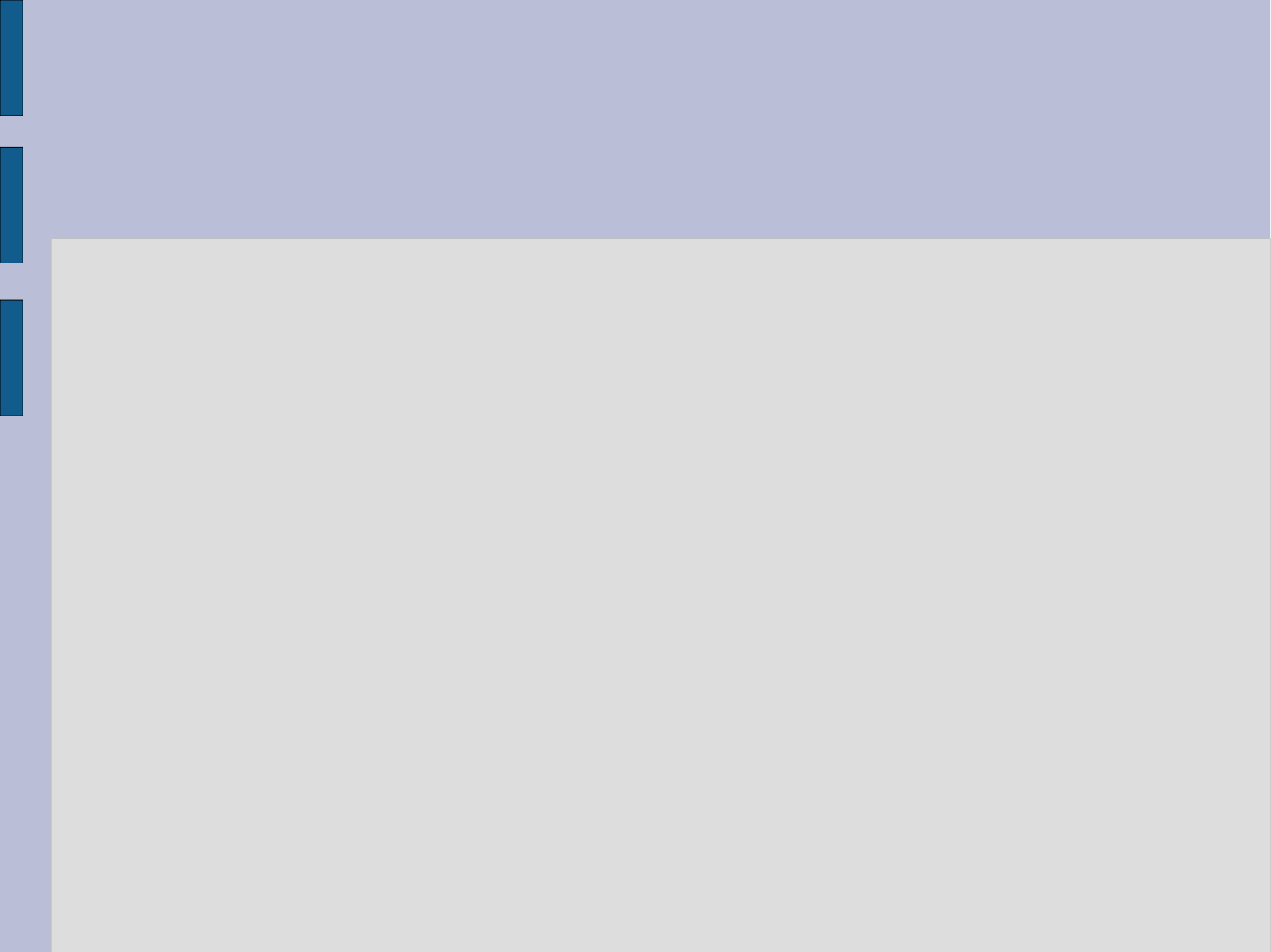
越常用的詞

其詞頻越快升等

$freq += (max\ freq - orig\ freq) / n;$

$n=5,$ if (current time - recent time < 4000)

$n=10,$ if (current time - recent time < 50000)



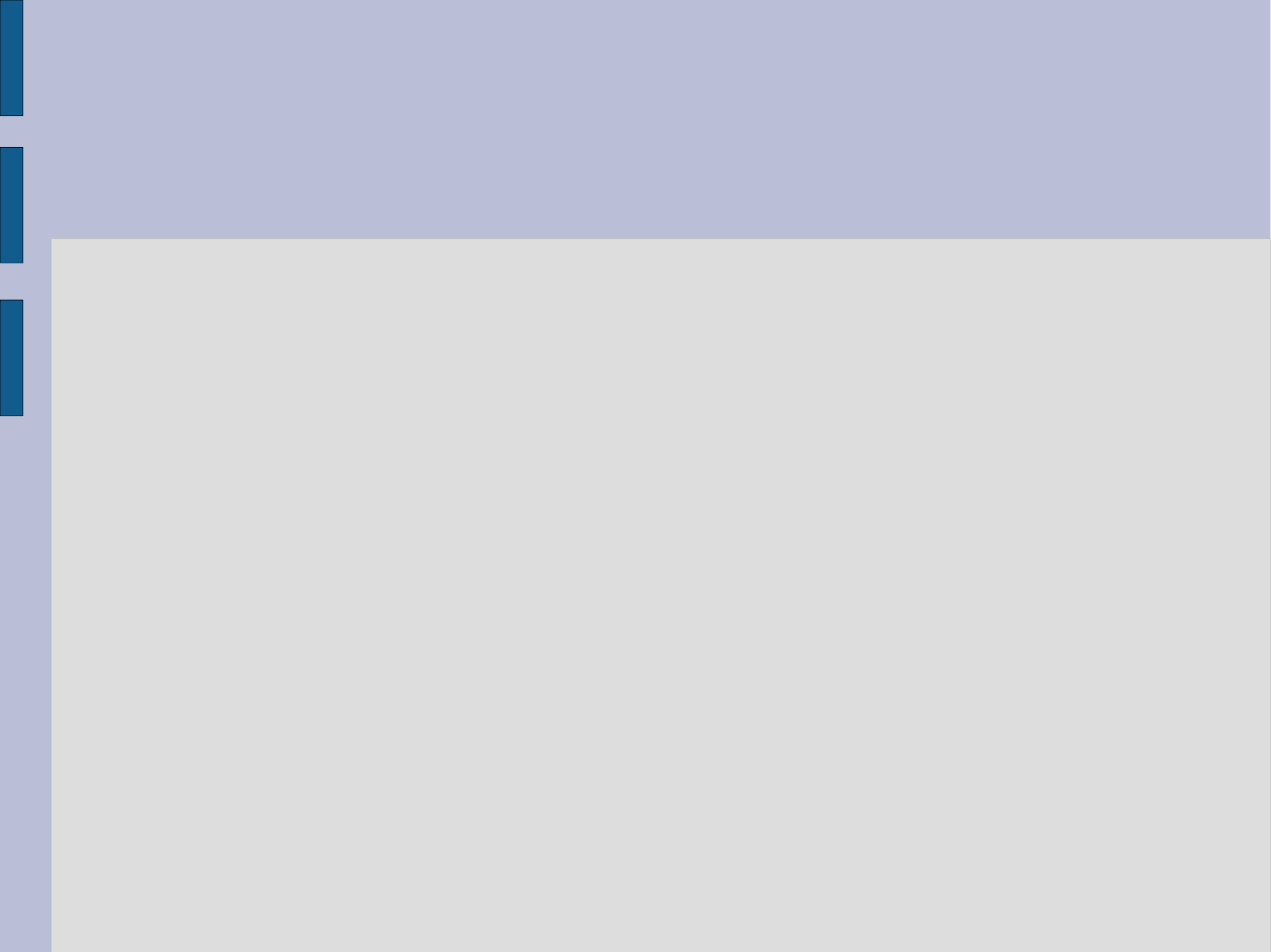
到底 (3023)+ 螫舌 (0)+ 麼 (2536/512)+
東西 (6024)
=9052

到底是 (829)+ 什麼東西 (337)
=1166

修改詞庫後

已解決上述問題

by sky00888



順便說說

斷詞演算法

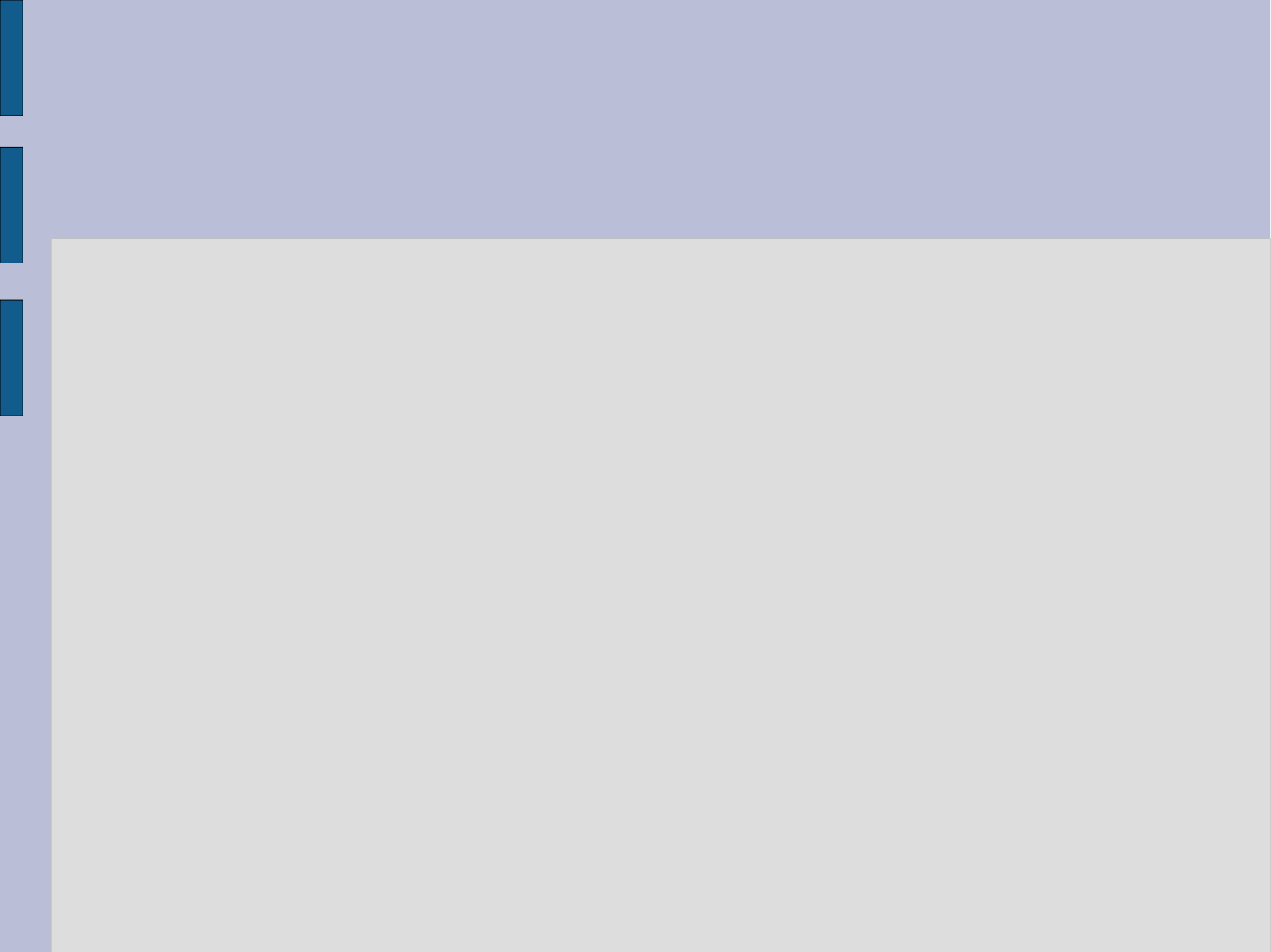
真的是順便說說

因為 一言以蔽

最短路徑

Pass 1:
優先錄用較長詞彙

Pass 2:
以詞頻總和最高者為結果



報告完畢