Build a search engine

(actually just retrieve values from a database, score them, and display results)

Input	DB records	Analysis
query	record 1	score 1
	record 2	score 2
	record 3	score 3
	.	.
	.	.
	.	.
	record n	score n

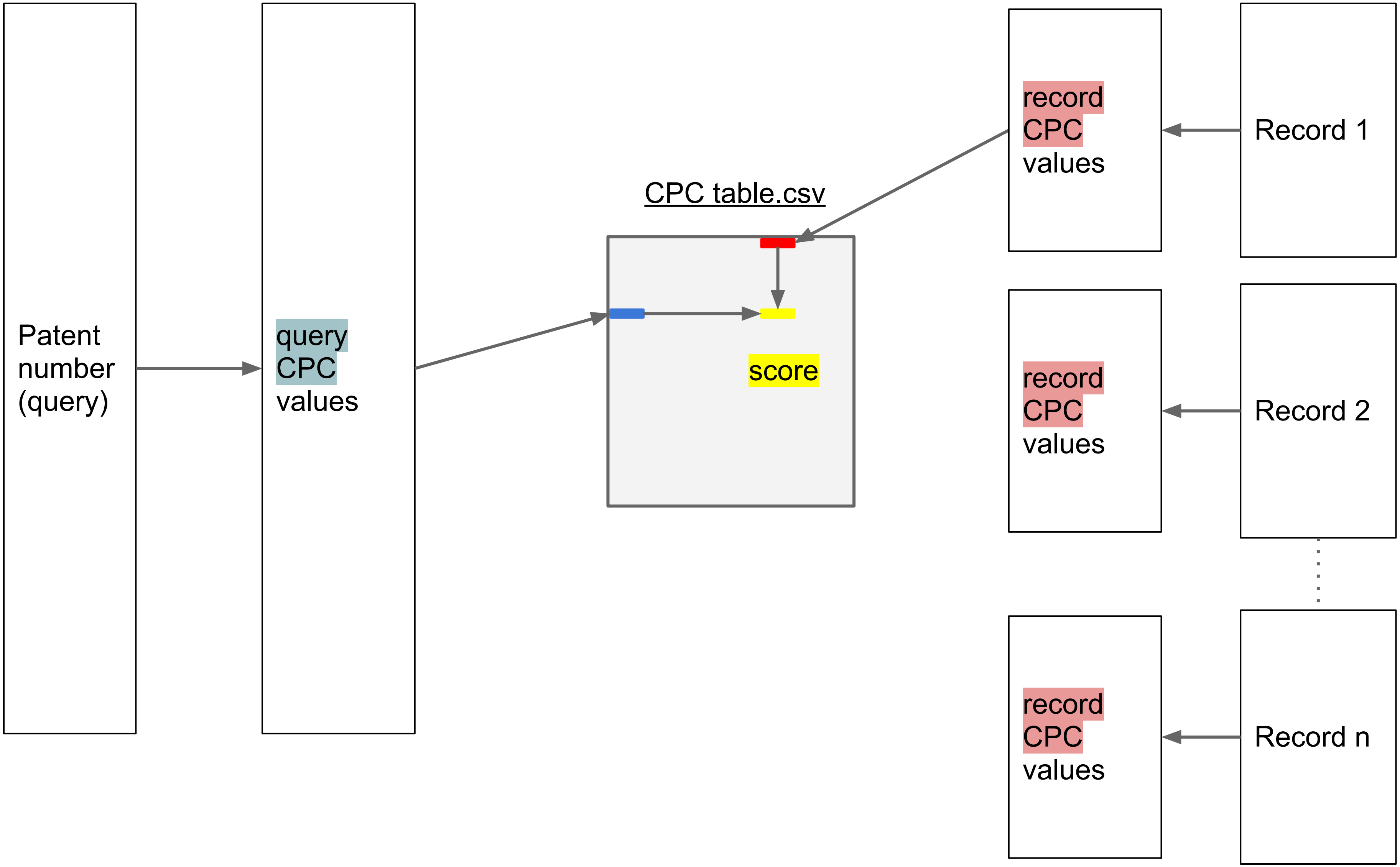
Output

Records 1-n in rank order based on their score

Understand the algorithm

1. user enters a patent number on web page (query)
2. system converts patent number into query CPC data
3. system scores query CPC data to record CPC data
4. system determines max score for each record
5. system returns records in score order

Summary: visualize the task



Patent number (query)

query CPC values

CPC table.csv

score

record CPC values

Record 1

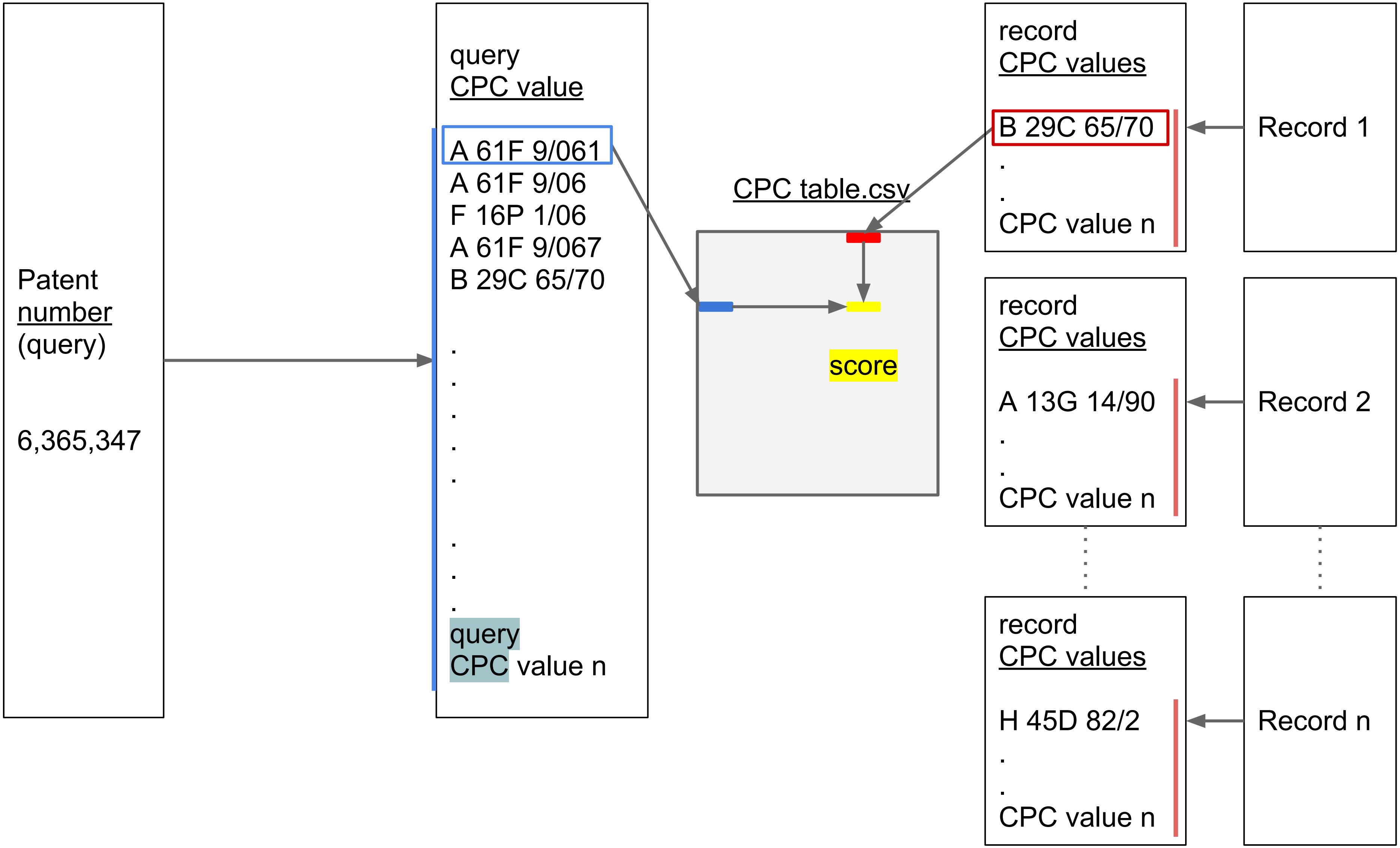
record CPC values

Record 2

record CPC values

Record n

Summary: visualize the task with details



Patent number (query)

6,365,347

query CPC value

A 61F 9/061
A 61F 9/06
F 16P 1/06
A 61F 9/067
B 29C 65/70
.
.
.
.
.
.
.
.
query CPC value n

CPC table.csv

score

record CPC values

B 29C 65/70
.
.
CPC value n

Record 1

record CPC values

A 13G 14/90
.
.
CPC value n

Record 2

record CPC values

H 45D 82/2
.
.
CPC value n

Record n

Summary: visualize the task of each CPC pair

Description

Column A: user input (query)

Column C: query CPC

(from Pat to CPC table.csv)

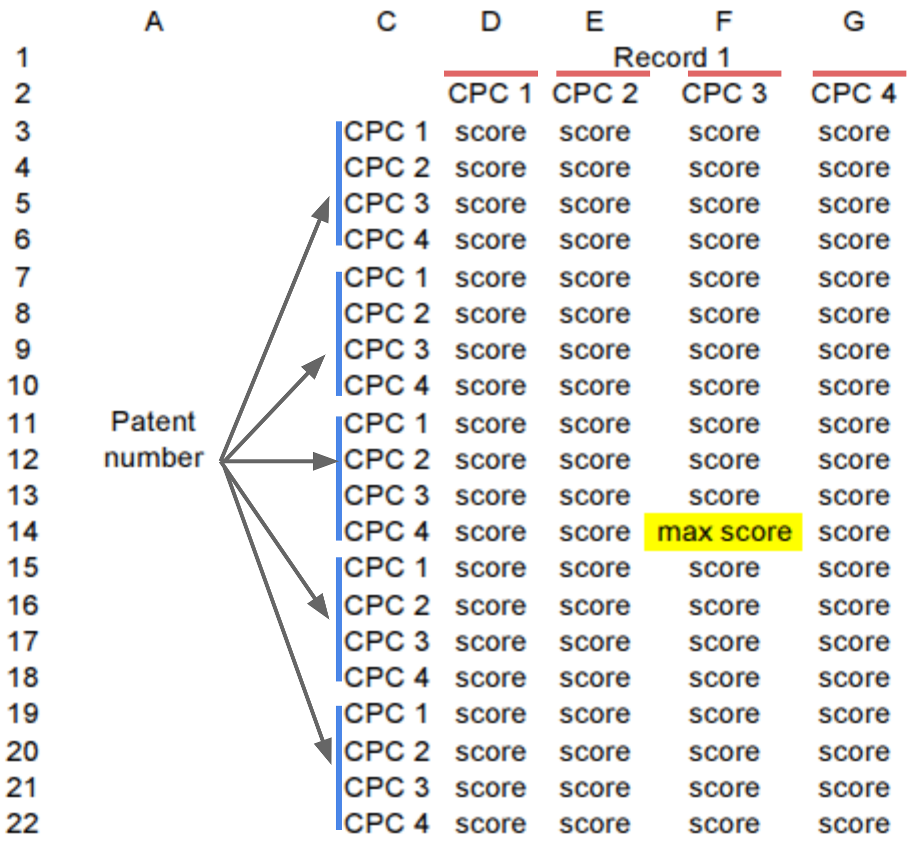
Row 1: records 1 - n

Row 2: record CPC value(s)

D3:G22: score

(from CPC table.csv)

F14: max score (record 1 score)



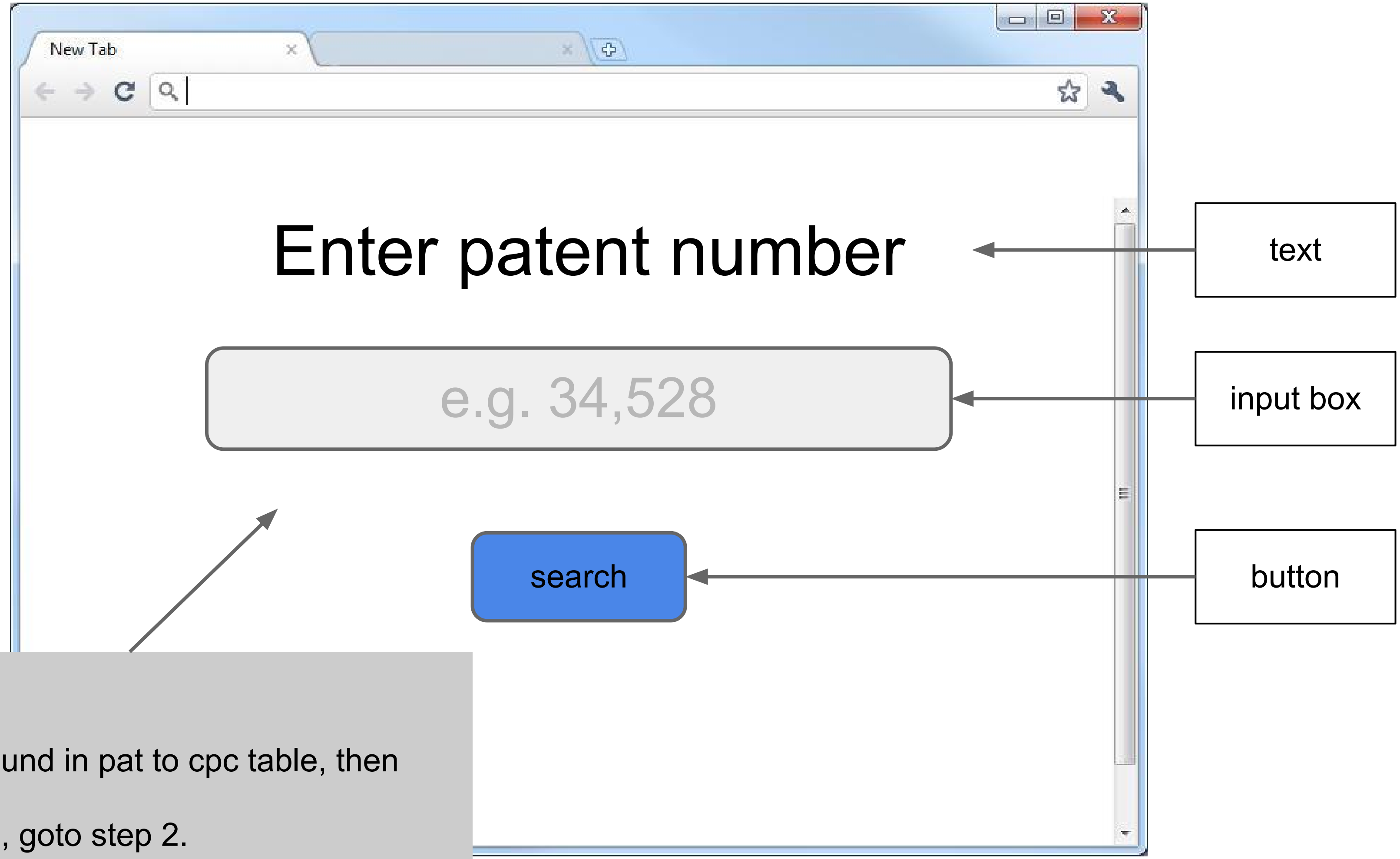
	A	C	D	E	F	G
1				Record 1		
2			CPC 1	CPC 2	CPC 3	CPC 4
3		CPC 1	score	score	score	score
4		CPC 2	score	score	score	score
5		CPC 3	score	score	score	score
6		CPC 4	score	score	score	score
7		CPC 1	score	score	score	score
8		CPC 2	score	score	score	score
9		CPC 3	score	score	score	score
10		CPC 4	score	score	score	score
11	Patent	CPC 1	score	score	score	score
12	number	CPC 2	score	score	score	score
13		CPC 3	score	score	score	score
14		CPC 4	score	score	max score	score
15		CPC 1	score	score	score	score
16		CPC 2	score	score	score	score
17		CPC 3	score	score	score	score
18		CPC 4	score	score	score	score
19		CPC 1	score	score	score	score
20		CPC 2	score	score	score	score
21		CPC 3	score	score	score	score
22		CPC 4	score	score	score	score

Understand the algorithm (review)

1. user enters a patent number on web page (query)
2. system converts patent number into query CPC data
3. system scores query CPC data to record CPC data
4. system determines max score for each record
5. system returns records in score order

Step 1: user enters patent number

Task: make a very simple web page.

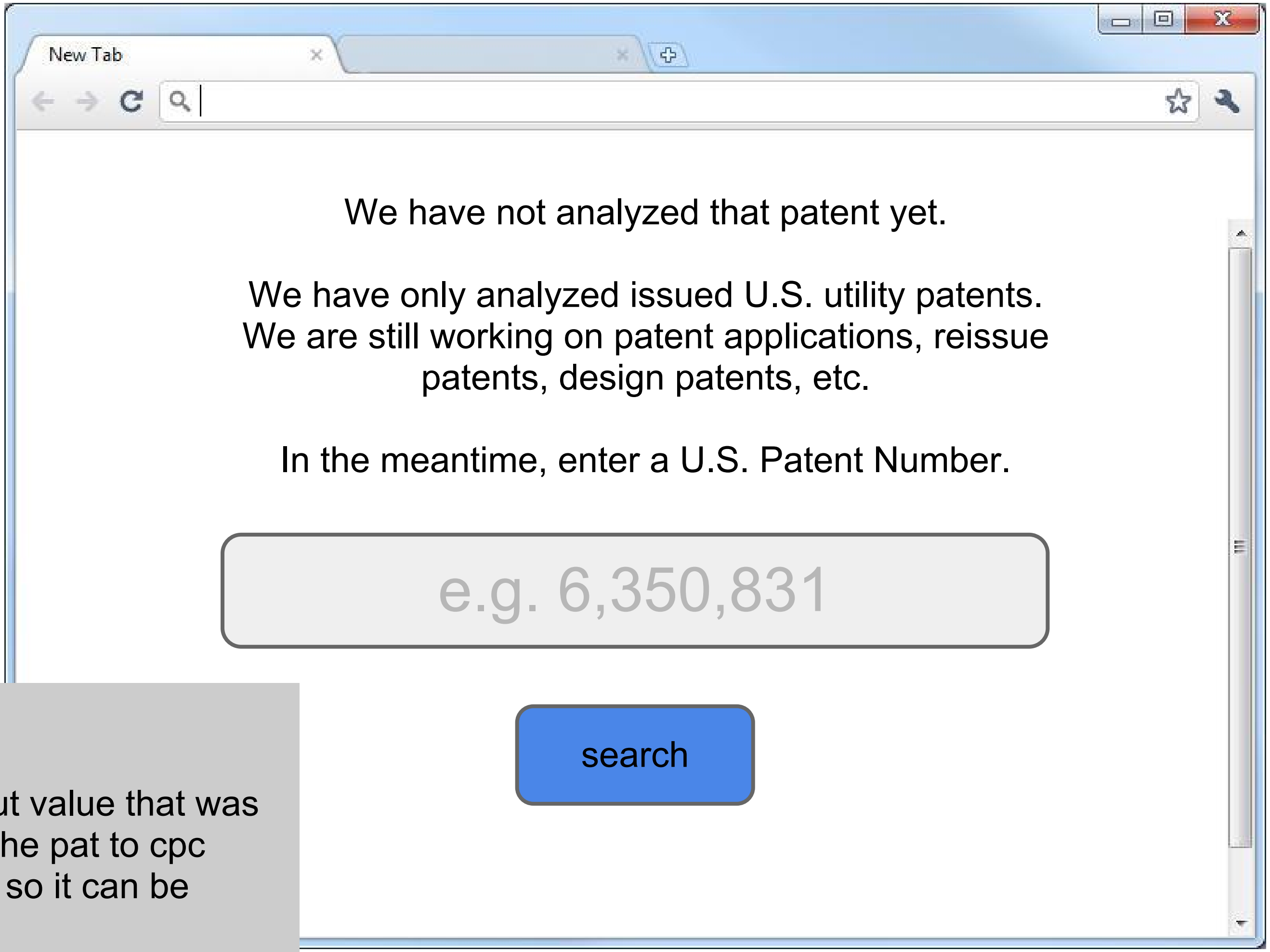


Note:

If query not found in pat to cpc table, then goto step 1a.
If query found, goto step 2.

Step 1a: user enters invalid format

Task: make a simple error web page.



We have not analyzed that patent yet.

We have only analyzed issued U.S. utility patents. We are still working on patent applications, reissue patents, design patents, etc.

In the meantime, enter a U.S. Patent Number.

Note:

store the input value that was not found in the pat to cpc table in a log so it can be investigated

Step 2: convert query patent number into query CPC data

Use the attached file to associate a query patent number to query CPC value(s).

See: Pat to CPC.csv

patent	CPC	CPC	CPC	CPC	CPC
0034528	A 01B 3 26 O	A 01B 3 06 X	A 01B 3 14 X	A 01B 13 02 X	A 01B 13 08 X
0034528	A 01B 3 12 O	A 01B 31 00 X	A 01B 19 02 X	A 01B 35 26 X	A 01B 37 00 X
0034528	E 01C 23 121 O	A 01B 13 16 X	A 01B 17 004 X	A 01B 35 06 X	A 01B 63 104 X
0034528	A 01B 17 00 O	A 01B 29 046 X	A 01B 35 16 X	A 01B 39 28 X	A 01B 63 163 X

Notes:

Some CPC values in the table may be blank. That is ok.

Each CPC value has a trailing "O" or "X". That value can be ignored in the analysis.

The final slide has details about this data, a spec to understand it, and sample values.

Step 3a: score query CPC value to record CPC value

The cpc values (query and record) start with a letter A-H, or Y.

If the query CPC value letter matches the record CPC value letter, then pull the corresponding CPC table and goto step 3b.

If the query CPC value letter does not match the record CPC value letter, then the score = 0. Goto next record CPC value.

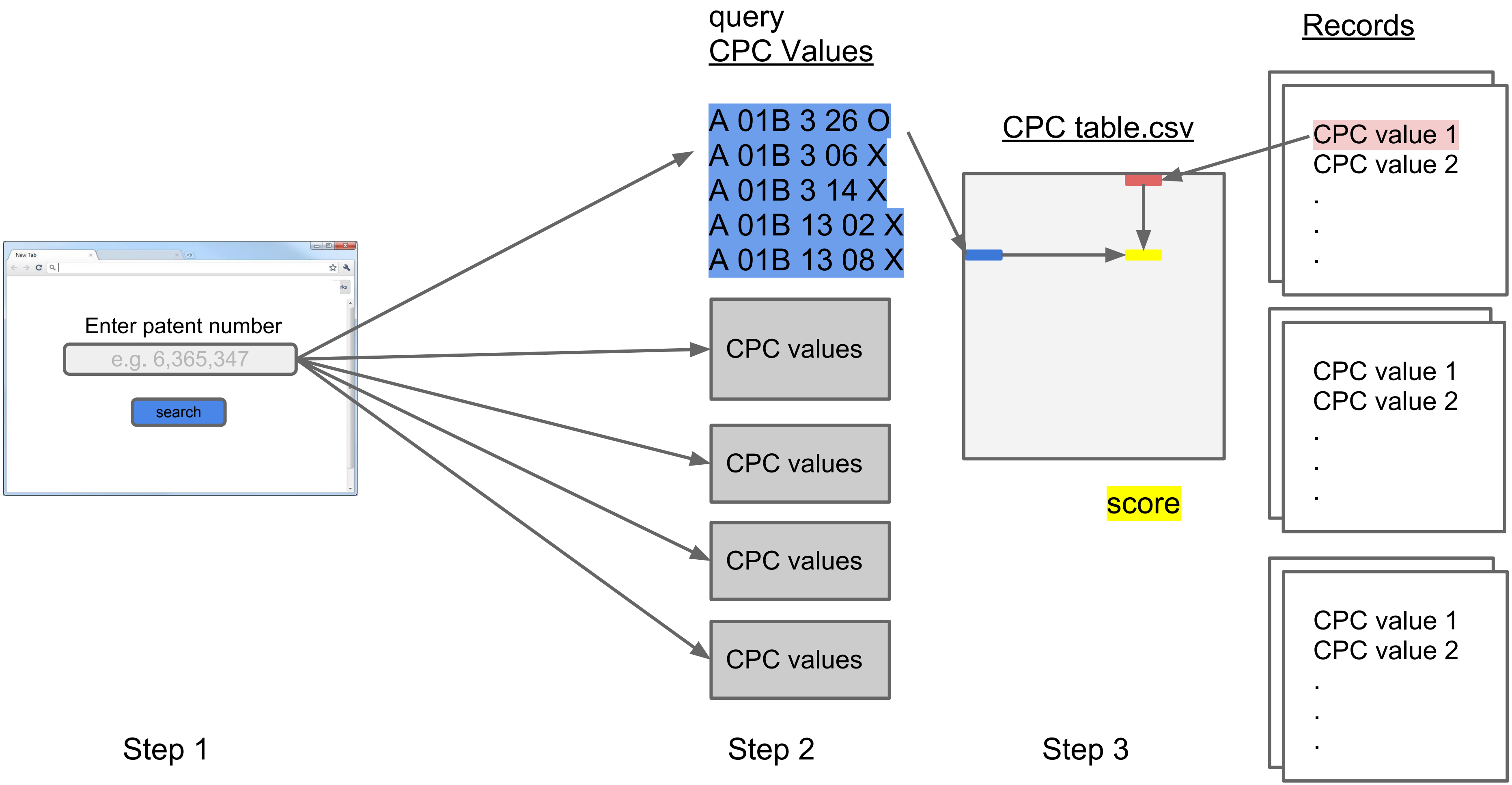
Example:

query CPC	record CPC	score	reason
A 01B 3 14	A 01B 3 26	goto step 3b.	A = A
E 01C 23 121	A 01B 3 26	0	E does not = A

Step 3b: score query CPC value to record CPC value

Each record lists record CPC value(s). Use the attached table to associate record CPC values to query CPC values. Each CPC pair results in a score.

See: CPC table.csv.



query
CPC Values

A 01B 3 26 O
A 01B 3 06 X
A 01B 3 14 X
A 01B 13 02 X
A 01B 13 08 X

CPC values

CPC values

CPC values

CPC values

CPC table.csv

score

Records

CPC value 1
CPC value 2
.
.
.

CPC value 1
CPC value 2
.
.
.

CPC value 1
CPC value 2
.
.
.

Step 4a: find max score for the first record

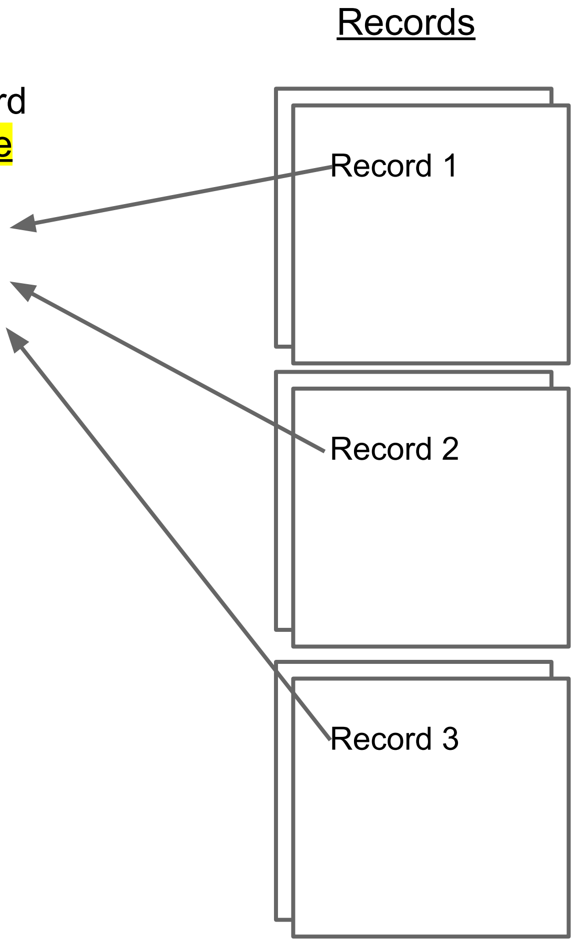
Find the highest query CPC value to record CPC value score.

query CPC value	record 1, record CPC value 1	score
A 01B 3 26 O	A 01B 3 26 O	200
A 01B 3 06 X	A 01B 3 26 O	*from step 3b*
A 01B 3 14 X	A 01B 3 26 O	*from step 3b*
A 01B 13 02 X	A 01B 3 26 O	*from step 3b*
A 01B 13 08 X	A 01B 3 26 O	*from step 3b*
	.	
	.	
	.	
	record 1, record CPC value **2**	
	.	
	.	
	.	

Step 4b: score all records

Repeat step 3a, 3b, and 4a for all records.

Patent number	Record	record score
6,365,347	1	200
	2	170
	3	120
	4	60
	5	150
	.	.
	.	.
	.	.
	n	n



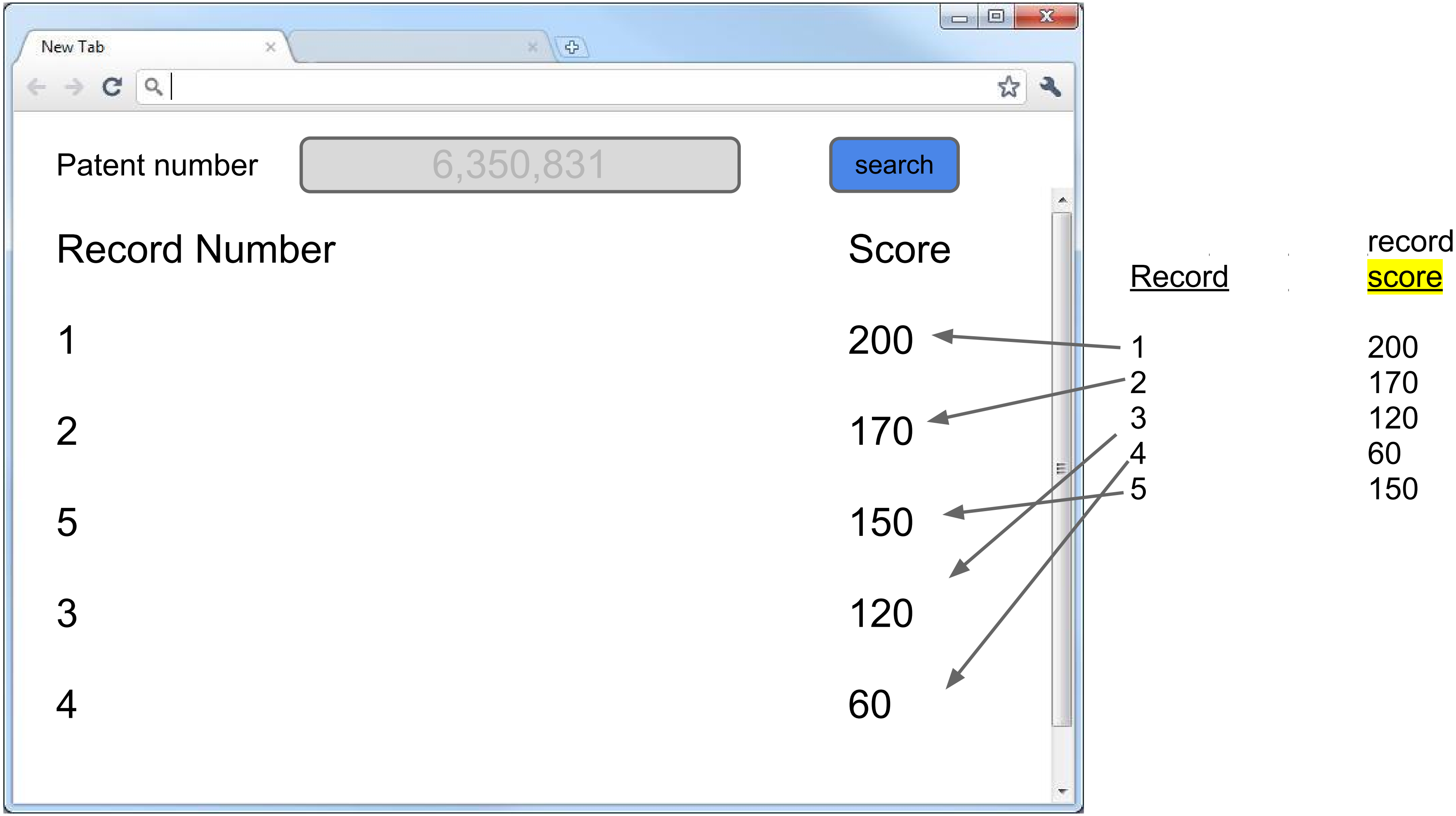
Just checking

Are you a human that is reading this,
or are you a spam bot?

Enter this code in the beginning of
your response if you are human.

Code: P8b4C

Step 5: system returns records to user in ranked order



Patent number: 6,350,831

Record Number	Score
1	200
2	170
5	150
3	120
4	60

Record	record score
1	200
2	170
3	120
4	60
5	150

Pseudo code (for your consideration only)

```
1   Covert patent number into CPC values
2   Retrieve all record CPC values from record n
3   Set record n score = 0
4   If query CPC letter = record CPC letter, goto next.  Else score = 0
5   Score first query CPC to first record CPC
6       Get score.  If score > record n score, then record n score = score
7       Else, goto next
8   Loop lines 4-7 until each query CPC scored against each record CPC
9   Save record n score; goto next record
10  Loop lines 3-9 until all records are scored.
11  Sort all records from largest to smallest score
12  Output records in score order to results page of website
```

Understand the data

Pat to cpc.csv*	This is the full set of data to use
Pat to cpc - sample	This is a small subset to use for testing
Pat to cpc spec.doc	This describes how this data is formatted

* Note: this file is large. About 500MB zipped, 2.5 GB unzipped and millions of lines long.

cpc table(s).csv*	There are the nine tables (A-H, and Y)
cpc table - sample	This is a small subset to use for testing

* Note: these tables are large. Only a few MBs zipped, but massive if unzipped. If they were all combined, it would be a square table of 300k x 300k or about 90 billion values.

Understand the data

record 1-10.csv	This is the full set of data to use
record spec.doc	This describes how this data is formatted

Note: in reality there are about 10,000 records. Just an FYI for considering performance issues.