

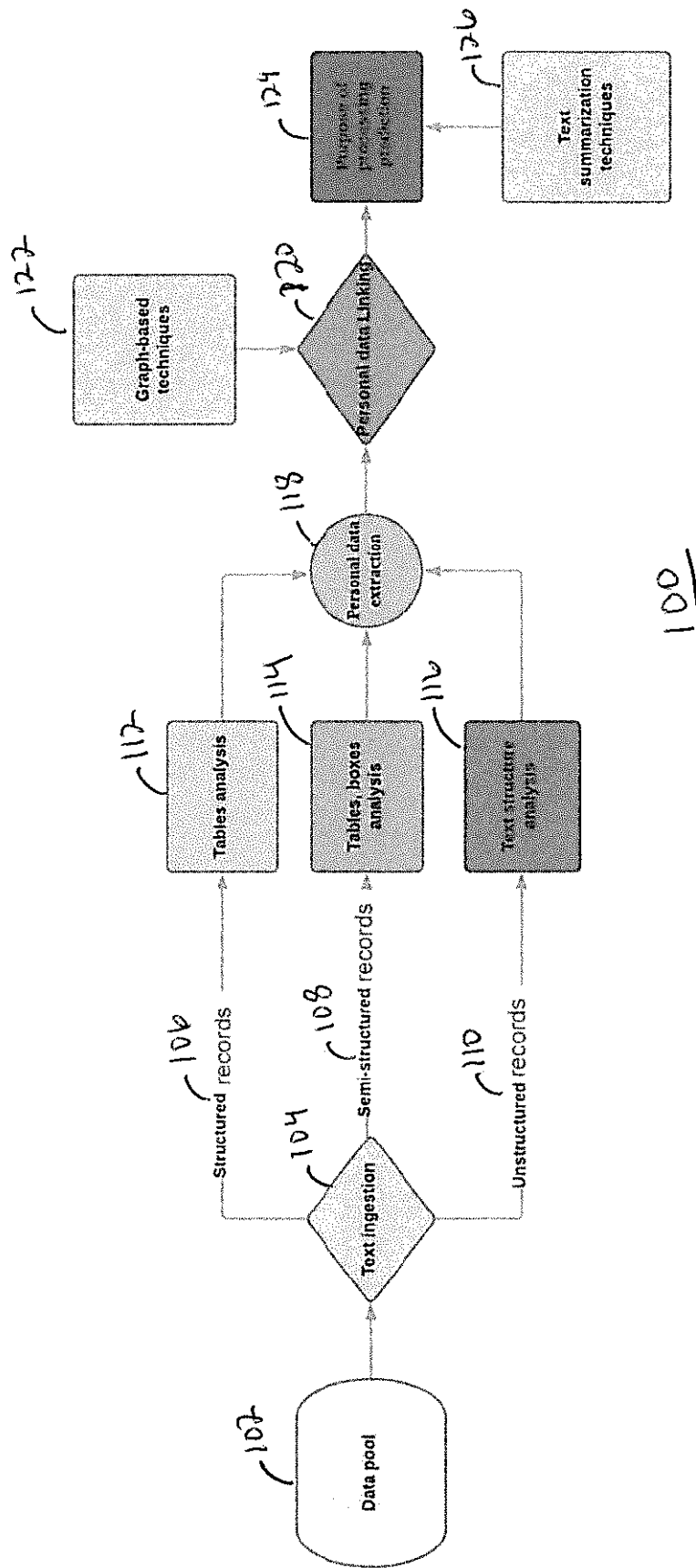
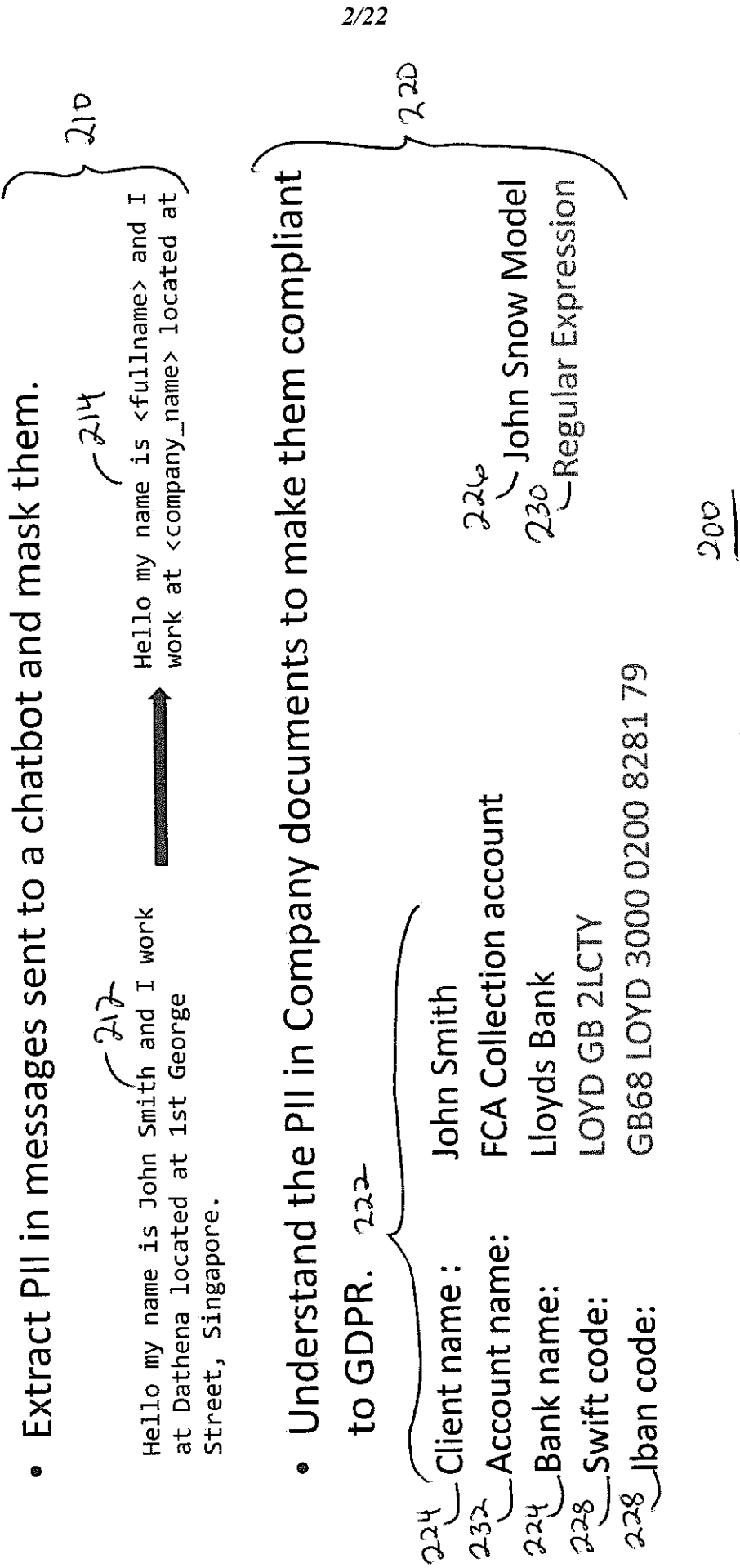
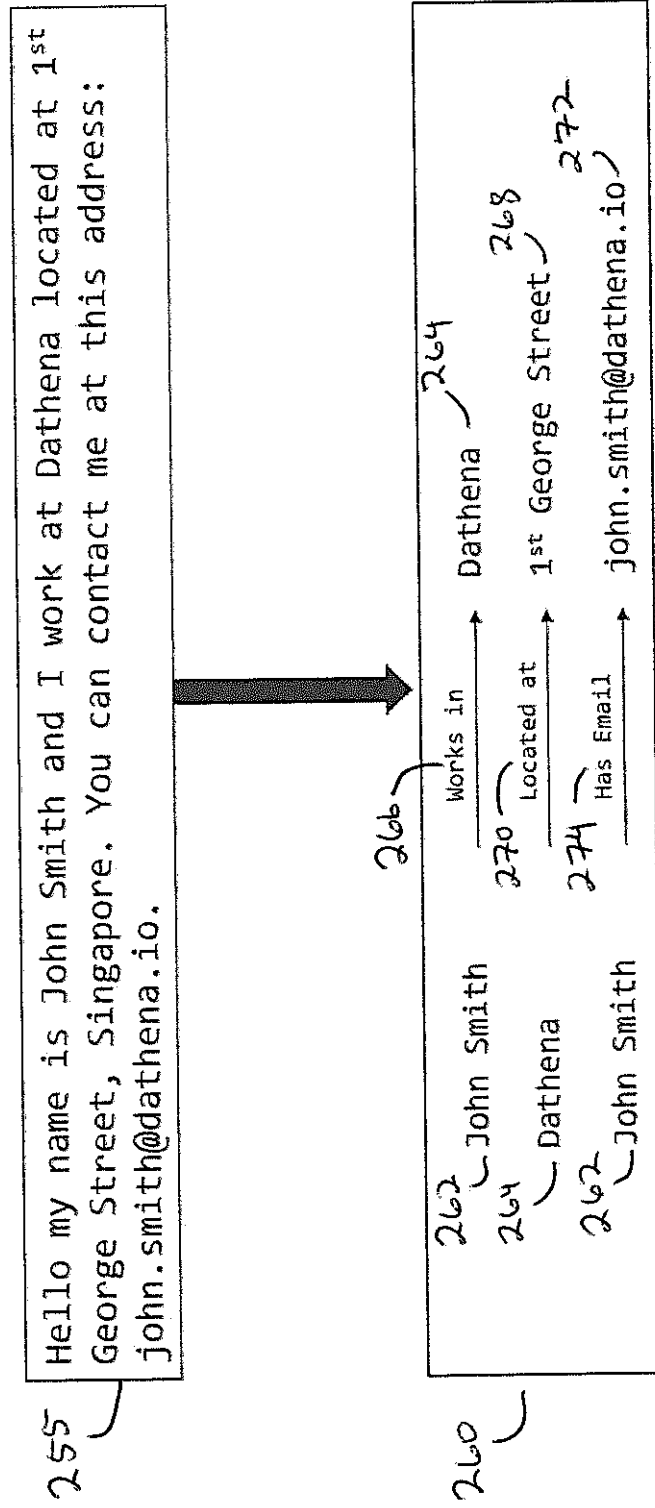


FIG. 1



PII Linking

- **Goal** : For each pair of entities find the relationship that link them.



250

FIG. 2B

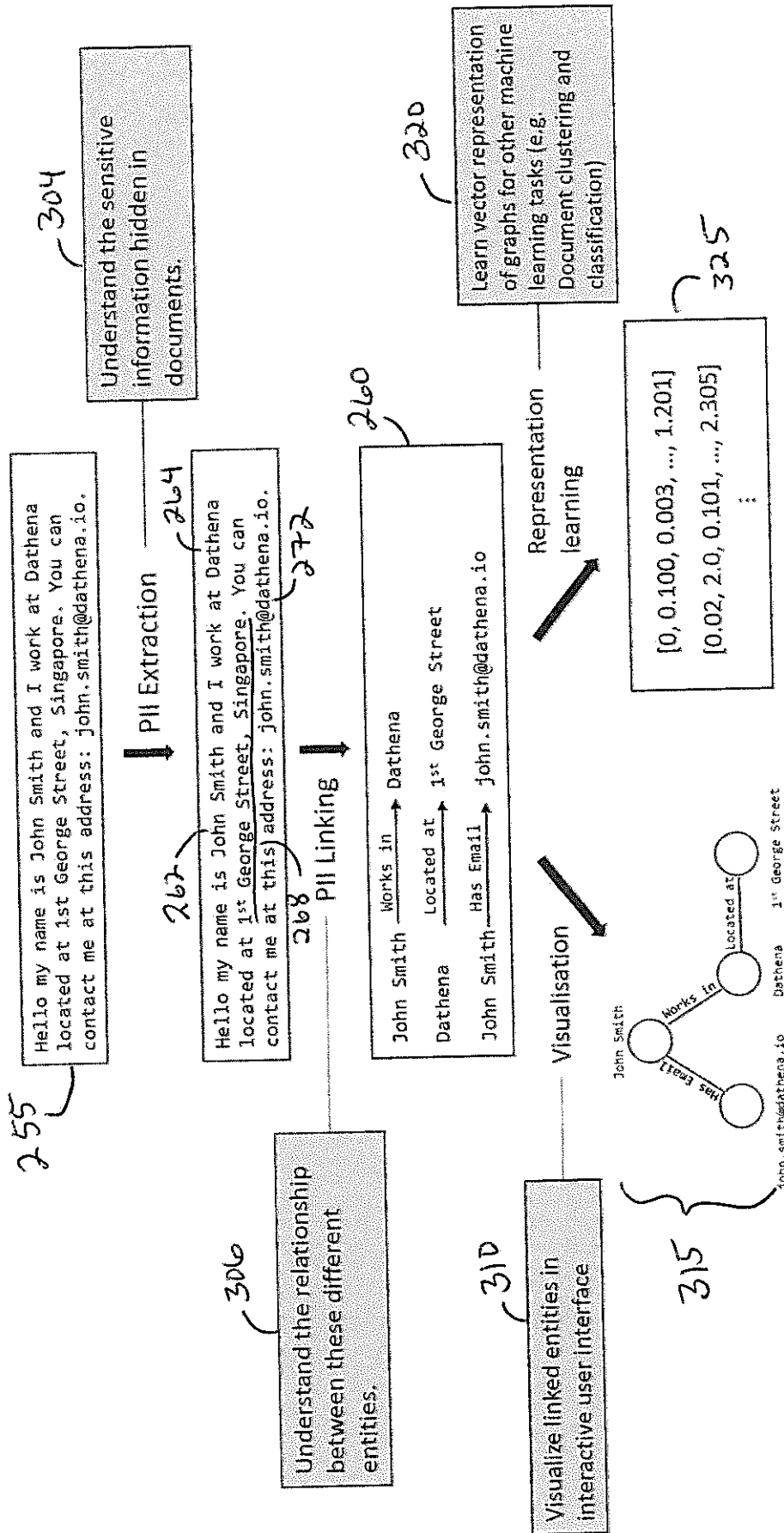


FIG. 3

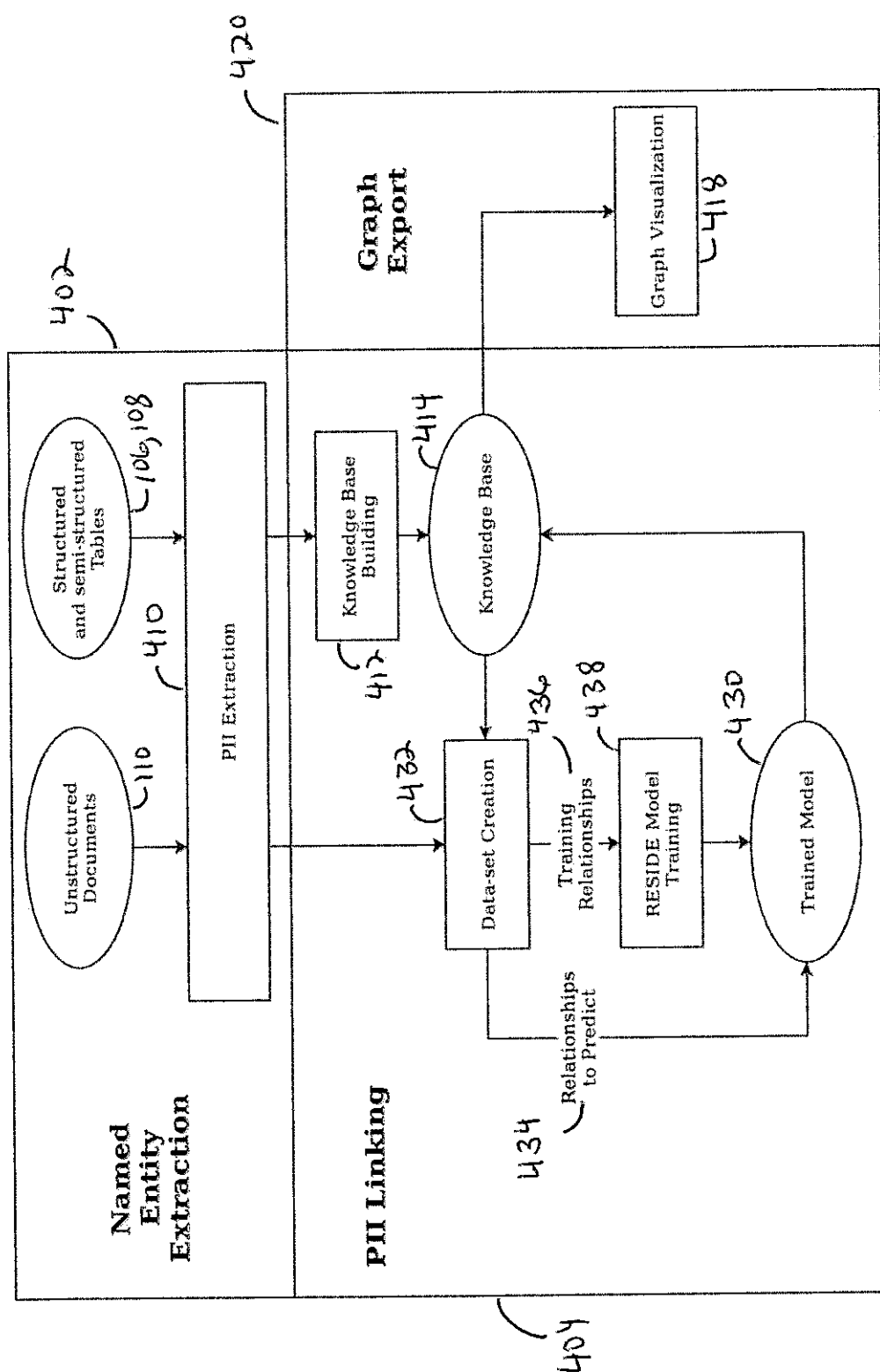


FIG. 4 400

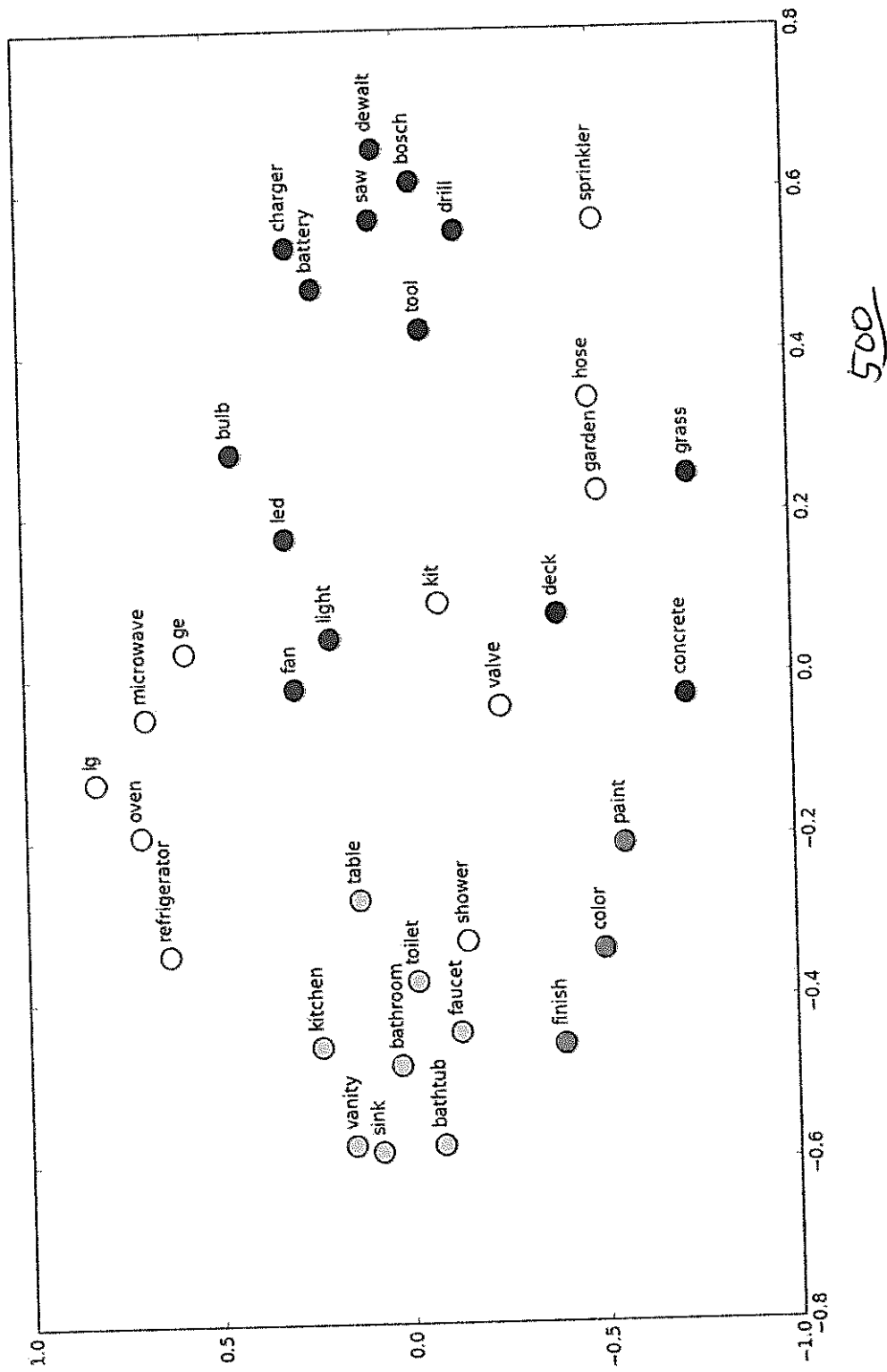


FIG. 5

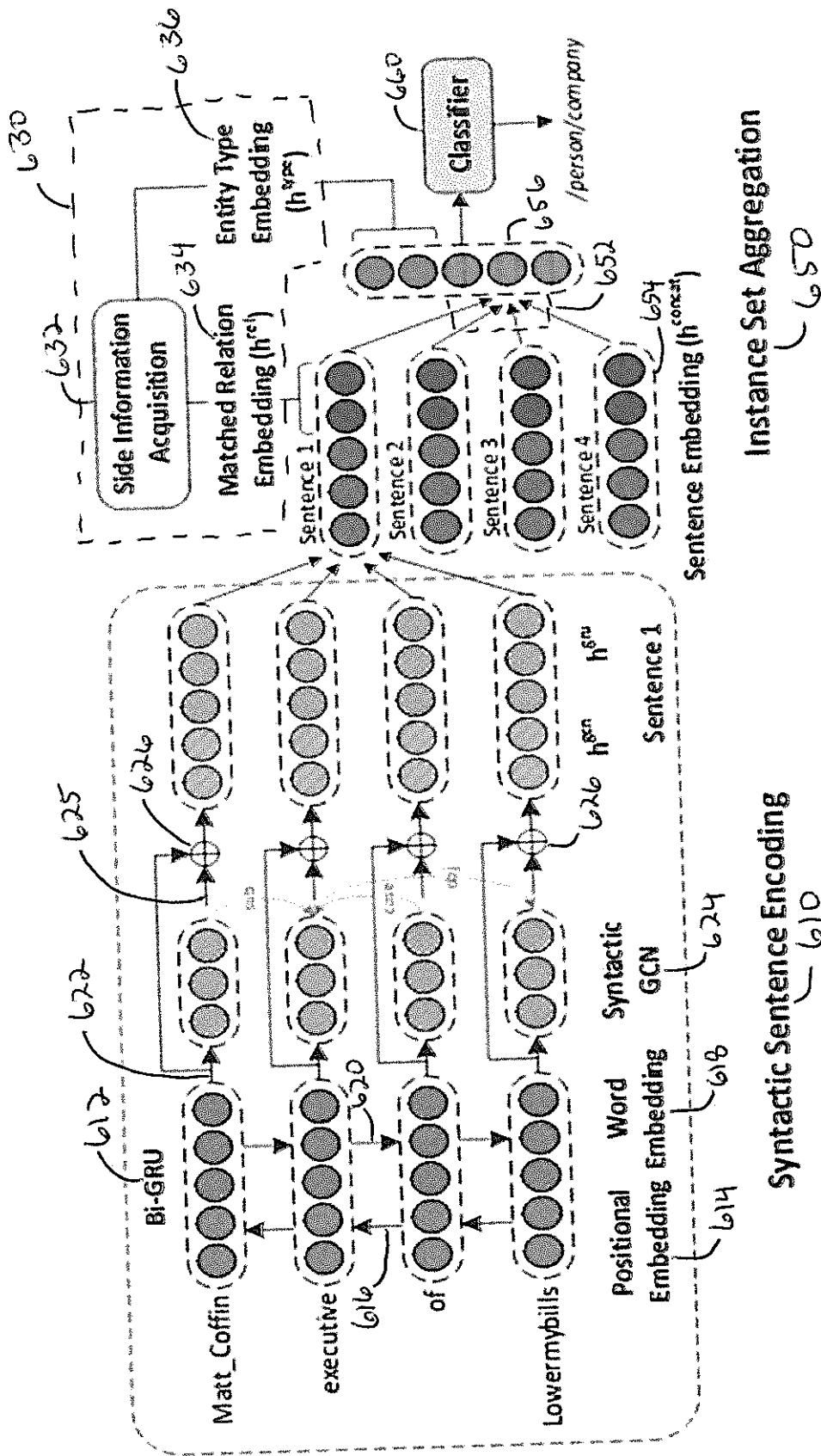


FIG. 6

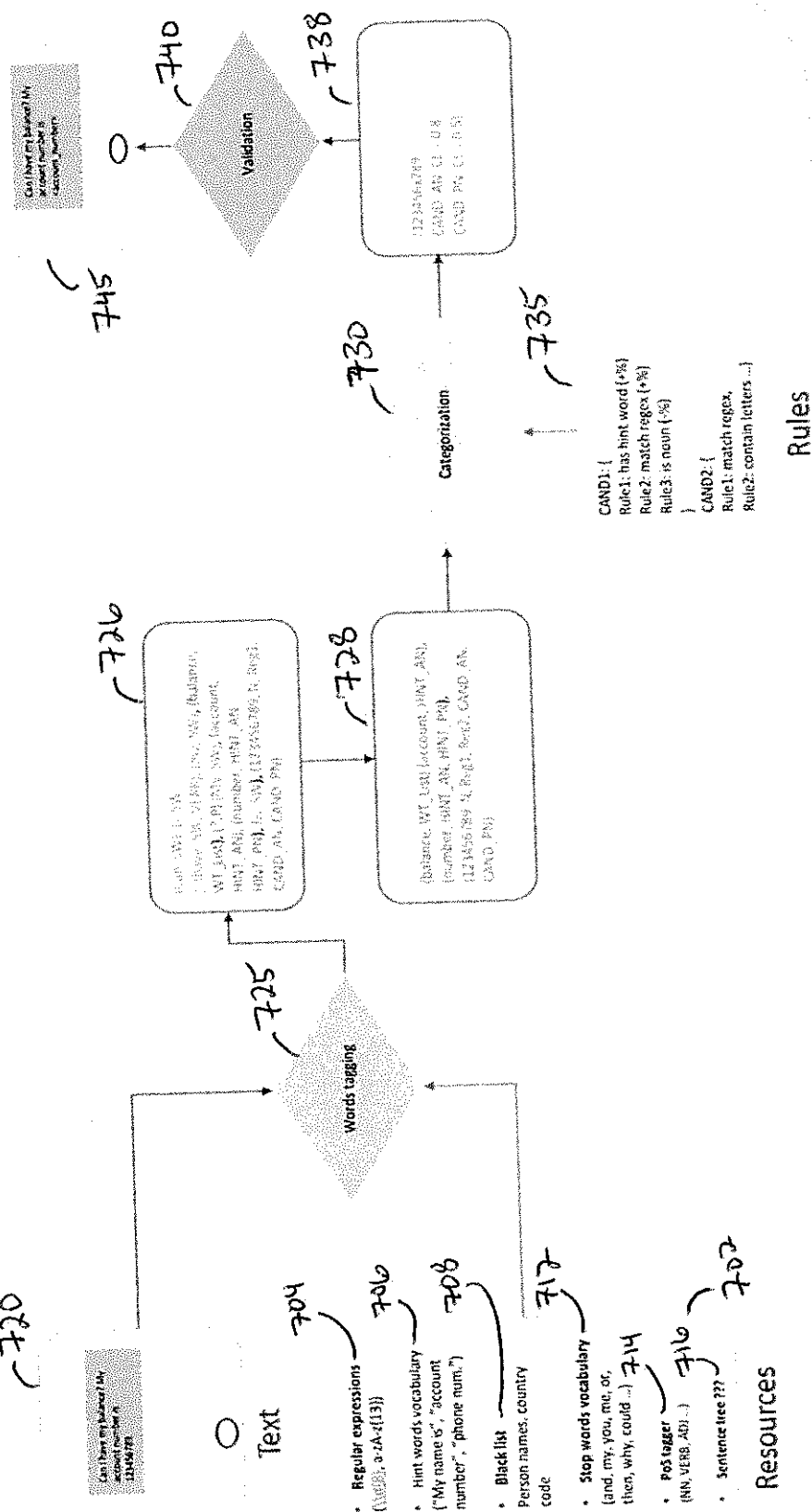


FIG. 7

9/22

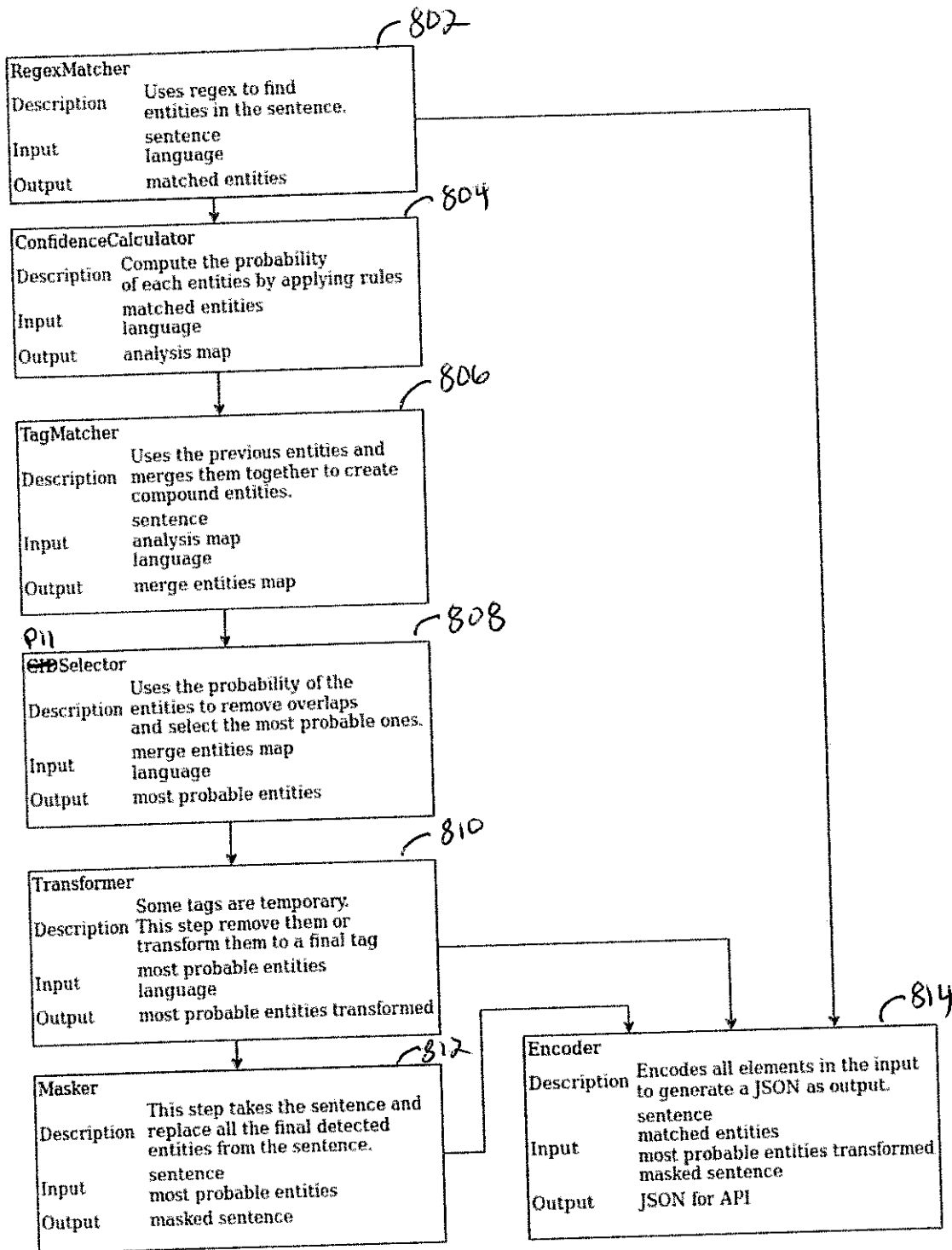


FIG. 8

800

10/22

```
"name": "DATHENA Cash Account",  
"regex": "\\b(?i)(0(\\d){3}[-\\.\\:\\\\; ]*([0-9A-Z]){2}(\\d){6}.([0-9A-Z]){2}[A-Z])\\b"  
  
"name": "DATHENA Portfolio Id",  
"regex": "\\b(?i)(0(\\d){3}[-\\.\\:\\\\; ]*([0-9A-Z]){2}(\\d){6}.(\\d){4})\\b"  
  
"name": "DATHENA Custody Account",  
"regex": "\\b(?i)(0(\\d){3}[-\\.\\:\\\\; ]*([0-9A-Z]){2}(\\d){6}.([A-Z][0-9]|[0-9]))\\b"  
  
"name": "DATHENA Client Relationship",  
"regex": "\\b(?i)(0(\\d){3}[-\\.\\:\\\\; ]*([0-9A-Z]){2}(\\d){6})\\b"
```

FIG. 9

900

```
WordWithIndex(I,Range 0 to 1) -> <proper_noun>  
WordWithIndex(Mr,Range 5 to 7) -> <proper_noun>  
WordWithIndex(Tree,Range 9 to 13) -> <proper_noun>  
WordWithIndex(Mr,Range 5 to 7) -> <uppercase_company_name>  
WordWithIndex(Tree,Range 9 to 13) -> <uppercase_company_name>  
WordWithIndex(Tree,Range 9 to 13) -> <first_name>  
WordWithIndex(Tree,Range 9 to 13) -> <last_name>  
WordWithIndex(Mr,Range 5 to 7) -> <title>  
WordWithIndex(I,Range 0 to 1) -> STOPWORD  
WordWithIndex(am,Range 2 to 4) -> STOPWORD  
WordWithIndex(Mr,Range 5 to 7) -> STOPWORD  
WordWithIndex(I am,Range 0 to 4) -> HINT_<PROPER_NOUN>  
WordWithIndex(I am,Range 0 to 4) -> HINT_<FIRST_NAME>  
WordWithIndex(I am,Range 0 to 4) -> HINT_<LAST_NAME>  
WordWithIndex(I am,Range 0 to 4) -> HINT_<TITLE>  
WordWithIndex(I,Range 0 to 1) -> POS_PRP  
WordWithIndex(am,Range 2 to 4) -> POS_VBP  
WordWithIndex(Mr.,Range 5 to 8) -> POS_NNP  
WordWithIndex(Tree,Range 9 to 13) -> POS_NNP
```

FIG. 10

1000

```

WordWithIndex(I,Range 0 to 1)      -> [<proper_noun>, POS_PRP, STOPWORD]
WordWithIndex(Mr,Range 5 to 7)     -> [<proper_noun>,
<uppercase_company_name>, <title>, STOPWORD]
WordWithIndex(Tree,Range 9 to 13) -> [<proper_noun>,
<uppercase_company_name>, <first_name>, <last_name>, POS_NNP]
WordWithIndex(am,Range 2 to 4)     -> [STOPWORD, POS_VBP]
WordWithIndex(I am,Range 0 to 4)   -> [HINT <PROPER_NOUN>,
HINT <FIRST_NAME>, HINT <LAST_NAME>, HINT <TITLE>]
WordWithIndex(Mr.,Range 5 to 8)    -> [POS_NNP]

```

1100

FIG. 11

```

WordWithIndex(I,Range 0 to 1)      -> [[<proper_noun>,
0.156000000000000003], [STOPWORD, 1.0], [POS_PRP, 1.0]]
WordWithIndex(am,Range 2 to 4)     -> [[STOPWORD, 1.0], [POS_VBP, 1.0]]
WordWithIndex(Tree,Range 8 to 12) -> [[<proper_noun>, 0.8156800000000001],
[<uppercase_company_name>, 0.43999999999999995], [<first_name>, 0.9475],
[<last_name>, 0.9475], [POS_NNP, 1.0]]
WordWithIndex(I am,Range 0 to 4)   -> [[HINT <PROPER_NOUN>, 1.0],
[HINT <FIRST_NAME>, 1.0], [HINT <LAST_NAME>, 1.0], [HINT <TITLE>, 1.0]]
WordWithIndex(Mr,Range 5 to 7)     -> [[<proper_noun>, 0.51088],
[<uppercase_company_name>, 0.132], [<title>, 0.925], [STOPWORD, 1.0],
[POS_NNP, 1.0]]

```

1200

FIG. 12

```

WordWithIndex(I,Range 0 to 1)    -> [[<proper_noun>,
0.156000000000000003], [STOPWORD, 1.0], [POS_PRP, 1.0]]
WordWithIndex(am,Range 2 to 4)   -> [[STOPWORD, 1.0], [POS_VBP, 1.0]]
WordWithIndex(Tree,Range 8 to 12) -> [[<proper_noun>, 0.8156800000000001],
[<uppercase_company_name>, 0.43999999999999995], [<first_name>, 0.9475],
[<last_name>, 0.9475], [POS_NNP, 1.0]]
WordWithIndex(I am,Range 0 to 4) -> [[HINT_<PROPER_NOUN>, 1.0],
[HINT_<FIRST_NAME>, 1.0], [HINT_<LAST_NAME>, 1.0], [HINT_<TITLE>, 1.0]]
WordWithIndex(Mr,Range 5 to 7)   -> [[<proper_noun>, 0.51088],
[<uppercase_company_name>, 0.132], [<title>, 0.925], [STOPWORD, 1.0],
[POS_NNP, 1.0]]

```

1300**FIG. 13A**

```

WordWithIndex(I,Range 0 to 1)    -> [WordWithIndex(am,Range 2 to 4)]
WordWithIndex(am,Range 2 to 4)   -> [WordWithIndex(Mr,Range 5 to 7)]
WordWithIndex(I am,Range 0 to 4) -> [WordWithIndex(Mr,Range 5 to 7)]
WordWithIndex(Mr,Range 5 to 7)   -> [WordWithIndex(Tree,Range 8 to 12)]

```

1330**FIG. 13B**

```

WordWithIndex(Mr Tree,Range 5 to 12) -> [<fullname>, 0.774325]
WordWithIndex(Mr Tree,Range 5 to 12) -> [<fullname>, 0.774325]
WordWithIndex(Mr Tree,Range 5 to 12) -> [<fullname>, 0.7194]
WordWithIndex(Mr Tree,Range 5 to 12) -> [<title_lastname>, 0.8919500000000001]
WordWithIndex(Mr Tree,Range 5 to 12) -> [<title_lastname>, 0.946875]
WordWithIndex(Mr Tree,Range 5 to 12) -> [<title_lastname>, 0.946875]
WordWithIndex(Mr Tree,Range 5 to 12) -> [<company_name>, 0.40499999999999997]

```

1360**FIG. 13C**


```

WordWithIndex(Davide Smith,Range 11 to 23) -> [<fullname>, 0.9125]
WordWithIndex(6741 4587,Range 47 to 56) -> [<phone_no>, 0.7]
WordWithIndex(Smith,Range 18 to 23) -> [<first_name>, 0.895]
WordWithIndex(Davide,Range 11 to 17) -> [<first_name>, 0.895]
WordWithIndex(My,Range 0 to 2) -> [<proper_noun>, 0.15600000000000003]

```

FIG. 14B1420

```

WordWithIndex(Davide Smith,Range 11 to 23) -> [<fullname>, 0.9125]
WordWithIndex(6741 4587,Range 47 to 56) -> [<phone_no>, 0.7]
WordWithIndex(Smith,Range 18 to 23) -> [<first_name>, 0.895]
WordWithIndex(Davide,Range 11 to 17) -> [<first_name>, 0.895]

```

FIG. 14C1440

```

WordWithIndex(Davide Smith,Range 11 to 23) -> 0.9542483660130718
WordWithIndex(6741 4587,Range 47 to 56) -> 0.7329842931937172
WordWithIndex(Smith,Range 18 to 23) -> 0.6090161610433796
WordWithIndex(Davide,Range 11 to 17) -> 0.6723906627314193

```

FIG. 14D1460

```

WordWithIndex(6741 4587,Range 47 to 56) -> [<phone_no>, 0.7]
WordWithIndex(Davide Smith,Range 11 to 23) -> [<fullname>, 0.9125]

```

FIG. 14E1480

```

Input: Hello my name is John Smith and I work at Dathena located at 1st
George Street, Singapore.
Output: Hello my name is <fullname> and I work at <company_name> located
at <address>.

```

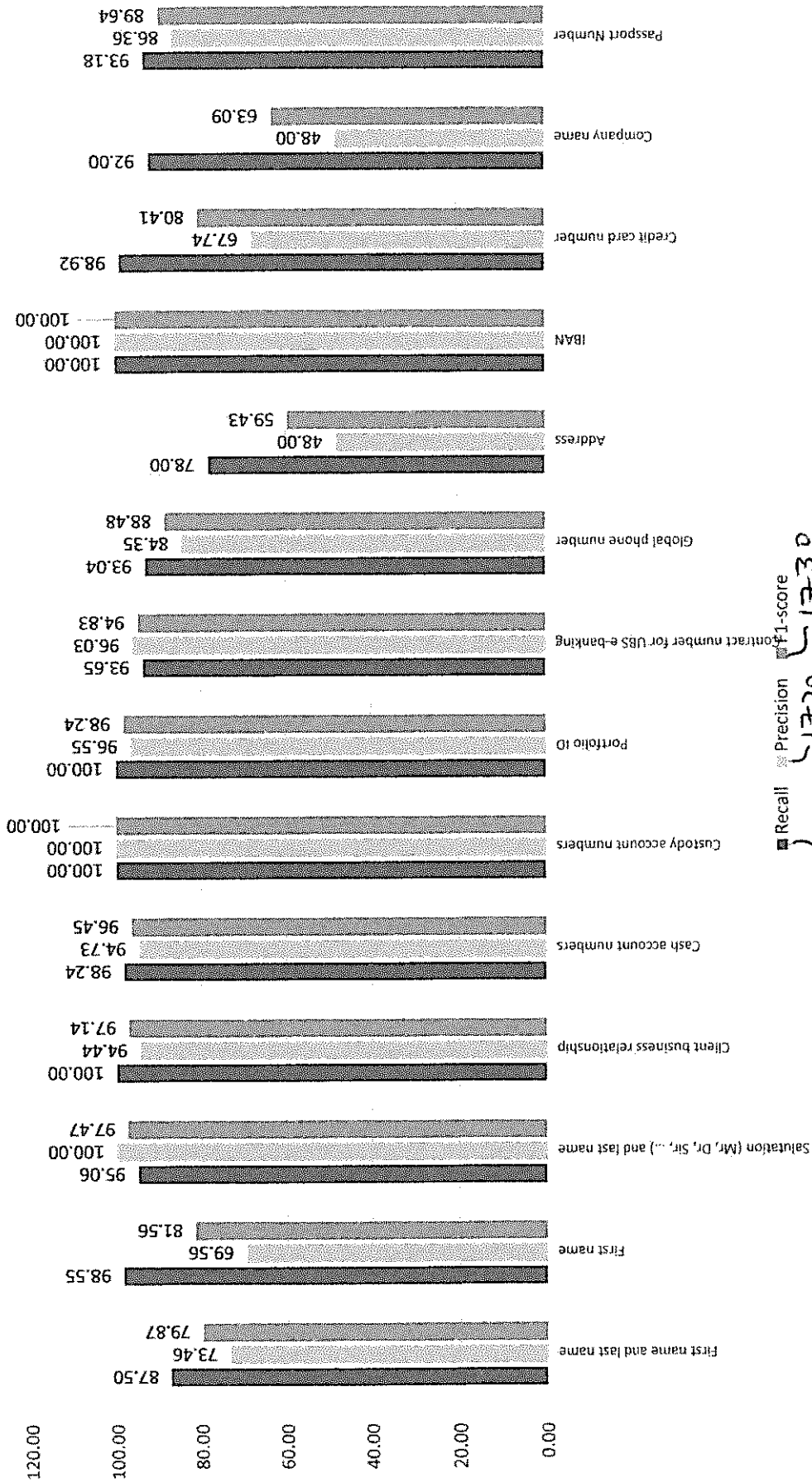
FIG. 151500

1502

1504

	1606 ↓ p	1610 actual value	1608 ↓ n	total
1602 → p'	True Positive 1620	False Positive 1640		P'
1612 prediction outcome				
1604 → n'	False Negative 1650	True Negative 1630		N'
total	P	N		
			<u>1600</u>	

FIG. 16



Recall Precision F1-score

FIG. 17

1700

17/22

EVALUATION - GERMAN

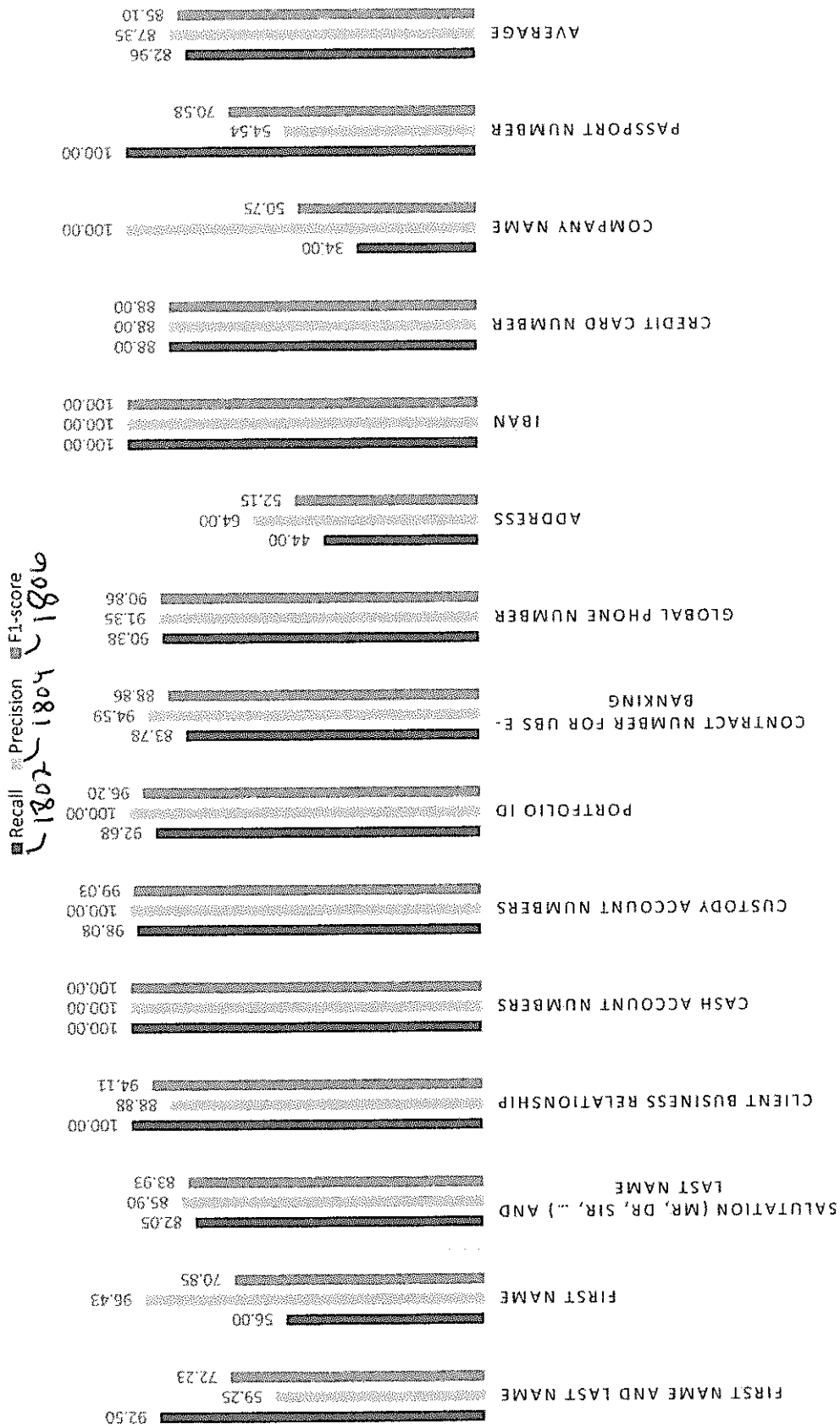
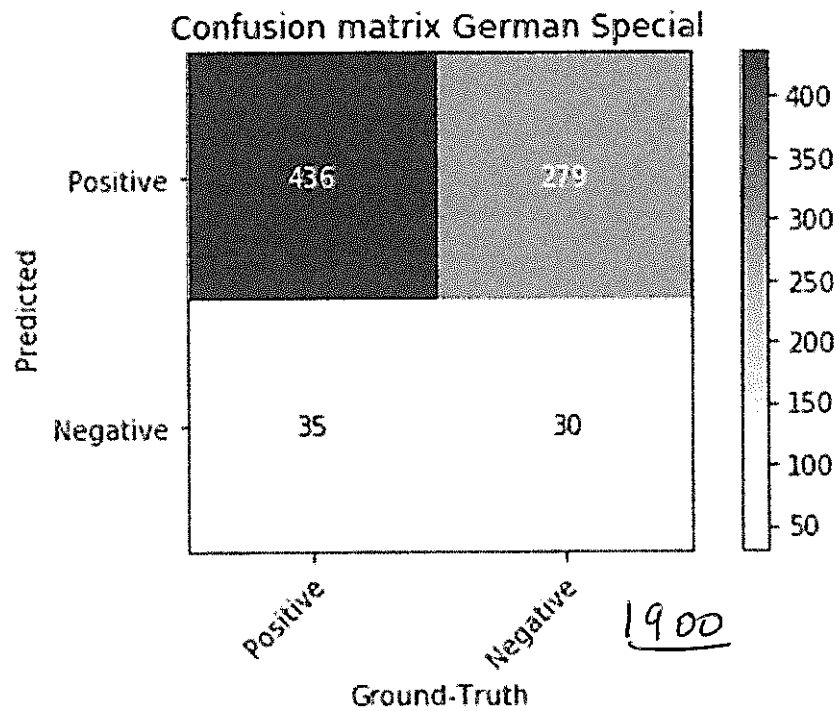
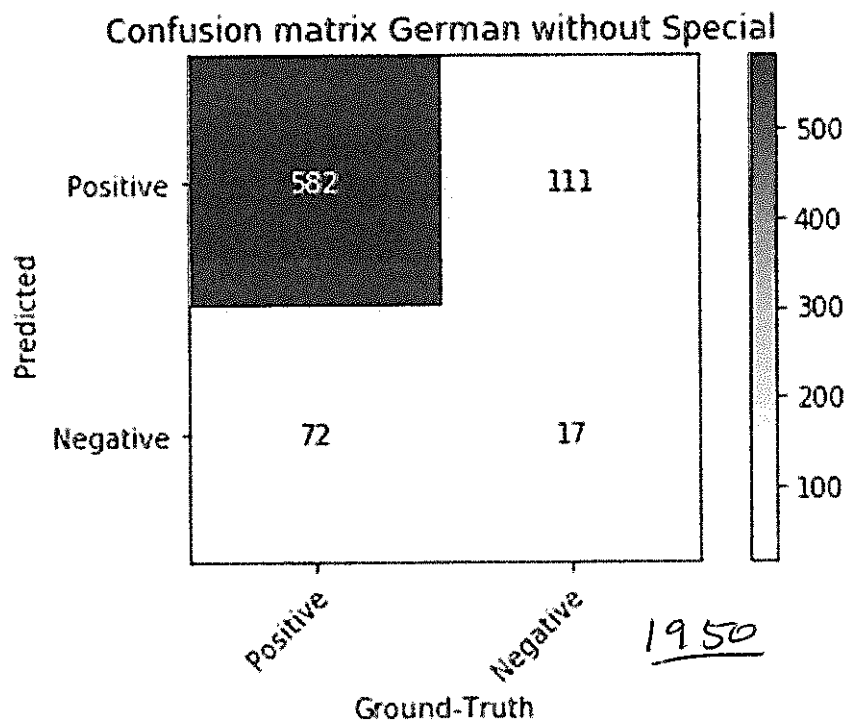
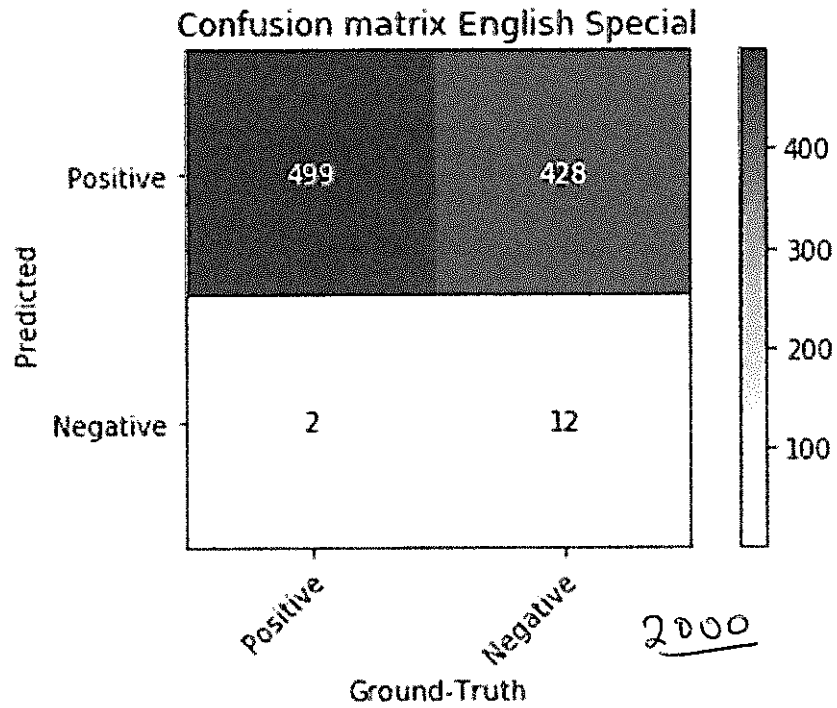
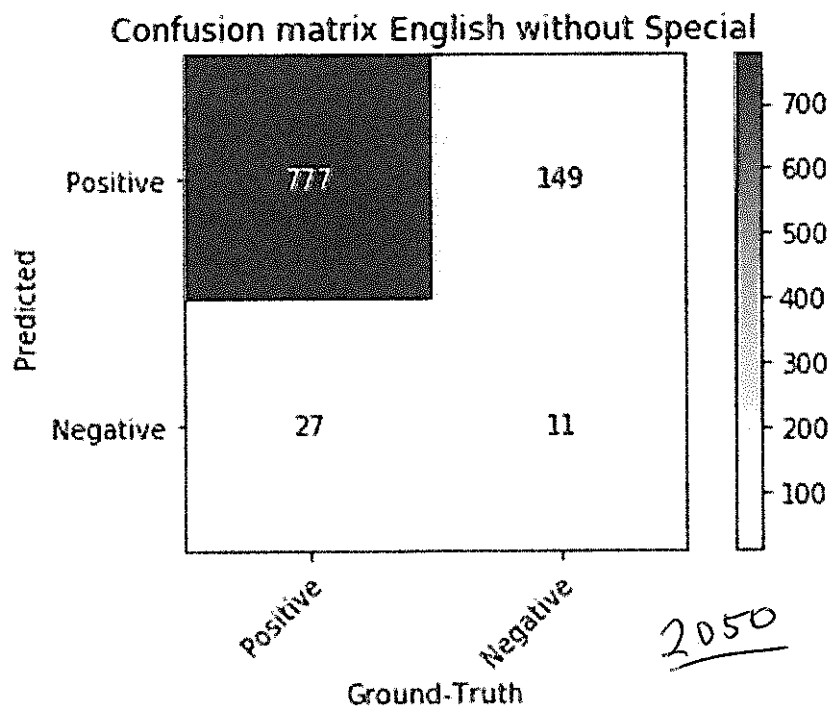
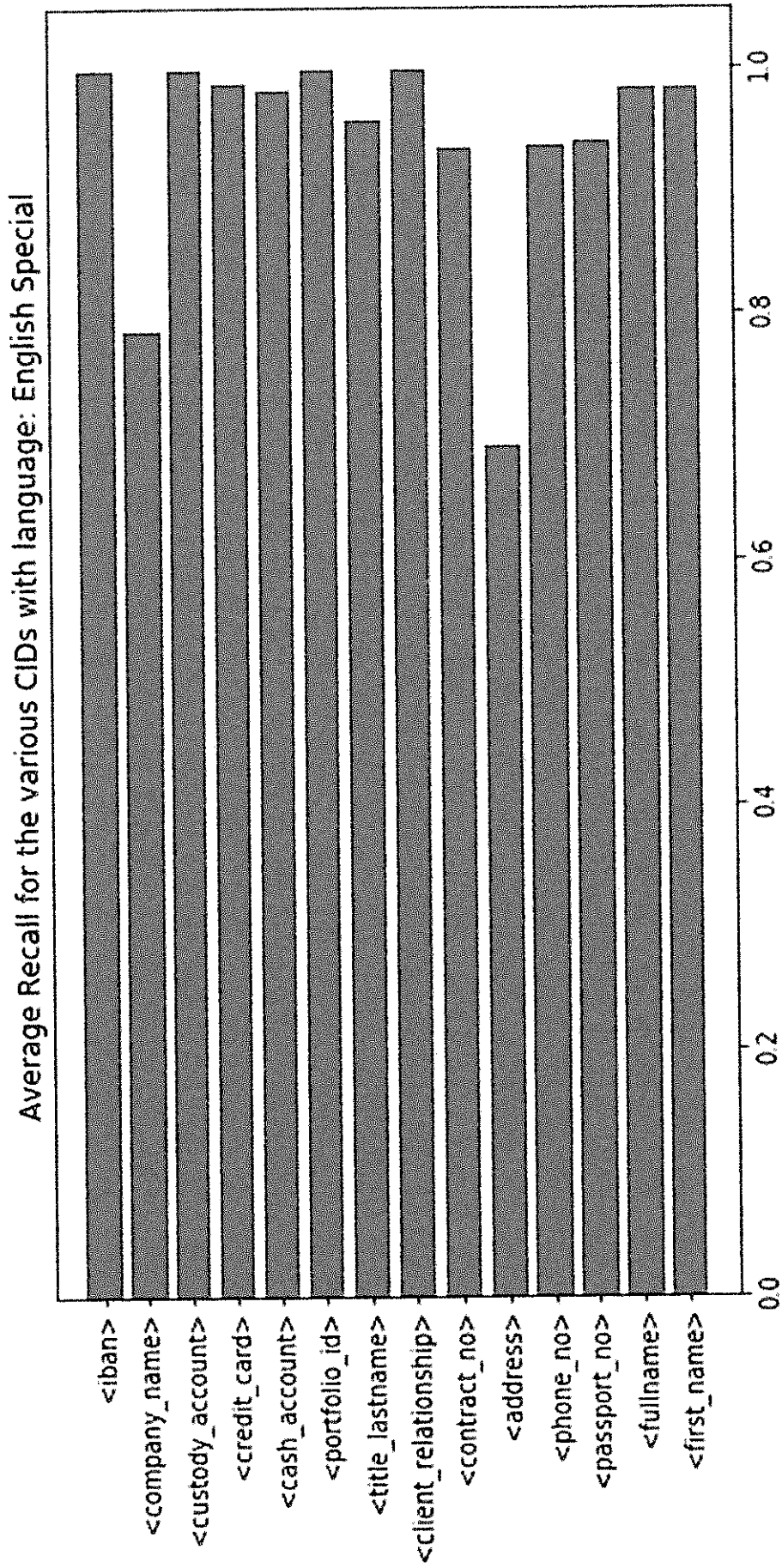


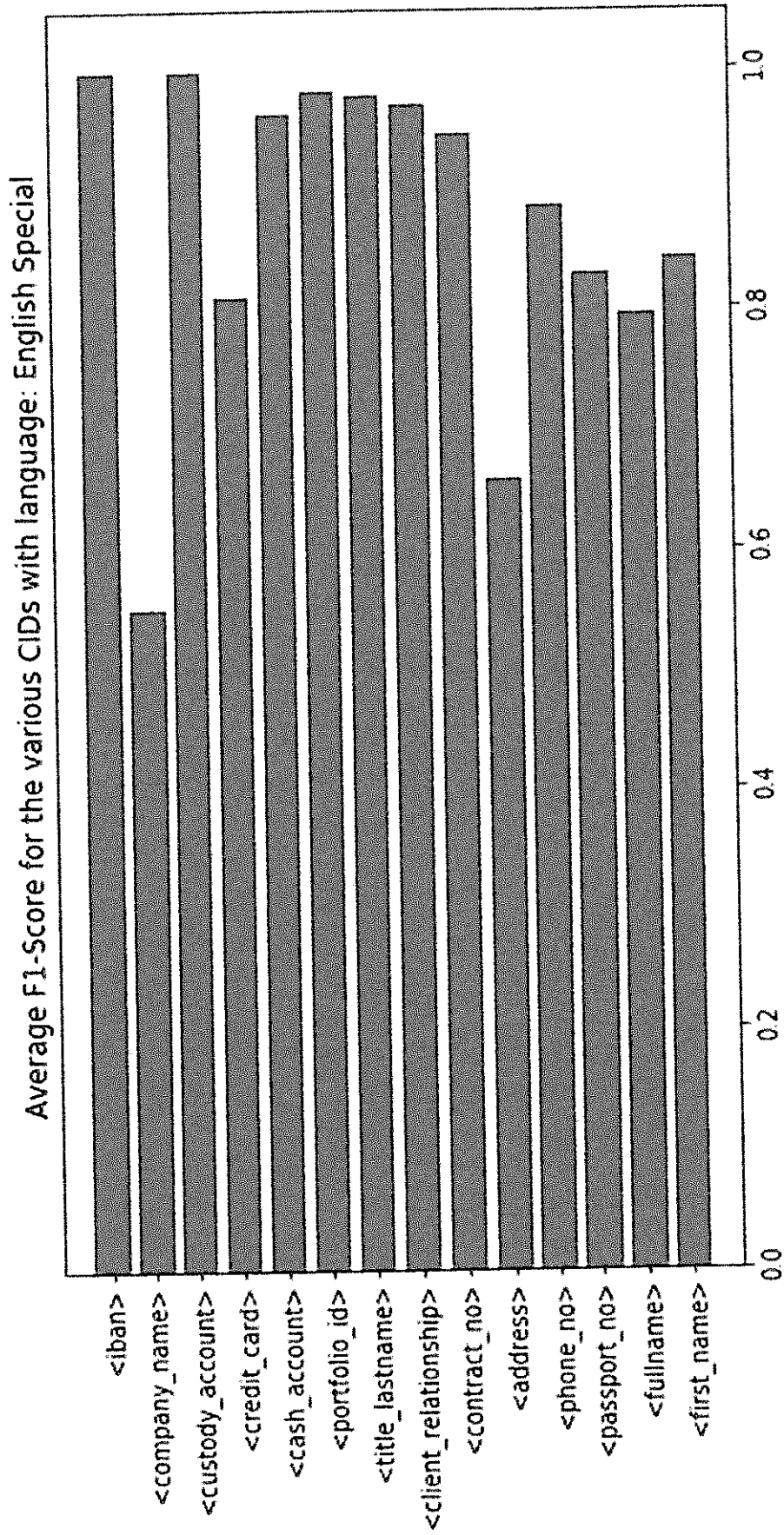
FIG. 18

1800

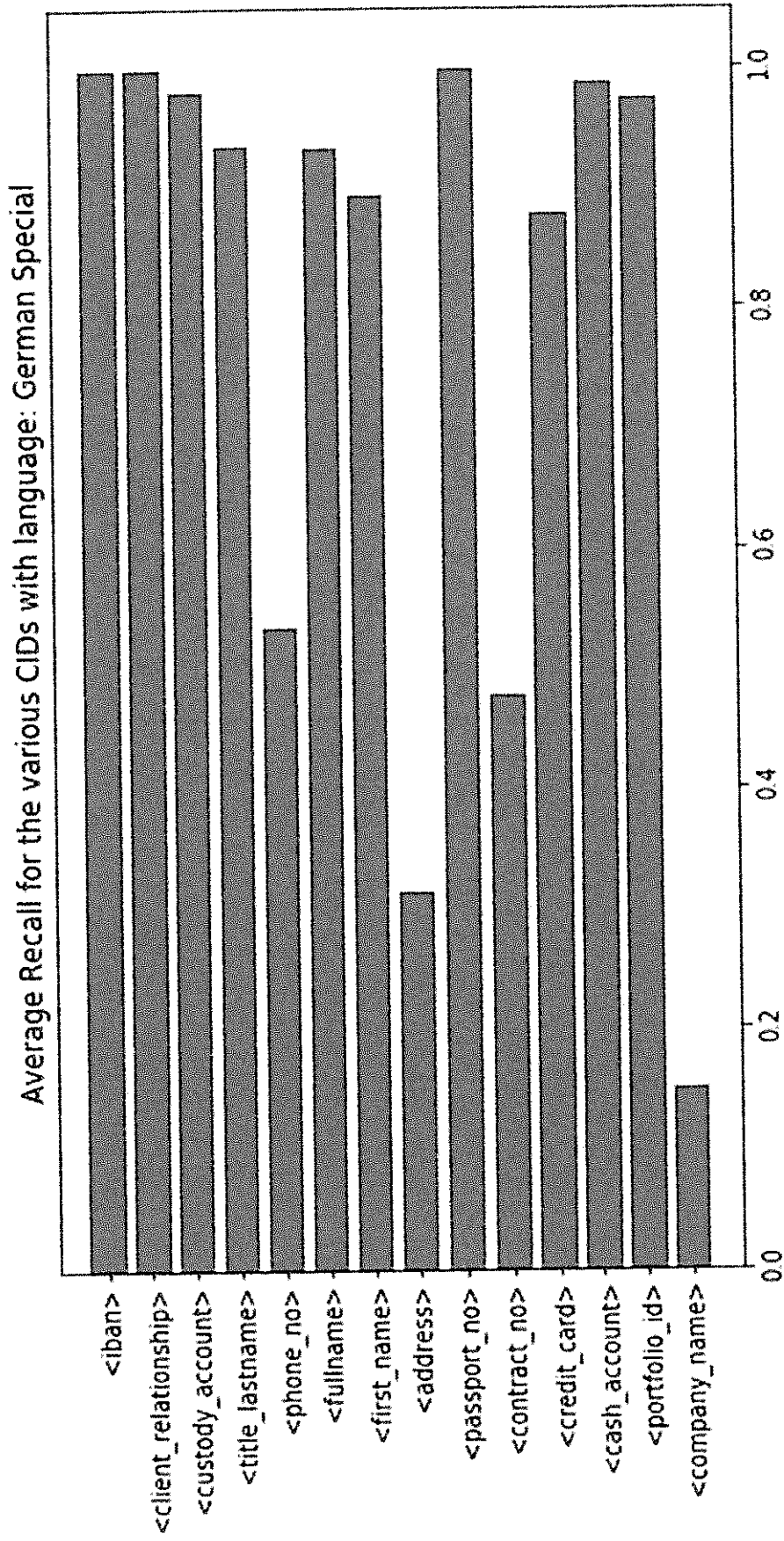
**FIG. 19A****FIG. 19B**

**FIG. 20A****FIG. 20B**

2100**FIG. 21A**

**FIG. 21B**

2130

2160**FIG. 21C**