

Extração de patentes com o atributo ISSN

```
SELECT DISTINCT tls201_appln.appln_id,  
                tls214_npl_publn.npl_biblio,  
                tls209_appln_ipc.ipc_class_symbol,  
                tls203_appln_abstr.appln_abstract,  
                tls201_appln.appln_auth,  
                tls201_appln.appln_nr_original,  
                tls203_appln_abstr.appln_abstract_lg  
FROM tls201_appln  
JOIN tls203_appln_abstr ON tls201_appln.appln_id =  
tls203_appln_abstr.appln_id  
JOIN tls209_appln_ipc ON tls201_appln.appln_id =  
tls209_appln_ipc.appln_id  
JOIN tls211_pat_publn ON tls211_pat_publn.appln_id =  
tls201_appln.appln_id  
JOIN tls212_citation ON tls211_pat_publn.pat_publn_id =  
tls212_citation.pat_publn_id  
JOIN tls214_npl_publn ON tls212_citation.cited_npl_publn_id =  
tls214_npl_publn.npl_publn_id  
WHERE LEN(tls212_citation.cited_npl_publn_id)>1  
      AND lower(tls209_appln_ipc.ipc_class_symbol) LIKE 'b22d%'  
      AND (tls214_npl_publn.npl_biblio LIKE '%ISSN:%')  
      AND tls203_appln_abstr.appln_abstract_lg = 'en'  
ORDER BY tls201_appln.appln_id  
OFFSET 10000 ROWS FETCH NEXT 10000 ROWS ONLY
```

Repare na cláusula WHERE da consulta SQL que filtramos o código IPC por quatro dígitos por meio da condição `lower(tls209_appln_ipc.ipc_class_symbol) LIKE 'b22d%'`. Realizamos um pull de queries limitando o número de patentes de retorno em 10000 linhas para contornar o limite de download permitido pelo repositório. Em cada uma das queries variamos os códigos ipcs de modo que todas as combinações possíveis de códigos ipcs fossem abrangidas.

Também impusemos uma restrição no perfil das NPRs extraídas. A query busca por referências que contenham uma subcadeia de caracteres ISSN no corpo. De modo a facilitar e aumentar a precisão do algoritmo de classificação de área científica por meio de NPRs.

Extração de patentes Sul Americanas para modelo de inferência

Um conjunto de dados de patentes Sul Americanas foi escolhido para a aplicação de nosso modelo de inferência. A query abaixo pode ser interpretada como extraia a aplicação (`appln_id`), o abstract (`appln_abstract`), os códigos ipc (`ipc_class_symbol`) e todos os aplicantes (`person_name`) de uma patente em que pelo menos um dos aplicantes é Sul americano (`person_ctype_code`).

```
SELECT DISTINCT tls201_appln.appln_id,
                tls209_appln_ipc.ipc_class_symbol,
                tls203_appln_abstr.appln_abstract,
                tls207_pers_appln.person_id,
                tls206_person.person_name,
                tls206_person.person_ctype_code,
                tls201_appln.receiving_office
FROM tls201_appln
JOIN tls203_appln_abstr ON tls201_appln.appln_id =
tls203_appln_abstr.appln_id
JOIN tls209_appln_ipc ON tls201_appln.appln_id =
tls209_appln_ipc.appln_id
JOIN tls207_pers_appln ON tls201_appln.appln_id =
tls207_pers_appln.appln_id
JOIN tls206_person ON tls207_pers_appln.person_id =
tls206_person.person_id
WHERE tls201_appln.appln_id IN
  (SELECT DISTINCT tls201_appln.appln_id
   FROM tls201_appln
   JOIN tls203_appln_abstr ON tls201_appln.appln_id =
tls203_appln_abstr.appln_id
   JOIN tls207_pers_appln ON tls201_appln.appln_id =
tls207_pers_appln.appln_id
   JOIN tls206_person ON tls207_pers_appln.person_id =
tls206_person.person_id
   WHERE tls203_appln_abstr.appln_abstract_lg = 'en'
   AND (  tls206_person.person_ctype_code = 'BR'
        OR tls206_person.person_ctype_code = 'AR'
        OR tls206_person.person_ctype_code = 'VE'
        OR tls206_person.person_ctype_code = 'CO'
        OR tls206_person.person_ctype_code = 'PE'
        OR tls206_person.person_ctype_code = 'BO'
        OR tls206_person.person_ctype_code = 'PY'
        OR tls206_person.person_ctype_code = 'UY'
        OR tls206_person.person_ctype_code = 'EC')
```

```
OR tls206_person.person_ctry_code = 'GY'  
OR tls206_person.person_ctry_code = 'SR'  
OR tls206_person.person_ctry_code = 'GF'))
```