



Importancia del acceso a datos en la ciencia

Erika Alejandra Peláez Coyotl
ALA-ICA 2017



De acuerdo con nuestros colegas canadienses...

50,000,000 artículos publicados hasta 2009

cada año en promedio se agregan **2,500,000**



Existen herramientas para hojear publicaciones



Nexus S, a pure Google experience. Get it FREE today only



Pero imaginemos

X Condiciones X Metabolito + Unificar mediciones

2958

The screenshot shows a PubMed search interface. At the top, the search bar contains the text "metabolites metabolyc syndrome" and a "Search" button. Below the search bar, there are options for "Create RSS", "Create alert", and "Advanced". The search results are displayed in a list format, with "Format: Summary", "Sort by: Most Recent", and "Per page: 20" visible. The search results section shows "Search results" and "Items: 1 to 20 of 2958". A red box highlights the number "2958". Below the search results, there is a "Showing results for" message and a list of search results. The first result is "Fatty Acids in Nephrotic Syndrome and Chronic Kidney Disease" by Turolo S, Edefonti A, Syren ML, Marangoni F, Morello W, Agostoni C, Montini G. The second result is "Higher intake of fish and fat is associated with lower plasma s-adenosylhomocysteine: a cross-sectional study" by Lind MV, Lauritzen L, Pedersen O, Vestergaard H, Stark KD, Hansen T, Ross AB, Kristensen M. The third result is "Follicular metabolic changes and effects on oocyte quality in polycystic ovary syndrome patients" by Zhang Y, Liu J, Yin T, Yang J, Xiong C. On the right side of the page, there are sections for "Results by year" (a bar chart), "Find related data" (a dropdown menu), and "Search details" (a text box containing the search query: ("Metabolites"[Journal] OR "metabolites"[All Fields]) AND ("metabolism"[Subheading] OR "metabolism"[All Fields] OR "metabolism"[MeSH Terms] OR



Para esto utilizamos un enfoque programático



Bases de datos

Name	Organism	Factor	Heatmap	Box plot	Variance	Anova F score
Mining for metabolic responses to long-term salt stress: a case study on the model legume Lotus japonicus (C)	Lotus	Salt			727.45	148.99
Metabolomic responses to long-term salt stress in related Lotus species (A)	Lotus	Germplasm, Salt			496.89	44.10
Metabolomic responses to long-term salt stress in related Lotus species (C)	Lotus	Germplasm, Salt			547.49	34.23
Mining for metabolic responses to long-term salt stress: a case study on the model legume Lotus japonicus (A)	Lotus	Salt			284.76	32.07
Metabolomic responses to long-term salt stress in related Lotus species (B)	Lotus	Germplasm, Salt			291.67	15.23
Mining for metabolic responses to long-term salt stress: a case study on the model legume Lotus japonicus (B)	Lotus	Salt			262.27	2.14
6 quantitative metabolite profile(s)						

API (Application Programming Interface)

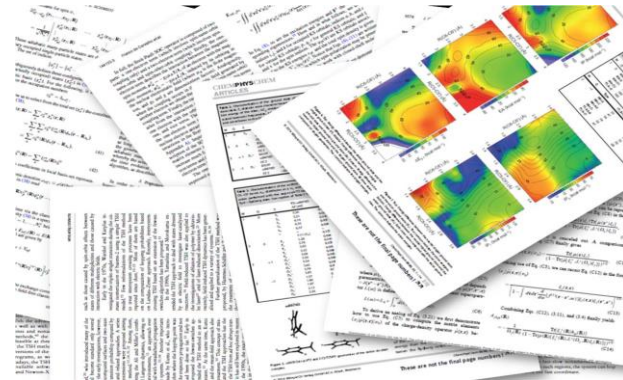
JSON

```
parse_response({
  "code": "/api/status/ok",
  "result": [
    {
      "id": "/topic/en/edinburgh",
      "key": [
        {
          "namespace": "/wikipedia/en_id",
          "value": "9602"
        }
      ]
    }
  ],
  "status": "200 OK",
  "transaction_id":
  "cache;cache02.p01.sjc1:8101;2011-10-07T15:51:49Z;0019"
})
```

```
▼<ArrayOfBlog xmlns:i="http://www.w3.org/2001/XMLSchema-instance"
  ▼<Blog>
    <Id>1</Id>
    <Name>C-SharpCorner</Name>
    <Url>http://www.c-sharpcorner.com/</Url>
  </Blog>
  ▼<Blog>
    <Id>2</Id>
    <Name>CodeProject</Name>
    <Url>http://www.codeproject.com/</Url>
  </Blog>
  ▼<Blog>
    <Id>3</Id>
    <Name>StackOverflow</Name>
    <Url>http://stackoverflow.com/</Url>
  </Blog>
</ArrayOfBlog>
```




Programa





Un ejemplo en reposicionamiento de fármacos



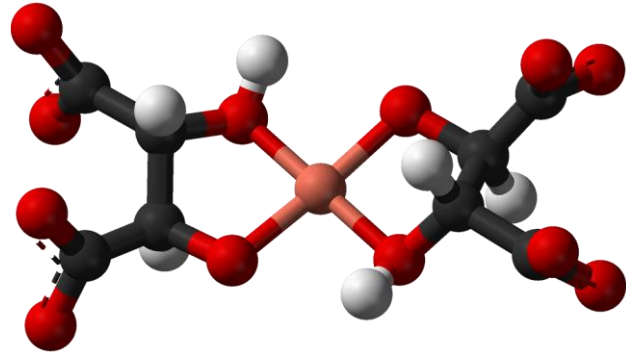
Human alpha defensin 1



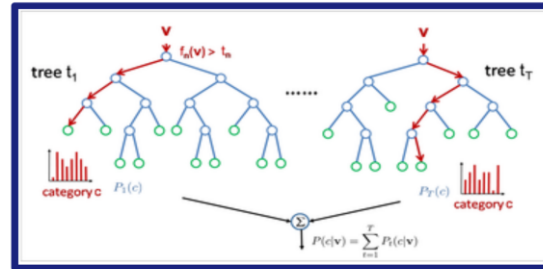
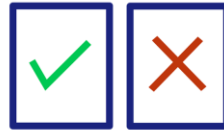
Human beta defensin 1



Human cathelicidin LL-37



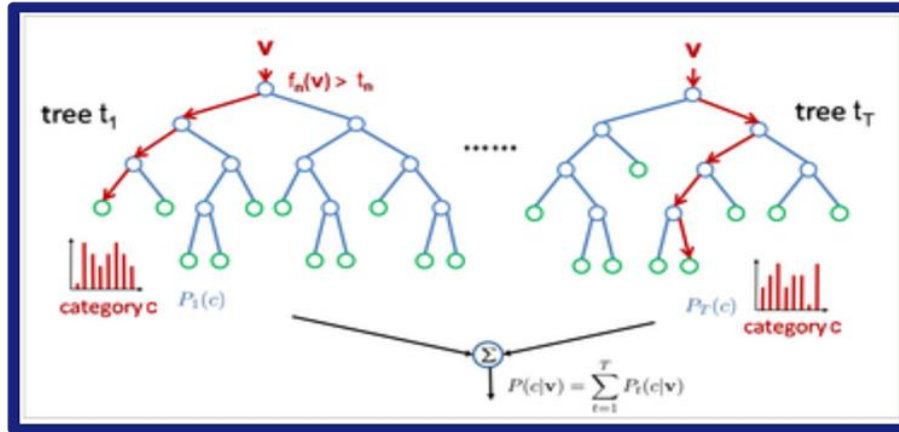
Generamos un modelo con Machine Learning



Antimicrobiano

No
Antimicrobiano

Utilizamos fármacos aprobados sobre el modelo



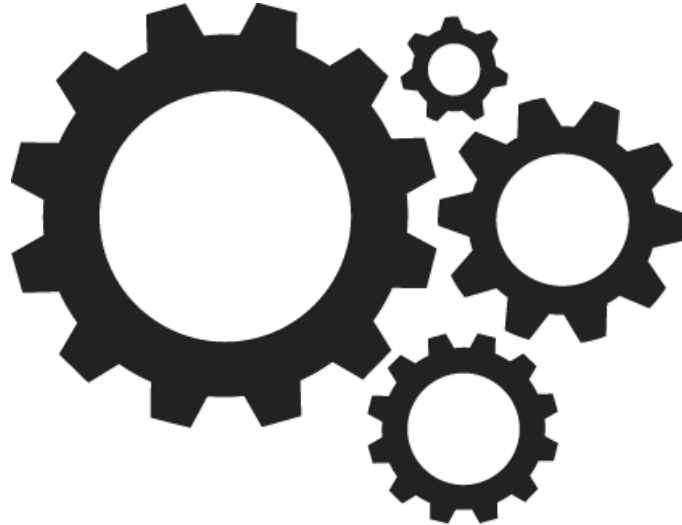
Antimicrobiano

No Antimicrobiano

**Validamos las predicciones
experimentalmente**



Para que esto funcione el proceso debe estar bien coordinado





La creación de fuentes de datos es vital

1. Crear recursos para el almacenamiento y la consulta de datos
2. Procesamiento de archivos existentes
3. Mejoramiento a la accesibilidad de los archivos (con miras al acceso programático)
4. Siempre tener en mente la escalabilidad



Muchas gracias por su atención