

ElasticSearch

A comrade in data exploration and analysis

What is it?

ElasticSearch is an open source search engine with powerful data analysis capabilities.

- Fast, almost automatic, cluster setup and querying
- Aggregations framework
- Less worries to start (mapping, missing data, normalisation)
- Not for data storage

Use Case

Open Gov API Exploration

Data Source

Beginning in October 1st 2010, all Greek government institutions are obliged to upload their acts and decisions on the Internet through a public API.

<https://diavgeia.gov.gr/en>

Use Case

Open Gov API Exploration

Experiment

- Scrape data using Python
 - Insert into ElasticSearch
 - Use Browser tool to see contents
 - Build up mapping
 - Use aggregations to find insights
-

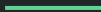
Average Expense: €29K

Max. Expense Decision: €900M

Coverage: €3B

Example 1

- Use “stats” aggregation to get basic stats about public expenses for 2016



Use Case

Open Gov API Exploration

Highway Construction S.A	262.189.980,00 €
Personal Insurance S.A	97.329.211,00 €
Sygiridis Car Parts G.P.*	92.880.863,00 €
National Deposits and Loans Fund	80.738.537,00 €
Manpower Employment Organization	75.067.723,00 €
Social Security Foundation	62.892.375,00 €
Information Society S.A	59.904.085,00 €
Ministry of Finance	42.358.165,00 €
Public Power Corporation S.A.	30.957.795,00 €
Pallas Stergios Georgios*	27.220.499,00 €

***amounts do not match published decisions**

Example 2

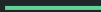
- Use “terms aggregation” to group decisions by grantee
- Sum expense amounts per bucket
- Get top hits per bucket to confirm the results. (“top_hits” aggregation)

Use Case

Open Gov API Exploration

Initial Conclusions

- Low Coverage
- Discrepancies
- Inexhaustive Expense Explanation



Thank you!

source code: <https://github.com/stathischaritos/diavgeia>