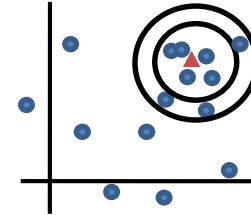


GOCat : our Machine Learning system



GOCat,
the Gene Ontology Categorizer
<http://eagl.unige.ch/GOCat>

$k = 100$



k-NN

Retrieves the k
most similar PMIDs
curated in GOA

Query

An abstract / piece
of free text

e.g. : «And similar
suppression was found
in emb12 and emb14
mutants which were
impaired in plastid
translation process...»

GOA

PMID
 P_1

GO:0005634
GO:0005886

PMID
 P_2

GO:0005634
GO:0005737

PMID
 P_3

GO:0005737

Output

Most prevalent
GO codes

#	GO Terms
1	GO0005634 Nucleus
2	GO0005737 Cytoplasm
3	GO0005886 Plasma membrane