# Evaluation summary

## Code fragments identification evaluation

For evaluating the performance of our approach, we compare the report automatically obtained by our implementation with the ground truth. For each project, we compute the precision and the recall depending on the minimum score threshold set.

This chart represents the trends in the evolution of precision and recall as a function of score thresholds.

Une image contenant texte, ligne, diagramme, Tracé

Description générée automatiquement

This chart represents raw data behind the previous chart.

Une image contenant texte, ligne, diagramme, Tracé

Description générée automatiquement

We can deduce, for each project, the optimal score threshold based on the best balance between precision and recall.

Une image contenant texte, capture d’écran, Caractère coloré, Police

Description générée automatiquement

The charts above are created based on the following table.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **API Express** |  | **Scores** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** |
|  | **comments-api** | **Precision** | 54,55% | 66,67% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
|  |  | **Recall** | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
|  | **cinema-microservice** | **Precision** | 73,68% | 93,33% | 100,00% | 100,00% | 100,00% | N/A | N/A | N/A |
|  |  | **Recall** | 100,00% | 100,00% | 100,00% | 85,71% | 85,71% | 0,00% | 0,00% | 0,00% |
|  | **overleaf** | **Precision** | 69,33% | 81,25% | 92,86% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
|  |  | **Recall** | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 92,31% | 87,82% | 84,62% |
|  | **cloudboost** | **Precision** | 29,33% | 61,68% | 95,65% | 100,00% | 100,00% | N/A | N/A | N/A |
|  |  | **Recall** | 100% | 100,00% | 100,00% | 98,48% | 98,48% | 0% | 0% | 0% |
|  | **robot-shop** | **Precision** | 78,57% | 78,57% | 78,57% | 78,57% | 100,00% | 100,00% | 100,00% | 100,00% |
|  |  | **Recall** | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
|  |  |  |  |  |  |  |  |  |  |  |
|  |  | **Precision average (** | 61,09% | 76,30% | 93,42% | 95,71% | **100,00%** | 100,00% | 100,00% | 100,00% |
|  |  | **Recall average (** | 100,00% | 100,00% | 100,00% | 96,84% | **96,84%** | 58,46% | 57,56% | 56,92% |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DB Mongo** |  | **Scores** | **1** | **2** | **3** | **4** | **5** | **6** |
|  | **comments-api** | **Precision** | 72,00% | 75,00% | 78,95% | N/A | N/A | N/A |
|  |  | **Recall** | 100,00% | 100,00% | 83,33% | 0,00% | 0,00% | 0,00% |
|  | **cinema-microservice** | **Precision** | 100,00% | 100,00% | 100,00% | N/A | N/A | N/A |
|  |  | **Recall** | 100,00% | 100,00% | 65,00% | 0,00% | 0,00% | 0,00% |
|  | **overleaf** | **Precision** | 78,36% | 92,92% | 99,06% | 100,00% | 100,00% | 100,00% |
|  |  | **Recall** | 100,00% | 100,00% | 100,00% | 72,38% | 11,43% | 10,48% |
|  | **cloudboost** | **Precision** | 53,16% | 63,92% | 97,94% | 100,00% | N/A | N/A |
|  |  | **Recall** | 100% | 100,00% | 94,06% | 36,63% | 0% | 0% |
|  | **robot-shop** | **Precision** | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% |
|  |  | **Recall** | 100,00% | 100,00% | 100,00% | 78,26% | 73,91% | 13,04% |
|  |  |  |  |  |  |  |  |  |
|  |  | **Precision average (** | 80,70% | 86,37% | **95,19%** | 100,00% | 100,00% | 100,00% |
|  |  | **Recall average (** | 100,00% | 100,00% | **88,48%** | 37,46% | 17,07% | 4,70% |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **DB Redis** |  | **Scores** | **1** | **2** | **3** | **4** | **5** | **6** |
|  | **overleaf** | **Precision** | 40,94% | 67,10% | 82,64% | 96,43% | 100,00% | 100,00% |
|  |  | **Recall** | 100,00% | 100,00% | 96,15% | 51,92% | 48,08% | 4,81% |
|  | **cloudboost** | **Precision** | 9,76% | 32,88% | 91,30% | N/A | N/A | N/A |
|  |  | **Recall** | 100% | 100,00% | 87,50% | 0% | 0% | 0% |
|  | **robot-shop** | **Precision** | 22,58% | 28,00% | 30,43% | 58,33% | 100,00% | N/A |
|  |  | **Recall** | 100,00% | 100,00% | 100,00% | 100,00% | 100,00% | 0,00% |
|  |  |  |  |  |  |  |  |  |
|  |  | **Precision average (** | 24,43% | 42,66% | **68,13%** | 77,38% | 100,00% | 100,00% |
|  |  | **Recall average (** | 100,00% | 100,00% | **94,55%** | 50,64% | 49,36% | 1,60% |

## Heuristics evaluation

For evaluating the individual relevance of our code fragment identification heuristics, we compute separately their precision and recall for each project codebase, based on the ground truth. We remove heuristics that make no contribution to the score.

Une image contenant texte, capture d’écran, Parallèle, nombre

Description générée automatiquement Une image contenant texte, capture d’écran, nombre, diagramme

Description générée automatiquement

Une image contenant texte, capture d’écran, diagramme, Rectangle

Description générée automatiquement