

# Self-supervised Information Retrieval Trained from Self-generated Sets of Queries and Relevant Documents

Gianluca Moro<sup>1</sup>[0000–0002–3663–7877], Lorenzo Valgimigli<sup>1\*</sup>[0000–0003–0309–771X],  
Alex Rossi<sup>1</sup>, Cristiano Casadei<sup>2</sup>, and Andrea Montefiori<sup>2</sup>

<sup>1</sup> Department of Computer Science and Engineering (DISI), University of Bologna, Cesena (FC), Via dell’Università 50, Italy  
{gianluca.moro,lorenzo.valgimigli}@unibo.it  
alex.rossi6@studio.unibo.it

<sup>2</sup> Maggioli S.p.A. Santarcangelo di Romagna (RN), Via del Carpino 8, Italy  
{cristiano.casadei, andrea.montefiori}@maggioli.it

## 1 Appendix

### 1.1 Query samples

**Passage:** gli operatori del diritto e il confronto con le esperienze europee ed internazionali, avvertono come gli interventi migliorativi debbano essere inseriti in modo sistemico, in tutti quegli ambiti regolamentari sensibili all’efficacia dello strumento di finanziamento in oggetto. Di seguito si riportano i dieci temi individuati nel documento quali aspetti di criticità e le rispettive proposte di ottimizzazione come sviluppate dall’UFP

**Query:** migliorativi ottimizzazione sensibili criticità esperienze

**Passage:** osservando in particolare che l’art. 223 C.d.S., nell’ipotesi di guida in stato di ebbrezza - ricorrente nella fattispecie - prevede l’emissione dell’ordinanza di sospensione cautelare della patente di guida da parte del Prefetto come un atto dovuto. Avverso tale sentenza l’automobilista ha proposto ricorso per Cassazione. La Suprema Corte ha esaminato anzitutto il secondo motivo di ricorso, ritenendolo fondato.

**Query:**223 C.d.S., nell’ipotesi di guida in stato di ebbrezza - ricorrente nella fattispecie - prevede l’emissione dell’ordinanza di sospensione cautelare della patente di guida da parte del Prefetto.

### 1.2 Translated Query Samples

**The not-fluent query Passage:** legal practitioners and the comparison with European and international experiences, warn how improvement interventions must be included in a systemic way, in all those regulatory areas sensitive to effectiveness of the financing instrument concerned. Below are the ten themes identified in the document as critical aspects and their respective proposals optimization as developed by the UFP

**Query:** improvement optimization sensitive critical experiences

---

\* Corresponding author.

**The fluent query Passage:** noting in particular that art. 223 C.d.S., in the hypothesis of driving while intoxicated - the applicant in the present case - provides for the issuance of the order precautionary suspension of the driving license by the Prefect as a act due. The driver brought an action against that judgment in order to Cassation. The Supreme Court first examined the second plea in law of appeal, considering it well founded. **Query:**223 C.d.S., in the hypothesis of driving while intoxicated - the applicant in the present case - provides for the issuance of the order precautionary suspension of the driving license by the Prefect .

**The entity-based query Passage:** ... total and, therefore, of the consideration for the contract. According to the T.A.R. such a hypothesis would be prosecutable only in the course of development of the service.... **Extracted name entity:** T.A.R.

### 1.3 Accuracy Measure

$$\begin{aligned} rank@n(Q, R) &= \sum_{q \in Q} \mathcal{R}(q, R_{:n}^q) \\ \mathcal{R}(q, R_{:n}^q) &= \begin{cases} 1 & \text{if } q \text{ in } R_{:n}^q \\ 0 & \end{cases} \end{aligned} \quad (1)$$

where  $Q$  is the set of  $q_0, \dots, q_j$  queries and  $R$  is the set of retrieved documents for each query  $R_0, \dots, R_j$ .  $R_{:n}^q$  indicates the top  $n$  retrieved documents for query  $q$ .

### 1.4 Metric space definition

### 1.5 Loss explanation

### 1.6 Dataset

### 1.7 Self-supervised vs Supervised learning

### 1.8 Model Architecture