Definition 2.2 (Query Result). The result of \( Q(Q_T, K_p) \) is a subset of ranking targets \( R_Q \subseteq R_T \) such that \( \forall r \in R_Q \), satisfying \( K_p \) consisting of the elements of \( R_T \) that satisfy every \( K_p \).

Example 2.1. Consider two example queries discussed in the introduction. The XPath expression 
\[
//\text{paper/author[./bio contains "DB, privacy"]} \text{- projects author elements. So author elements are ranking targets of this query.}
\]
The bio elements under an author element form the textual content of the "author document". The expression 
\[
//\text{paper[./year > 2004 and ./content contains "MapReduce hash join"\}
\]
projects paper elements. However, not all papers satisfy the pattern. The ranking targets of this query should only be those papers published after 2004.

While in Definition 2.1 we define ranking targets to be the same as query targets, in some scenarios, the two may be different. For example, consider the query:

\[
Q_3: //\text{paper/abstract contains "XML mapping"}/\text{author}
\]

By Definition 2.1, the query projects author elements, and thus all authors are ranking targets. The abstract of a paper written by an author is modelled as the content of an "author document", and is used to evaluate the relevance between the author and the keyword query. But there can be another interpretation of the query: we rank papers and project authors associated with them. In such a case, there is no relevance ranking for authors, but only papers. In general, it is helpful to separate ranking targets and query targets in query syntax. Since the main focus of the paper is not a new query language or a syntax, we stick to Definition 2.1 in the following of this paper.

The techniques discussed later can be applied when the two concepts are separated.

2.2 Ranking Model

While various IR ranking models have quite different theoretical foundations, they all boil down to a small set of statistics. IR heuristics [6] either describe roles of individual statistics or regulate their interactions. After taking a close look at the existing ranking functions, we classify the statistics into two types:

- **Local Statistics**: A local statistic \( S(d, Q_k) \) is a parameter that describes the property of the input query \( Q_k \) or an individual document \( d \). For example, query length is a parameter