

Experiment on Combining Sources of Evidence for Passage Retrieval *

Alexander Gelbukh,¹ NamO Kang,² SangYong Han²⁺

¹ National Polytechnic Institute, Mexico
gelbukh@gelbukh.com, www.Gelbukh.com

² Chung-Ang University, Korea
kang@archi.cse.cau.ac.kr, hansy@cau.ac.kr

Abstract. Passage retrieval consists in identifying short but informative runs of a long text, given a specific user query. We discuss the sources of evidence that help choosing likely high-quality passages, such as relevance to the user query and self-containedness. These measures are different from the traditional information retrieval procedure due to the use of the context of the passage.

1 Introduction

Unlike full document retrieval—a traditional task of information retrieval—passage retrieval task [2, 5] consists in identifying in a long document (or collection of long documents) short text runs relevant for a specific user query, which—unlike in question answering task—do not allow a simple factual answer.

In this paper we discuss the parameters affecting selection of such passages from the text of the document. We intentionally do not give any specific formulas since our experimental result do not yet allow us to reliably argue in favor of a specific way of calculation of these parameters.

2 The Method

Our algorithm consists in the following steps:

- Preprocessing,
- Candidate passage generation,
- Assessing various properties of each candidate passage,
- Combining the obtained scores for each property in a single value overall score.

Then the passages are presented to the user in the order of obtained scores. The steps of the algorithm and the specific quality measures are described in the following subsections.

* This research was supported by ITRI of the Chung-Ang University. Work was done when the first author was on Sabbatical leave at Chung-Ang University. ⁺Corresponding author.