

# Creating a Novel Semantic Video Search Engine Through Enrichment Textual and Temporal Features of Subtitled YouTube Media Fragments

Babak Farhadi

Department of Computer engineering  
University of Tehran, International Campus  
Kish Island, Iran  
babakfarhadi@ut.ac.ir

M. B. Ghaznavi-Ghoushchi

Dept. of E.E. School of Engineering  
Shahed University  
Tehran, Iran  
ghaznavi AT shahed.ac.ir

**Abstract**— Semantic video Annotation is an active research zone within the field of multimedia content understanding. With the stable increase of videos published on the famous video sharing platforms such as YouTube, more efforts are spent to automatically annotate these videos. In this paper, we propose a novel framework to annotating subtitled YouTube videos using both textual features such as all of portions extracted from web natural language processors in relation to subtitles, and temporal features such as the duration of the media fragments where particular entities are spotted. We implement SY-VSE (Subtitled YouTube Video Search Engine) as an efficient framework to cruising on the subtitled YouTube videos resident in the Linked Open Data (LOD) cloud. For realizing this purpose, we propose Unifier Module of Natural Language Processing (NLP) Tools Results (UM-NLPTR) for extracting main portions of the 10 NLP web tools from subtitles associated to YouTube videos in order to generate media fragments annotated with resources from the LOD cloud. Then, we propose Unifier Module of Popular API's Results (UM-PAR) containing the seven favorite web APIs to boost results of Named Entities (NE) obtained from UM-NLPTR. We will use dotNetRDF as a powerful and flexible API for working with Resource Description Framework (RDF) and SPARQL in .Net environments.

**Index Terms** — Subtitled YouTube video, NLP web tools, textual metadata, semantic web, video Annotation, video search.

## I. INTRODUCTION

For better understanding the discussions located in introduction section, we have divided them in six subsections that can be summarized as followings:

### A. Influence of Natural language processors on on-line video platforms

Nowadays, on-line video-sharing platforms, especially YouTube shows that video has become the medium of choice for many people interchanging via the net. On the other hand, the amazing increase of online video content confronts the consuming user with an indefinite amount of data, which can only be accessed with advanced multimedia search and management technologies in order to retrieve the few needles from the giant haystack. The most on-line video search engines provide a keyword-based search, where lexical ambiguity of natural language often leads to imprecise and defective results. For instance, YouTube supports a keyword-based search within the textual metadata provided by the users, accepting all the

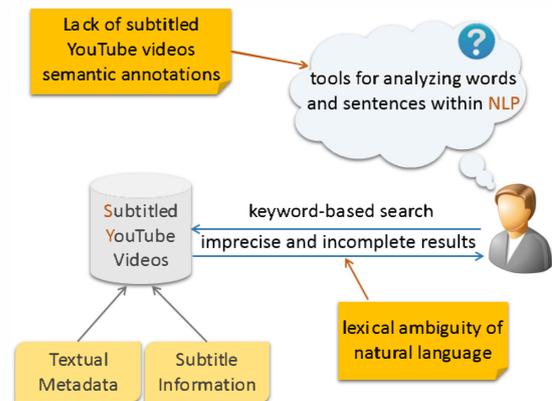


Figure 1. Structure of subtitled Youtube keyword-based search.

shortcomings caused by e.g. homonyms (see Figure 1). Optimal results are possible through the analysis of the available textual metadata with NLP web tools and popular web APIs, especially given the availability of subtitles on YouTube videos.

### B. Semantic multimedia Annotations and Linked Data

Role of semantic web technologies is to make implicit meaning of content explicit by providing suitable metadata annotations based on formal knowledge representations. In this area, Linked Data (LD) means to expose, share and connect pieces of data, information and knowledge on the semantic web using Uniform Resource Identifiers (URI) for identification of resources and RDF as a structured data format. It is creating the relationships from the data to other sources on the Web. These data sets are not only accessible by human beings, but also readable for machines. LOD can aims to publish and connect open but heterogeneous databases by applying the LD principles. The aggregation of all LOD data set is denoted as LOD-Cloud. In this direction, the term “media fragment” refers to the inside content of multimedia objects, such as a certain zone within an image, or a five-minute segment within a one-hour video. Most research about media fragments focus on exposing the closed data, such as tracks and video segments, within the multimedia file or the host server using semantic Web and LD technologies.