Ferri F., Grifoni P., Padula M, **Using Shape to Index and Query Web Document Contents**, Journal of visual languages and computing, Academic Press. - London, UK, 2002.

**Abstract**: The mass of information now available on web sites has greatly increased the web's popularity and, consequently, the demand for its convenient access and use. The Internet/Intranet phenomenon provides some fifty million people with access to multiple sources of information. However the current techniques for retrieving and navigating do not make it easy for the data user to satisfy his/her needs. This paper aims to show how user communities can be enabled to unlock the information stored in web documents. W e start from the observation that authors shape their documents to clarify the intended meaning, and readers in turn exploit document shape to synthetically grasp this meaning. Thus the Web document can be seen as a structure composed of different types of information units (such as images, tables, movies, videos, sounds, titles, and paragraphs). These units are shaped, represented, and organized in the document so that it transmits its message according to the cultural formation of its author. From our investigation of collections of web documents we have derived some heuristics to use in shaping the document in order to emphasize its content. These heuristics can be exploited to manage and retrieve semantic information on the Web.
Since human computer interaction with the Web is preponderantly visual, we propose a visual approach in customizing web documents, and in indexing and querying the Web through the browser. This approach is based on a method of annotating HTML documents so that their shape and contents can be reorganized to satisfy the requirements of different readers.