Semantic indexing of sports videos is a subject of great interest to researchers working on multimedia content characterization. Sports programs appeal to large audiences and their efficient distribution over various networks should contribute to widespread usage of multimedia services. In this paper, we propose a semantic indexing algorithm for soccer programs which uses both audio and visual information for content characterization. The video signal is processed first by extracting low-level visual descriptors from the MPEG compressed bit-stream. The temporal evolution of these descriptors during a semantic event is supposed to be governed by a controlled Markov chain. This allows to determine a list of those video segments where a semantic event of interest is likely to be found, based on the maximum likelihood criterion. The audio information is then used to refine the results of the video classification procedure by ranking the candidate video segments in the list so that the segments associated to the event of interest appear in the very first positions of the ordered list. The proposed method is applied to goal detection. Experimental results show the effectiveness of the proposed cross-modal approach.