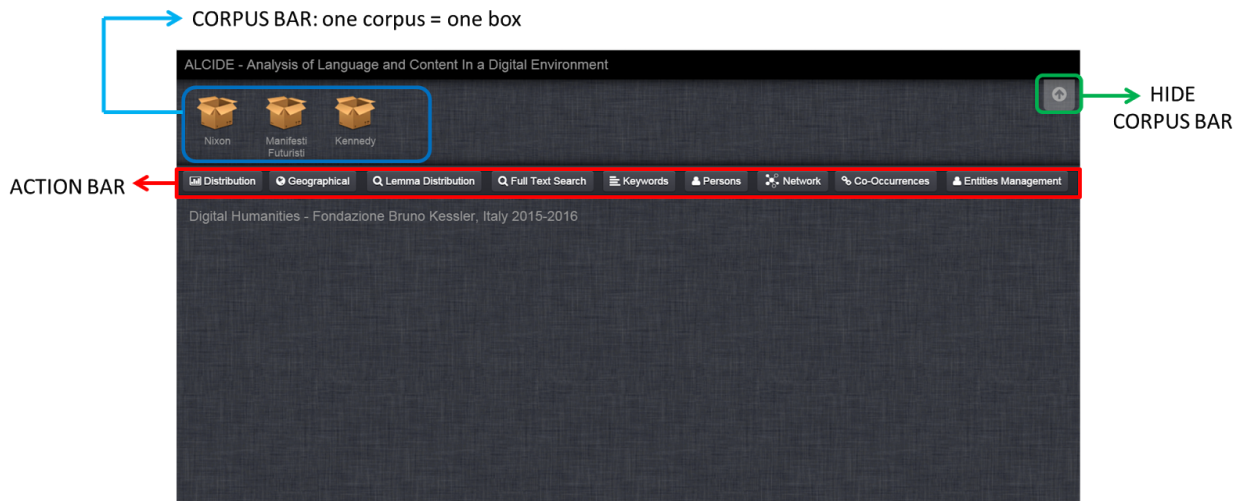


ALCIDE: Quick Start Guide

Digital Humanities Research Group - Fondazione Bruno Kessler, Trento (Italy)

<http://dh.fbk.eu/>

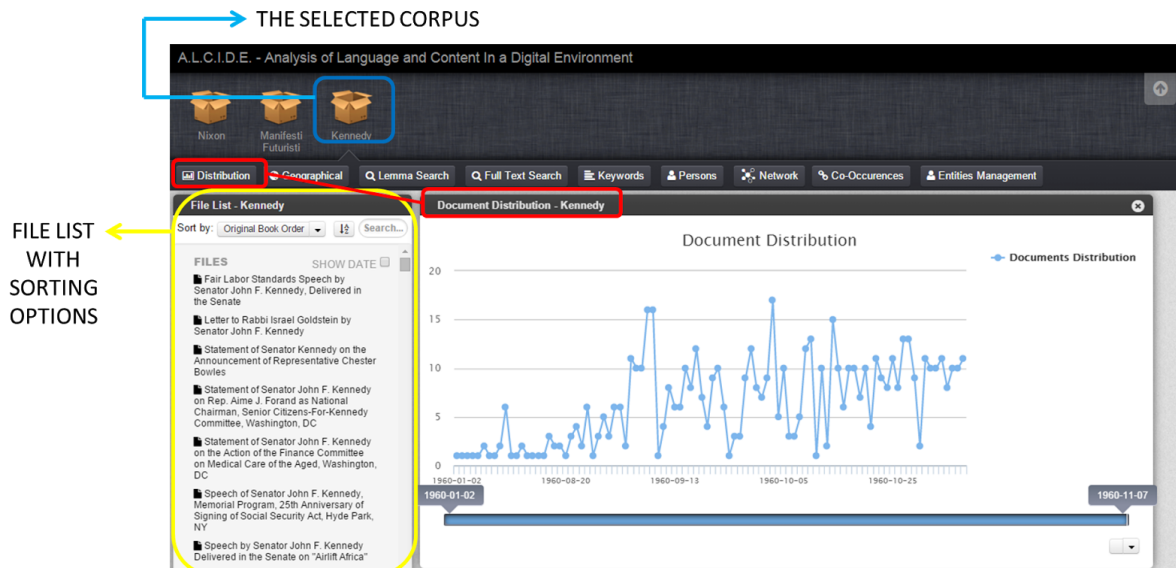
- Log in and explore the interface



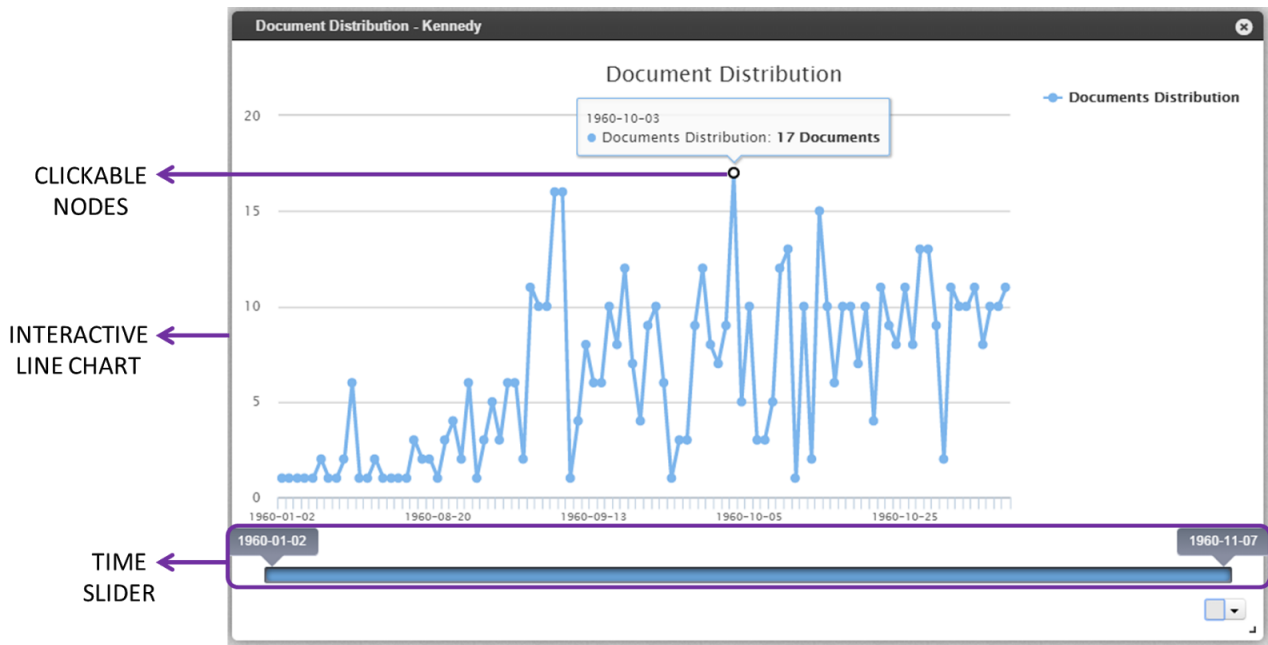
- The following buttons (i.e. analyses) are available in the *Action Bar*:
 - Distribution*: number of documents per day or year (depending on the granularity of the corpus).
 - Geographical*: map with the geographical distribution of the documents according to their place of publication (*Document metadata*) and map of the geo-political entities automatically extracted from the documents with a named entity recognizer (*Cited Geonames*).
 - Lemma Search*: one or more lemmas of nouns, verbs and adjective can be searched. Lemmas should be separated by a space. Lemmatization and part-of-speech tagging are automatically performed.
 - Full Text Search*: search for words and expressions made by more than one word. For an exact match, use double quotes (e.g. "peace" / "world war"). Avoid double quotes to search for substrings (e.g. searching for *peace* will retrieve also *peaceful*). Searching for words/expressions in double quotes separated by a space corresponds to an AND operation (e.g. the search "nuclear" "atomic" retrieves the documents containing both words).
 - Keywords*: key-concepts automatically extracted from the corpus.
 - Persons*: proper names of persons mentioned in the corpus and extracted with a named entity recognizer. Persons' names are clustered in a semi-automatic way (see *Entity Management* button below).
 - Network*: persons co-occurrence network automatically extracted from the corpus. This analysis is based on the idea that two persons who tend to be

mentioned together in a corpus share some commonality or relation from the author's perspective.

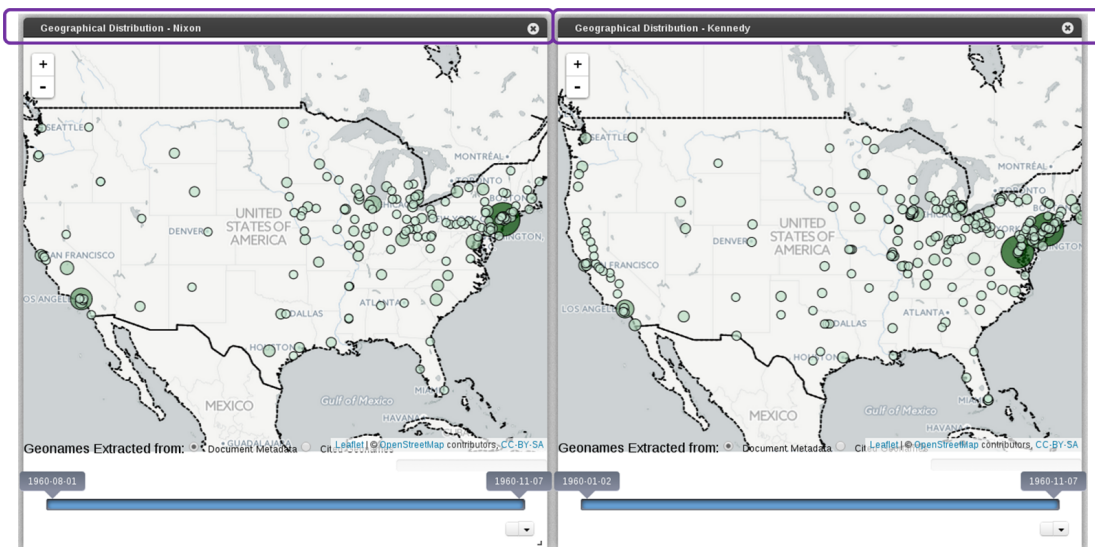
- h. *Co-Occurrences*: search for expressions, persons' names, geonames, nouns, adjectives and verbs that appear in the context of a given word (or a sequence of words).
 - i. *Entity Management* - NOT FULLY AVAILABLE IN THE DEMO VERSION: remove wrongly identified persons' names and cluster persons' names (putting "Rockefeller" and "Nelson Rockefeller" under the same person entity "Nelson Rockefeller") using an agglomerative and flat algorithm and a GUI for manual revision. Clusters and changes performed here are instantly stored in the database thus the analyses available under the *Persons* and *Network* buttons are updated on the fly.
- Click on a corpus in the *Corpus Bar* : documents are listed on the left, the *Distribution* widget opens on the right. The list of documents on the left is updated every time a search is performed and/or an interactive visualization is manipulated.



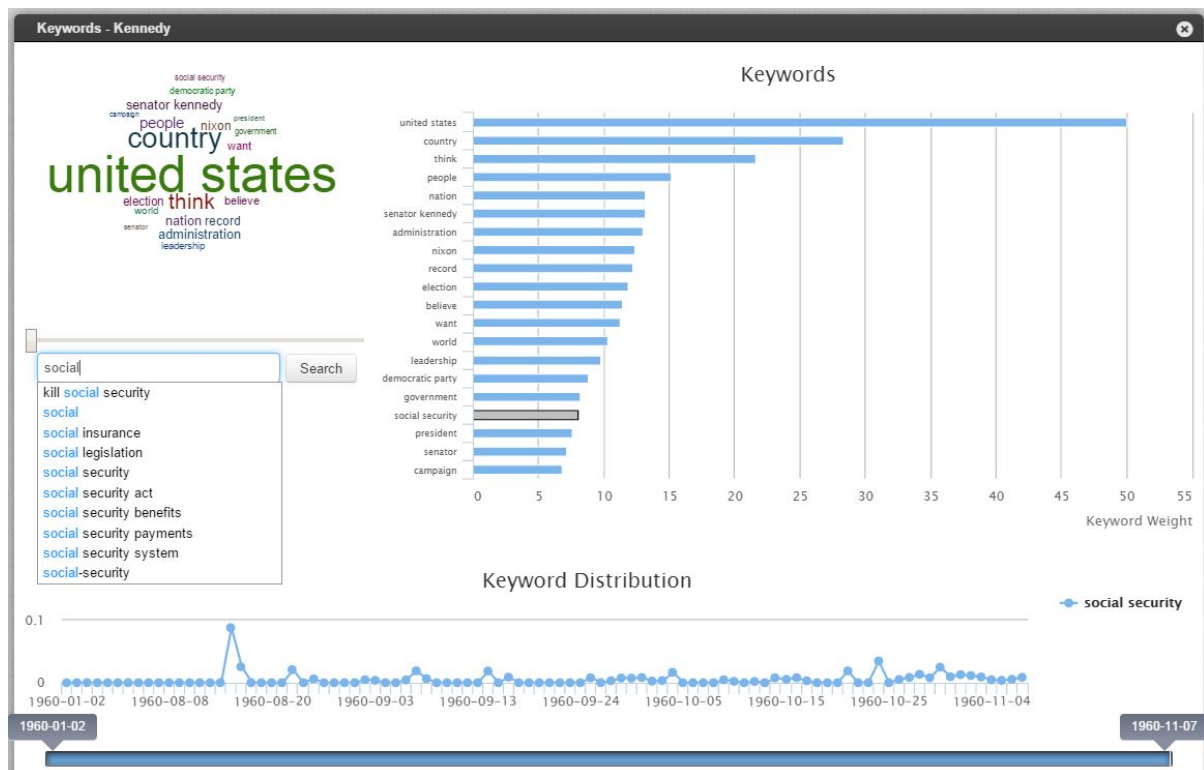
- ALL *widgets* contain a horizontal time slider and one or more visualizations. All visualizations (with the only exception of word clouds) are interactive. Line charts, as in the figure below, can be zoomed in by choosing a specific time span (thus moving the start/end point of the time slider) or by clicking and dragging the chart itself.



- Each button in the *Action Bar* opens the corresponding widget. Widgets can be closed before starting a new action. Alternatively widgets can be arranged one under the other or one next to the other so as to perform a visual comparison of the results. Possible comparisons can be performed to display::
 - the same action on the same corpus (e.g. two *Distribution* widgets for the “Kennedy” corpus, one zoomed in to September 1960 and one zoomed in to October 1960);
 - different actions on the same corpus (e.g.) *Distribution* and *Geographical* widgets for the “Kennedy” corpus;
 - different actions on two different corpora (e.g. *Lemma Search* for the “Kennedy” corpus and *Full Text Search* for the “Nixon” corpus);
 - same action but on two different corpora (e.g. one *Geographical* widget for the “Kennedy” corpus and one *Geographical* widget for the “Nixon” corpus, see Figure below).

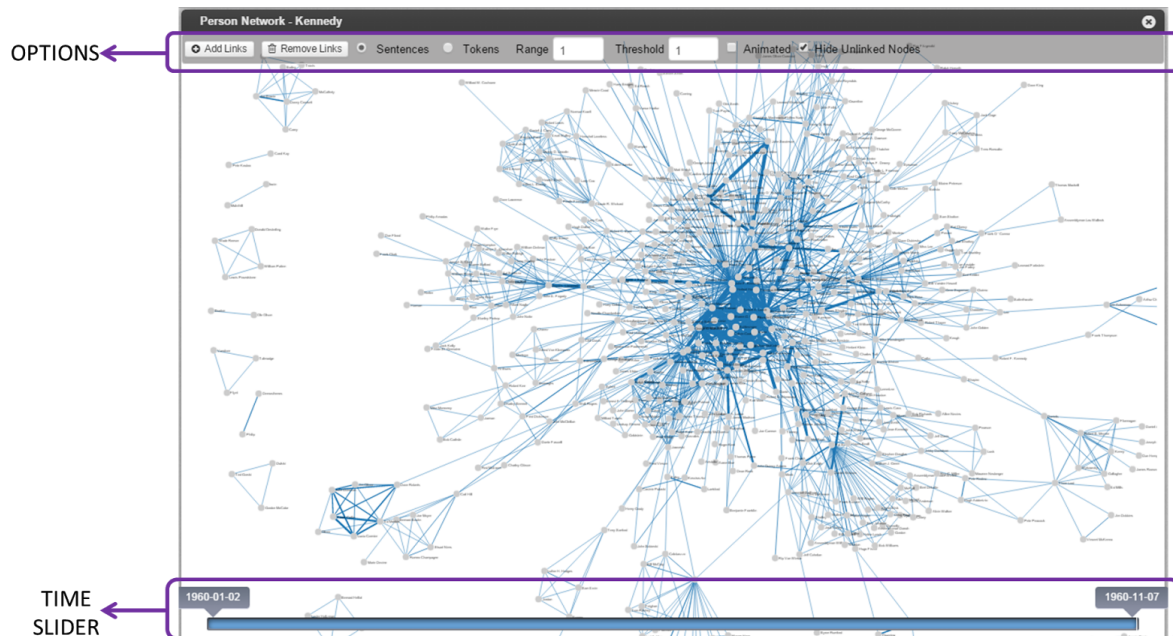


- The widgets of *Keywords* and *Persons* present the same type of visualization. The slider under the word cloud can be used to increase the number of keywords/persons' names shown in the word cloud and in the bar chart. Bars in the chart can be clicked to see how the relevance of the selected keyword changes over time: this distribution is displayed in a line chart above the time slider. The search field allows to search for a specific keyword/persons' name: suggestions are proposed and can be selected from a top-down menu.



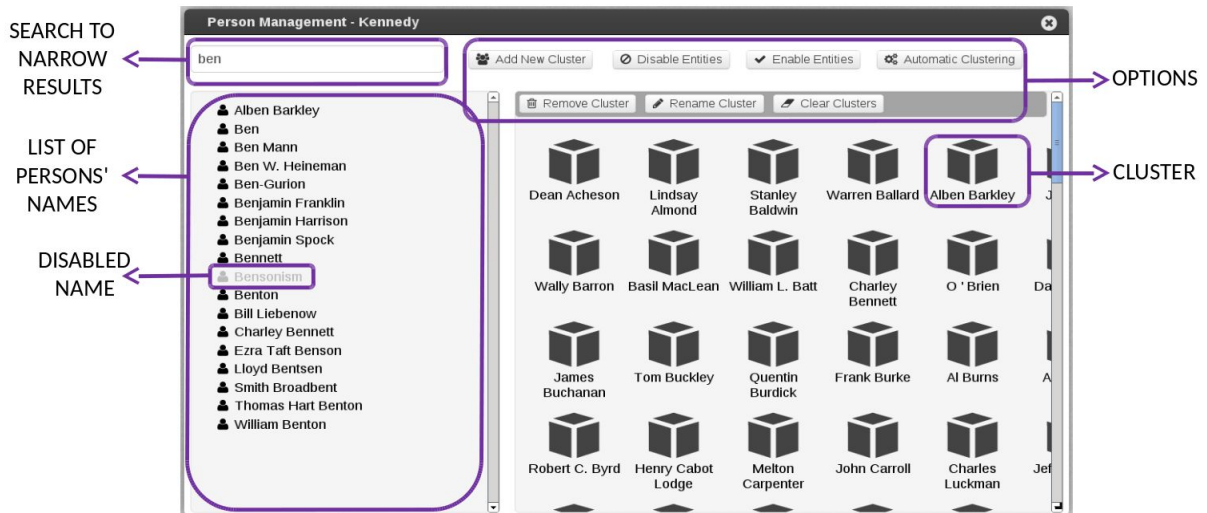
- The *Network* widget shows co-occurring persons' names. Depending on the computer speed, this analysis can take few minutes when it is run for the first time. Many options are available to reduce the network complexity. All changes made to these options modify the network on the fly:
 - Add Links/ Remove Links*: show and hide edges between nodes;
 - Sentences/Tokens*: change the co-occurrence window type;
 - Range*: co-occurrence window width, i.e. number of sentences/tokens;
 - Threshold*: minimal number of co-occurrences;
 - Animated*: stop the animation;
 - Hide Unlinked Nodes*: remove isolated nodes, i.e. nodes with no connections.
 - Time slider: select a specific time span or point in time.

Selecting one or more nodes, all documents containing the corresponding persons' names are retrieved and listed in the file bar on the left.

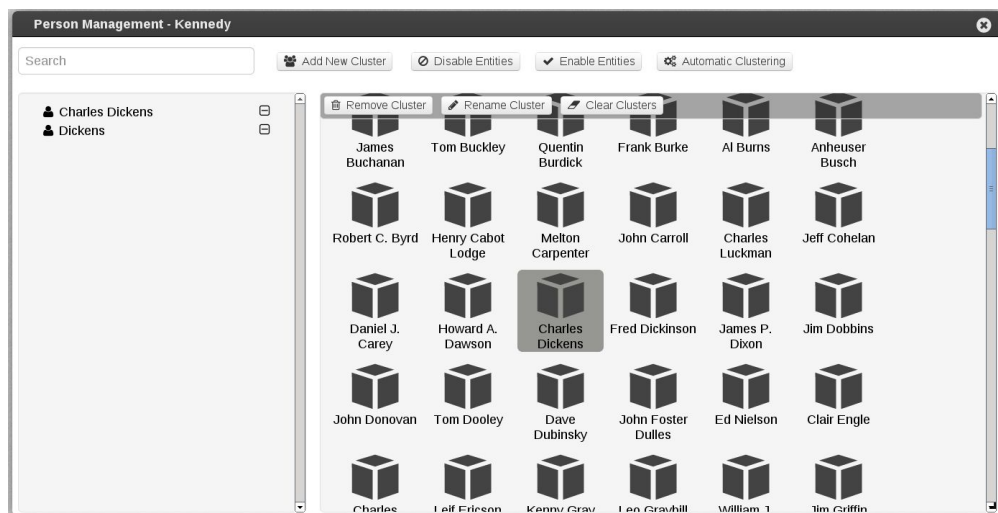


- *Co-Occurrence* widget allows to search for words, expressions, persons' names and geo-names that appear in the context of a given word(s). Co-occurrences are displayed as a network with the central node being the word(s) of interest. The size of the surrounding nodes reflects the frequency of each co-occurrence. Nodes are clickable so to retrieve documents containing the co-occurrence. The following options are available:
 - a. *Range*: co-occurrence window width, i.e. number of words (from 1 to 60);
 - b. *Before/Both/After*: search only in the right context or in the left context, or both;
 - c. *Expressions*: extract co-occurring expressions with specific part-of-speech patterns (e.g. verb+noun);
 - d. *Persons*: extract co-occurring persons' names;
 - e. *Geonames*: extract co-occurring names of geo-political entities;
 - f. *Nouns/Verbs/Adjectives*: extract only nouns, verbs or adjectives co-occurring with the word(s) of interest.
 - g. *Colorize*: color the nodes on the basis of the frequency of each co-occurrence (*Dimension*) or on the basis of the sentiment conveyed by the co-occurring word(s) (*Polarity*: red is negative, green is positive and yellow is neutral).
- Several actions (NOT AVAILABLE IN THE DEMO VERSION) can be performed using the *Person Management* widget:
 - a. *Add New Cluster*: create a cluster of persons' names after selecting two or more names in the list on the left. To select a name, just click on it;
 - b. *Disable Entities*: deactivate in the database a name wrongly identified as a name of person;
 - c. *Enable Entities*: activate again a name previously disabled;
 - d. *Automatic Clustering*: run an algorithm performing the clustering of persons' names;
 - e. *Remove Cluster*: delete a cluster without disabling names contained in it;

- f. *Rename Cluster*: change the name of a cluster;
- g. *Clear Clusters*: remove all clusters.



Names grouped together through clustering can be visualized on the left hand side of the widget by clicking on a cluster (i.e. a box on the right). THIS FUNCTION IS AVAILABLE ALSO IN THE DEMO VERSION.



- In every moment it is possible to read a single document by clicking on its title in the file list. Three analyses and the corresponding visualizations are available at the document level:
 - a. *Keywords*: bar chart with the 20 most relevant key-concepts automatically extracted from the document;
 - b. *Persons*: bar chart with all persons' names automatically extracted from the document;
 - c. *Sentiment*: a gauge with the overall sentiment value of the document calculated as the sum of all the polarity scores assigned to the single words in the document.

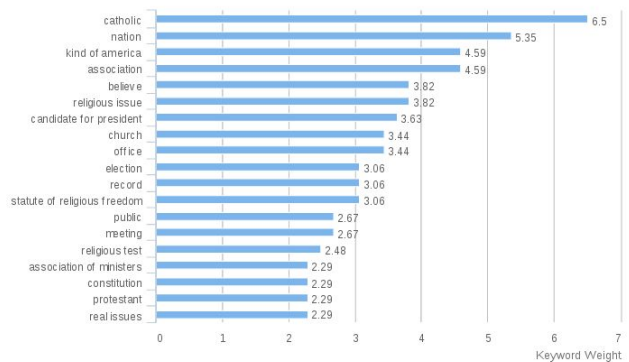
Speech of Senator John F. Kennedy, Greater Houston Ministerial Association, Rice Hotel, Houston, TX

Houston, Texas, 1960-09-12

ANNOUNCER. From Houston's Rice Hotel, Senator John Kennedy is about to address a special meeting of the Greater Houston Ministerial Association to which he has been invited. During this telecast, Senator Kennedy will participate in an informal question and answer period. The telecast of this meeting is sponsored by the Kennedy-Johnson Texas Campaign Committee, and is being seen throughout Texas on a special 22-station network. The audience you are seeing is composed of clergymen of the Houston area who have been invited by the association. Rev. Herbert Meza will introduce the Democratic presidential candidate. The meeting is about to be called to order by the president, Rev. George Peck. Mr. RECK. May I call this special meeting of the Association of Ministers of Greater Houston to order. Let us stand for prayer. God be merciful unto us and bless us and cause His face to shine upon us, that Thy way may be known upon earth. Thy saving health among all nations. God shall bless us and all the ends of the earth shall praise Him. With these words of the psalmist, we stand before Thee, O God, as our only Sovereign Lord. Forgive us, Good Lord, and show us Thy mercy. Let Thy grace rest upon our Nation and do not take our Gospel life from us, incline our ears to Thee and to Thy will and show us always the truth that makes and keeps men free, in the name of our Lord Jesus Christ we pray, amen. We are very happy that so many of you ministers are present at this meeting. The treasurer of our association, Pastor E. S. Harrison, has wondered to me if some of you would not like to pay your dues for this time, this year which begins with September. I am sure that he will be in the lobby after this session, ready to shake your hands. [Laughter.] We are very happy to see so many of you ministers present, and we want this to be a true meeting of the association, under the policy of your executive committee this year, we wish to have as guests for regular and special meetings as many personalities of note and reputation as possible. The purpose, of course, is to provide not only a good program, but to give knowledge and enlightenment to the spiritual leaders of our community. Thus, a similar invitation was extended by the association to Mr. Nixon. Please understand that this is not a political rally. This is a meeting of the association of ministers, and we rely upon your sense for good order, proper respect for the nominee to the highest office of our land, and good Christian behavior generally. Our little mouse has grown into a lion of significance. This has not been our original intention, but things happen these ways. Nevertheless, the atmosphere will be informal here, an informal gathering of ministers, and may such atmosphere be maintained. May I speak a welcome to all of you. I recognize at this time the Reverend Herbert Meza, vice president of the

Text Classification

Keywords



Relevance

Persons

