Finding Genealogy Facts with Linguistic Analysis

Peter Lindes, Deryle W. Lonsdale, David W. Embley
Brigham Young University

© 2014 Peter Lindes
Semi-Structured Text

243314. Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865, son of Mary Ely and Gerard Lathrop; m. 1856 Mary Augusta Andrus, 992 Broad St., Newark, N. J., who was b. 1825 dau. of Judge Caleb Halstead Andrus and Emma Sutherland Goble. Mrs. Lathrop died at her home, 992 Broad St., Newark, N. J., Friday morning, Nov. 4, 1898. The funeral services were held at her residence on Monday, Nov. 7, 1898, at half-past two o'clock P. M. Their children:

1. Charles Halstead, b. 1857, d. 1861
2. William Gerard, b. 1858, d. 1861
3. Theodore Andrus, b. 1860
4. Emma Goble, b. 1862

Miss Emma Goble Lathrop, official historian of the New York Chapter of the Daughters of the American Revolution, is one of the youngest members to hold office, but one whose intelligence and capability qualify her for such distinction.
A Simple Ontology

Charles Christopher Lathrop

born on 1817

has

DeathDate

died on 1865
Unstructured Text

Children of James Harwood, No. 103.

229. M. K., born July 20, 1865 in Eden, Vt. He married Elijah Spencer, Dec. 25, 1851. They had five children: Arvila, born in 1852, is not living; Mariette, born Dec. 25, 1854, married Jonathan Snyder, have a family; Leverett, born Feb. 6, 1857, married Cora Smith, Nov. 2, 1879, had two children, Perry F. and Ida I. Leverett died May 21, 1910; Rosa E., born Jan. 13, 1860, married Emmett Byers, and have children; and Harrison, born about 1862, is not living. Elijah Spencer died in the Union army in 1863, and his widow married Jonathan Squires, who was born in Ohio, July 25, 1823, by whom he had one son, J. Wilbur, born June 16, 1865, in DeKalb county, Ind., married Cora M. Thomas, Aug 24, 1887, they reside in St. Joseph, Mich., five children. Myra Squires died in Allen county, Ind., Feb. 13, 1874.
A More Complex Ontology

Feb. 13, 1874

Myra Harwood
Jonathan Squires
J. Wilbur Squires
OntoSoar Architecture
Two Cases

• Case 1:

    Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865, son of Mary Ely and Gerard Lathrop ;

• Case 2:

    his widow married JONATHAN SQUIRES, who was born in Ohio, July 25, 1823, by whom she had one son, J. Wilbur, born June 16, 1865, in DeKalb county, Ind.
Charles Christopher Lathrop, ..., b. 1817, d. 1865,
Charles Christopher Lathrop, …, son of Mary Ely and Gerard Lathrop;
Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865, son of Mary Ely and Gerard Lathrop;
his widow married JONATHAN SQUIRES, ..., by whom she had one son, J. Wilbur, born June 16, 1865, ...
his widow married JONATHAN SQUIRES, who was born in Ohio, July 25, 1823, by whom she had one son, J. Wilbur, born June 16, 1865, in DeKalb county, Ind.
his widow married JONATHAN SQUIRES, who was born in Ohio, July 25, 1823, by whom she had one son, J. Wilbur, born June 16, 1865, ...
84 GENEALOGICAL HISTORY.
228. JAMES, born July 23, 1835, in Wayne county, Michigan. He married BRIDGET HACKETT, a sister of his brother Edmund’s wife, June 20, 1858. She died about 1875, was a member of the Catholic Church. Mr. Harwood died in Fenwick, Michigan, Aug. 19, 1910.
Children of JAMES HARWOOD, no. 103.
229. MYRA, born July 26, 1835, in Eden, Vt. She married ELIJAH SPENCER, Dec. 25, 1851. They had five children: Averilla, born in 1852, is not living; Mariette, born Dec. 25, 1854, married Jonathan Snyder, have a family; Leverett, born Feb. 6, 1857, married Cora Smith, Nov. 2, 1879, had two children, Perry F. and Ida I. Leverett died May 21, 1910; Rosa E., born Jan. 13, 1860, married Emmett Byers, and have children; and Harrison, born about 1862, is not living. Elijah Spencer died in the Union army in 1863, and his widow married JONATHAN SQUIRES, who was born in Ohio, July 25, 1823, by whom she had one son, J. Wilbur, born June 16, 1865, in DeKalb county, Ind., married Cora M. Thomas, Aug. 24, 1887, they reside in St. Joseph, Mich., five children, Mrs. Myra Squires died in Allen county, Ind., Feb. 13, 1874.
230. HARRISON, born May 21, 1837, in Allen county, Ind. He enlisted Sept. 5, 1861, in the 44th regiment, Indiana volunteers, which went south from Indianapolis, Nov. 26, 1861. He was killed in the Stone River fight, in the great battle of Murfreesborough, Dec. 31, 1862. He was a member of the Methodist church.
231. EDMUND, born April 25, 1840, in Allen county, Ind.
Conclusions

Accomplishments
• Preprocessor/Segmenter
• Adapted LG Parser for genealogy text
• Built a construction grammar meaning analyzer
• Mapped meaning schemas to user-defined ontologies
• Integrated all into a single processing pipeline
• Partially implemented – works well so far

Future work
• Complete implementation
• Evaluate on a broader corpus
• Explore using a named entity recognizer
• Build in much more linguistic knowledge
• Research how to make the system learn linguistic knowledge