indium
Make Technology Work

Information Extraction from survey text in land deed documents

Business

Data Annotation

Domain

Real Estate

Tools

Gate Software, Named entity recognition, LSTM -CRF

CUSTOMER BACKGROUND

The customer is a real estate and infrastructure consulting services provider. Their expertise allows their customers (real estate property owners – both public and private) make better informed decisions about infrastructure projects and cut costs in project implementation. They help efficiently design and provide services to real estate property owners in the U.S.

BUSINESS REQUIREMENT

The client being in the real estate space wanted to automate the process of extracting information of interest from survey text in land deed documents. The information required was in the form of a dozen entities like distance, direction, and curve details for each traverse. To effectively train the models to extract this information accurately, it was imperative to have accurate annotated training data.

CHALLENGES

- Training the model required good quality annotated data
- Accuracy and consistency of data
- Controlling the cost of labelling the data

## SOLUTION OVERVIEW AND IMPLEMENTATION

- LSTM CRF was used to solve the problem of named entity recognition.

  Annotation Details
- GATE software was used to streamline the tagging process.
- Over 12 different entity types were tagged in the documents.
- Total of 2000+ documents were annotated and converted to XML format, which can then be directly consumed by the model.

BUSINESS IMPACT

- 30% reduction in time taken to annotate the data, owing to streamlined processes
- Effective quality control mechanisms minimized human errors, delivering a high level of accuracy
- Model trained using the annotated data was able to achieve 85% F1 score
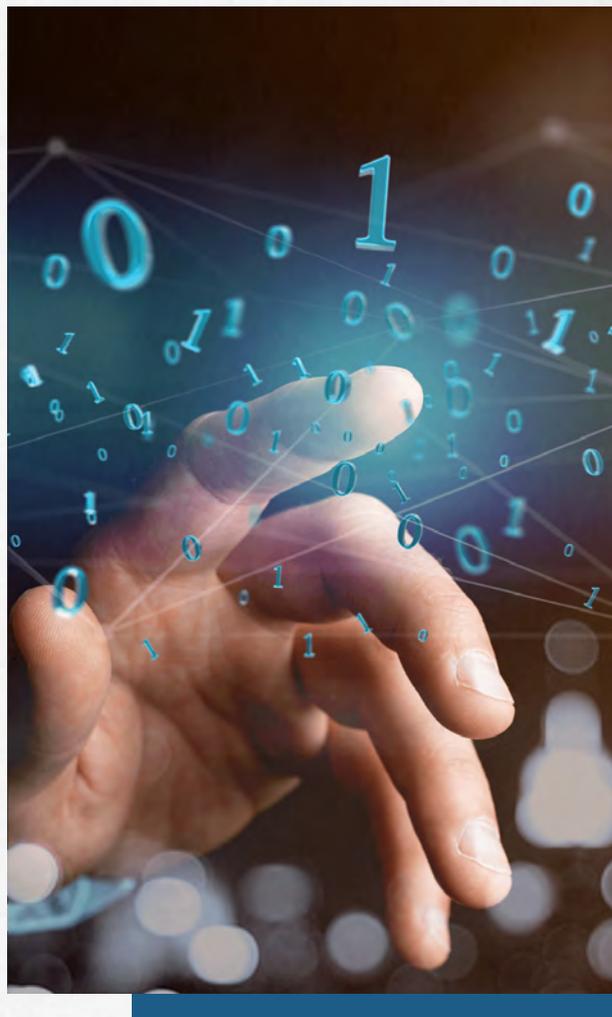
APPENDIX

GATE software was used for the annotation of the files. Screenshot of the app below
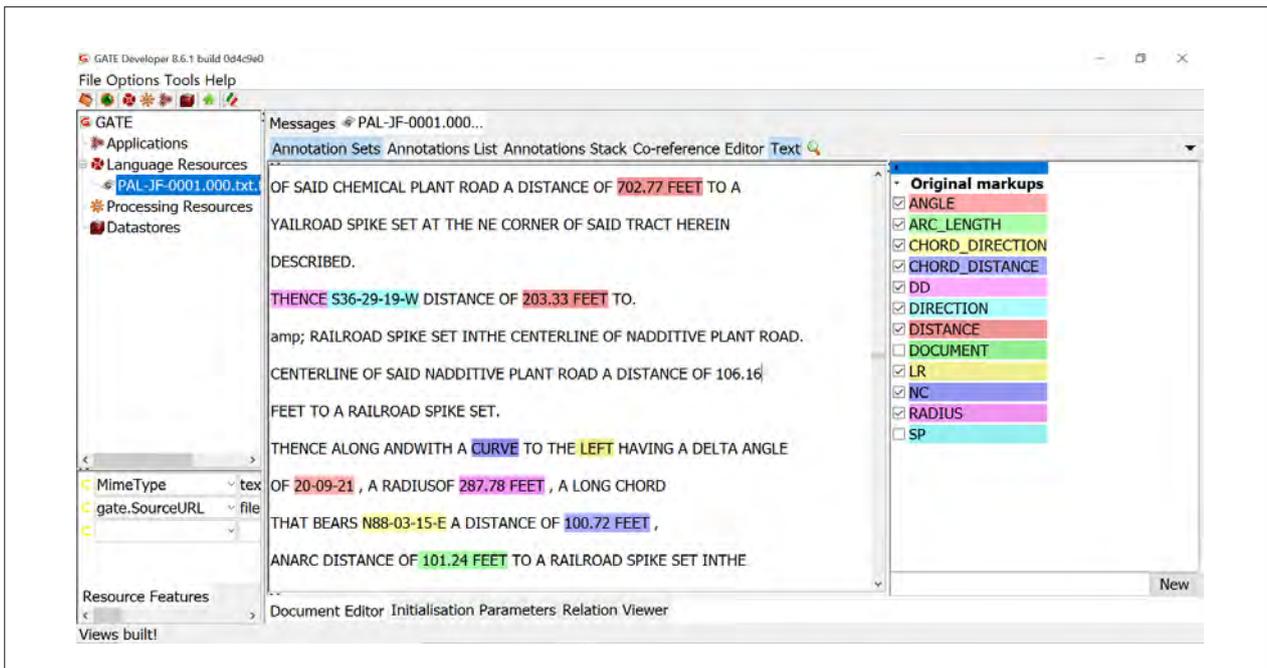
ANNOTATED FILES

Here are some sample files before and after annotation:

BEGINNING AT AN IRON STOB ON THE SOUTH SIDE OF HAW RIVER-HOPEDALE ROAD, CORNER WITH BULLARD AND BALLARD AND THOMPSON; RUNNING THENCE N.63-W. 110 FT. TO AN IRON STOB, CORNER WITH R.B. BAKER; THENCE S.66-50-E. 164 FT. TO AN IRON STOB, CORNER WITH BULLARD AND THOMPSON; THENCE S.51-W. 398 FT. WITH THE LINE OF BULLARD AND THOMPSON TO THE BEGINNING, CONTAINING 0.87 ACRES, MORE OR LESS, AS SURVEYED BY G.P. HOGAN, SURVEYOR, MARCH 12, 1948. A MORE MODERN DESCRIPTION OF THE ABOVE PROPERTY IS AS FOLLOWS: A CERTAIN TRACT OR PARCEL OF LAND LYING AND BEING IN HAW RIVER TOWNSHIP, ALAMANCE COUNTY, N.C., ADJOINING THE LANDS OF NCSR 1737 (HAW RIVER-HOPEDALE ROAD), ROBERT J. MULLIS, AND PROPERTY OF E. THOMPSON AND THELMA T. WATTS AND BEING MORE PARTICULARLY DESCRIBED AS FOLLOWS: SEE ATTACHED SHEET FOR CONTINUED DESCRIPTION BOOK 504 PAGE 346 BEGINNING AT A P.K. NAIL IN NCSR 1737 (HAW RIVER-HOPEDALE ROAD), CORNER WITH ROBERT J. MULLIS; RUNNING THENCE ACROSS NCSR 1737 (HAW RIVER-HOPEDALE ROAD) AND WITH SAID MULLIS N.41-36-19-E. 271.99 FT. TO AN OLD IRON STAKE, CORNER WITH ROBERT J. MULLIS; RUNNING THENCE WITH SAID MULLIS S.66-50-E. 163.82 FT. TO AN OLD IRON STAKE IN THE LINE OF THE PROPERTY OF E.L. THOMPSON AND THELMA T. WATTS, CORNER WITH ROBERT J. MULLIS; RUNNING THENCE WITH SAID THOMPSON AND WATTS S.51-00-W. 398 FT. TO A POINT IN THE SOUTHERN PORTION OF THE ROAD RIGHT OF WAY OF SAID NCSR 1737 (HAW RIVER-HOPEDALE ROAD), CORNER WITH SAID THOMPSON AND WATTS (AN IRON STAKE BEING SET 38.53 FT. FROM THE CORNER ON THIS LINE); RUNNING THENCE WITH AND ACROSS SAID NCSR 1737 (HAW RIVER-HOPEDALE ROAD) N.63-08-W. 110.00 FT. TO THE BEGINNING;

ABOUT INDIUM

Indium Software is a leading provider of Digital Engineering solutions with deep expertise in Application Engineering, Data and Analytics, Cloud Engineering, DevOps, Digital Assurance and Gaming.

Over the past decade, Indium has built strong relationships with over 100 clients spanning ISVs, Global 2000 as well as born-digital companies across North America, India, Europe and the Asia-Pacific region, and with ecosystem partners such as AWS, Mendix, Striim, Denodo and Claris.

indium
Make Technology Work

USA	INDIA	UK	SINGAPORE
Cupertino | Princeton	Chennai | Bengaluru | Mumbai	London	Singapore
Toll-free: +1-888-207-5969	Toll-free: 1800-123-1191	Ph: +44 1420 300014	Ph: +65 6812 7888

www.indiumsoftware.com

For Sales Inquiries
sales@indiumsoftware.com

For General Inquiries
info@indiumsoftware.com