



特許情報に関する言語生成モデルを 活用した知的財産創造手法の開発

Development of Intellectual Property Creation Method
Using Language Generation Model on Patent Information

Shigeaki Onoda

Graduate School of Information Engineering, Toyama Prefectural University
t855005@st.pu-toyama.ac.jp

Wednesday, 4 17, 2019,
Toyama Prefectural Univ.



行ったこと

当初の目的

- 特許（英語）データをより集める
- テストコレクションを NII に要請
- 集めたデータでプロトタイプモデルを構築

トラブル

様々なトラブルに見舞われ、テストコレクション以外はクリアできなかった

- GPU のロードの失敗
- スクレイピング時のロボット判定によりアクセス拒否される



ソリューション

GPU マシンについて

今回のトラブルに際して GPU ドライバー設定の資料を書いたので
小野田が修了後はそちらを参考にトラブルを対処

スクレイピング時のアクセス拒否

どのような状況か実際に行ってみる



スクレイピングのアクセス拒否回避方法

GPU マシンについて

アクセスが拒否されるということは同じ IP から一定期間中に大量のアクセスを行ったため、よって以下の解決策が考えれる

- プロキシサーバーを噛ませて IP を大量作成
- 匿名通信システム Tor を経由してスクレイプ
- 静的データを使う（NII の返答待ち）