Summer Internship 2024

Introduction

University of Maryland – College Park

Major: Computer Science

Development Intern

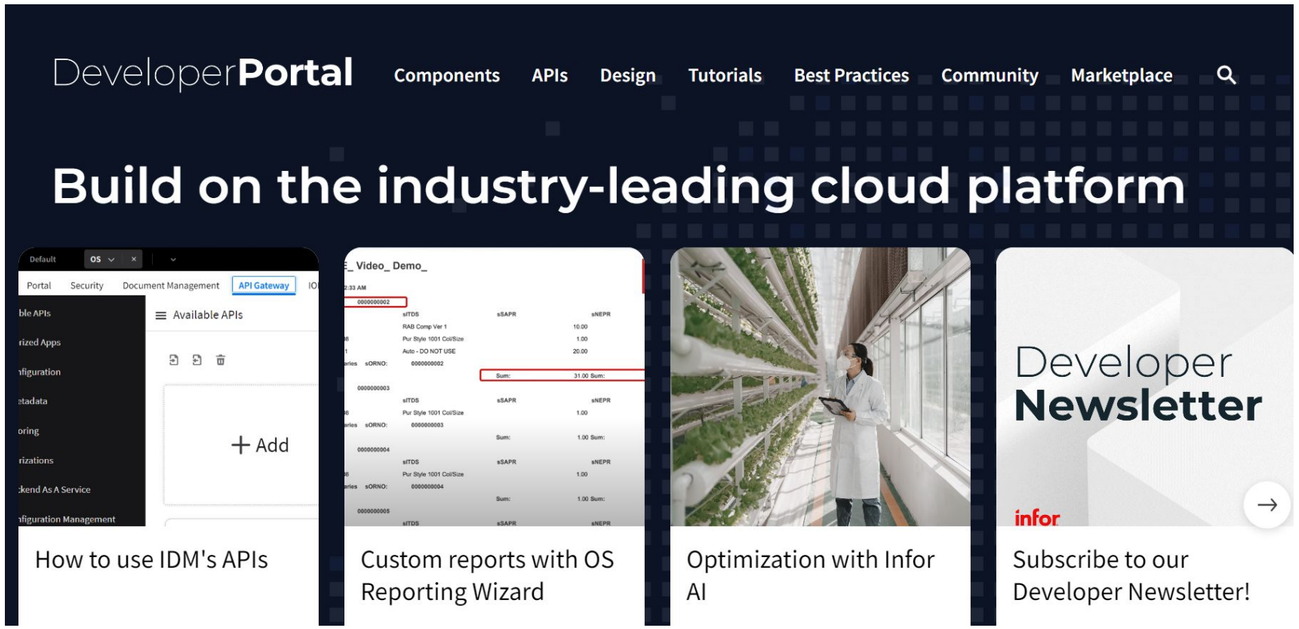
Overview

An Outsider's Perspective of the Developer Portal

Task: Explore the Developer Portal from an outsider's perspective and provide first impressions.

Observations:

- Inconsistency with content formatting and tutorials
- Outdated information
- Missing written tutorials for some videos



Problem Statements

To address those observations, solutions were proposed through two projects.

Project 1: Tutorials

This project addresses the following:

Gaps in content

Misalignment between the Developer Portal and YouTube content

Project 2: AI Model

This project addresses the following:

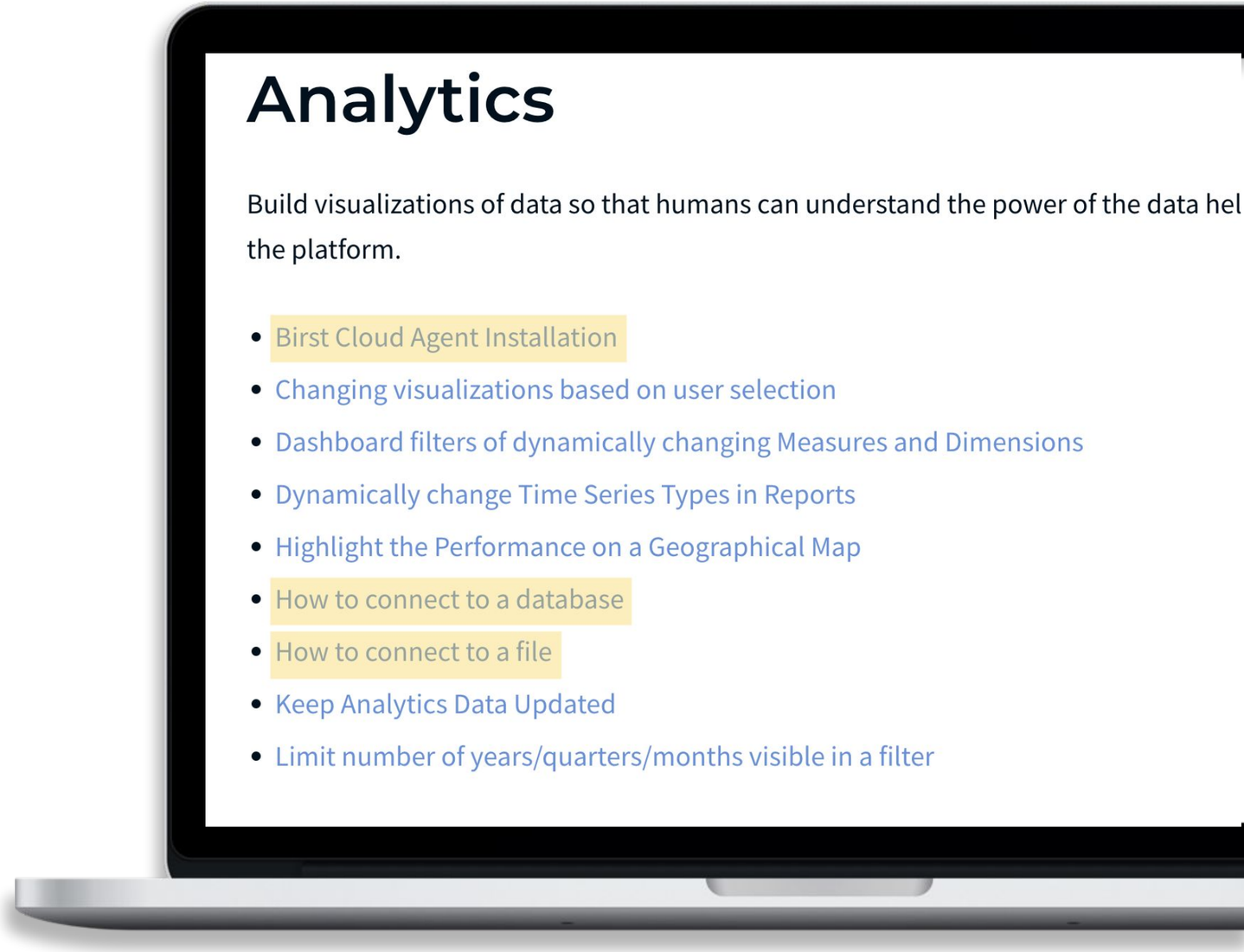
Inconsistency in content across multiple contributors

Tutorials

Task: Write tutorials to accompany YouTube videos on Developer Portal

Example:
https://developer.infor.com/tutorials/analytics/birst-cloud-agent-installation/

- 4 tutorials live
- 8 tutorials under editorial review



Meet InDevA

Infor Developer Assistant (InDevA)

A small-scale deep learning AI model that takes a written tutorial as an input and generates feedback.

- Model type: T5-small (Transformer)
- Tools used: Jupyter Notebook
- Libraries used:
  Transformers, BeautifulSoup, Datasets

InDevA's Story

Project Steps Overview

Pre-planning

Initial Goals:

- Enforce uniformity across existing and new content
- Improve the quality of all content with minimal supervision

Obstacles:

- Limited AI/ML knowledge and experience
- Difficulty with defining the rules and guidelines for the model to follow
- Time (9 weeks)
- Unclear overall purpose

The "Game" Plan

1. Research basic AI/ML concepts, natural language processing, and deep learning.
2. Build functions that will format and clean data, train the model, and generate feedback.
3. Fix bugs.
4. Repeat steps 2 and 3.

Research

AI/ML Introduction

Supervised & unsupervised learning, machine learning, deep learning, artificial neural networks, and recurrent neural networks.

AI Building Process

Data collection, data processing & cleaning, pre-training, fine-tuning, and evaluation metrics.

Natural Language Processing

Tokenization, stemming, lemmatize, sentiment analysis, topic modeling, and N-grams.

Transformers

Attention systems, encoder-based & decoder-based models, such as BERT, RoBERTa, XLNet, and T5, and architecture.

Grammarly

- A writing tool that uses AI to improve grammar and writing style
- Algorithms are trained by linguists and deep learning engineers to combine machine learning, natural language processing, and human expertise
- AI model reads each sentence and looks for ways to improve then it by correcting a verb tense, suggesting a stronger synonym, or recommending a clearer sentence structure
- Grammarly Premium features include rewrite sentences for clarity, adjust tone, catch accidental plagiarism, and generate text based on prompts

grammarly.com

Data Collection

Task: Gather the information to build the pre-training and training datasets.

- Pre-training dataset: "Components" section from Developer Portal (Analytics)
  - Used to provide the model with additional information about InforOS applications
- Training dataset: Analytic tutorials from the "Tutorials" section in the Developer Portal
- Coded functions that web-scrape text from Developer Portal pages using BeautifulSoup library

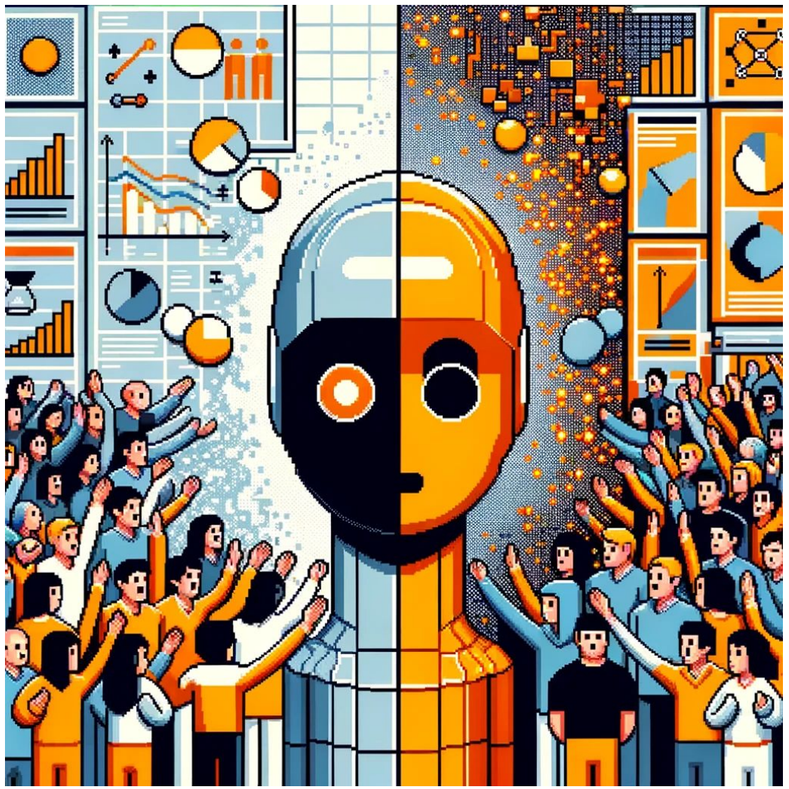
Data Pre-processing & Cleaning

ThePhoto by PhotoAuthor is licensed under CCYYSA.

Task: Format datasets as JSON strings.

- Clean the datasets by coding functions to remove duplicate and unwanted text
- Put the text in datasets and format them as JSON strings

Model Training & Evaluation



ThePhoto by PhotoAuthor is licensed under CCYYSA.

Task: Train the model, evaluate the results, and fix bugs.

- Pre-training phase: Use the pre-training dataset to train InDevA for more accurate feedback
- Training phase: Use the training dataset to train InDevA to generate feedback

Demo

InDevA's Future

Features to be included:

- Ability to notify the user (Developer Portal end) about outdated content
- Ability to adapt to new situations with less supervision
- Ability to use image and video processing to enhance pre-training and training datasets

Internship Outcomes

Contribution Motivated

- Asked for help and feedback when struggling
- Strove to realize potential with projects
- Contributed creatively

Self-Actualize

- Strove for constant improvement
- Learned about strengths and how to apply them
- Recognized need for continuous research and learning

Network

- Worked with amazing people
- Connected with other interns at Koch & Infor
- Collaborated with Infor interns on a project

Contributions to Career

- Explored AI/ML fields
- Developed & strengthened technical writing and CS skills
- Gained valuable working experience

Resources

Python libraries

- BeautifulSoup: https://beautiful-soup-4.readthedocs.io/en/latest/
- Transformers: https://huggingface.co/docs/transformers/en/index
- PyArrow: https://pypi.org/project/pyarrow/
- Datasets: https://huggingface.co/docs/datasets/en/index
- NLTK: https://www.nltk.org/
- Regular expressions (RE): https://docs.python.org/3/library/re.html
- PyTorch: https://pytorch.org/docs/stable/library.html

Thank you!

Kevin Cole

Khalid Marbou & Platform Product Strategists Team

Max Fisher & Team404

Yasmin Barnes

Sara Provencal

Remote Interns Group

Bradley Stillwell

Thank you!