

POP: Process mining On Python

이승훈, 임정은, 임지택, somebody

이름 공모 받습니다....

- 제가 이름 잘 만드는 것에 상당히 집착합니다....
 - POP(피오피): Process mining On Python
 - PMP(피엠피): Process Mining on Python
 - POPY(포파이): Process mining On PYthon
 - PROPY(프로파이): PROcess mining on Python
 - PAOP(파업): Process Analytics On Python
 - PASON(파손): Process AnalyticS On pythoN
 - 추천받아요.

이(판)걸 왜때문에 합니까?

- 솔직히, 잘 모르겠습니다 하하.
- 굳이 이유를 만들어 봅니다.
 - **ProM X같아여.** 느리고, 좋은 라이브러리와 연결하는 것도 어렵고, GUI도 별로고, 이름도 구리고(사실 POP도 그닥..), 플러그인이 많긴 한데 다 쓸데없는 것들만 잔뜩 붙어있어요.
 - 물론 java쪽에도 deeplearning4j 라는 라이브러리가 있지만 자바 싫어여
 - **프로세스 마이닝 분야의 알고리즘을 바닥부터** 하나하나 코딩하다 보면 재밌지 않을까요?
 - 연구 아이디어가 샘솟는다거나(설마)
 - **PKM, AIM lab 모두 프로세스를 연구하는데, 함께 뭐라도, 해보면 좋지 않을까요?**
 - **오픈소스 프로젝트를 하나 진행해서 경험을 쌓고 싶었는데 적당할 것 같음**

이미 진행된 비슷한 게 없나요?

- 사실 우리가 재밌으려고 하는 거니까 있든 말든 상관없지만 그래도.
- **Open source project 기준**
 - 구글에서 검색해 봤을 때, 몇 개 나오긴 하는데, 대부분 얇은 수준의 구현 (alpha algorithm 정도만 구현)
 - Python2 기반으로 구현한 쓸만한 뭔가 있었는데, 기억 안 남 (마지막 commit이 오래됨)
- **책**
 - A Primer on Process Mining
 - 읽어 봤는데, 별거 없음. 파이썬 초보자가 쓴 느낌
 - <http://www.springer.com/kr/book/9783319564265>

왜 없을까요?

- 프로세스 마이닝 자체가 별로 유명하지 않거나, 필요없다고 생각해서 아무도 만들지 않았거나.
- ProM이 이 분야에서 도미넌트한 툴이라서, 시작할 때 다들 java로 만들고
- ProM, Disco 등을 보면 알 수 있듯이 프로세스 마이닝을 원하는 클라이언트들은 대부분 not hard-coding, just soft-coding(GUI)을 더 좋아함
 - Python으로 코딩하면, 어쨌든 라인바이라인으로 짜야 하는데, 이게 안 맞을 수도.
- 뭐 어떤 이유든, 아무튼, 상관없습니다. 그냥 우리끼리 재밌게 하면 됩니다.

앞으로 무엇을 발표하면 될까요?

- **프로세스 마이닝에 존재하는(혹은 필요한) 알고리즘, ProM에 있는 플러그인 등을 구현**
 - “여러분, 제가 드디어 Alpha algorithm을 구현해봤습니다 하하하” 라거나.
 - “프로세스, 리소스 등을 클래스로 만들어봤습니다” 라거나.
 - “아직 구현은 못했지만, 이 플러그인을 만들어보려구요”라고 있는 논문 간단히 소개
 - “제가 지금 이것 만들어보려고 해요” 소개
- **남이 짠 코드 예쁘게 만들기**(훈수가 제일 좋아요)
 - 클래스를 이렇게 구성하면 더 좋지 않을까요?
 - 요렇게 바꾸었더니, 속도가 빨라졌어요!
 - 이렇게 만드는 게, 확장성 측면에서 좋지 않을까요?
- **프로세스 마이닝 측면에서 (어쩌면)도움이 될 것 같은 라이브러리 소개 및 유즈케이스**
 - Numpy: 매트릭스 연산이 좀 빨라요(ex: built-in sum보다 numpy.sum이 80배 이상 빠름)
 - Pandas: 엑셀 가져오고 처리할 때 편해요
 - Sklearn: trace clustering 등을 할 수 있지 않을까요??...
 - Matplotlib: 그림을 그리기 좋아요
 - Networkx: organizational mining 하기 좋아요

최소한의 rule

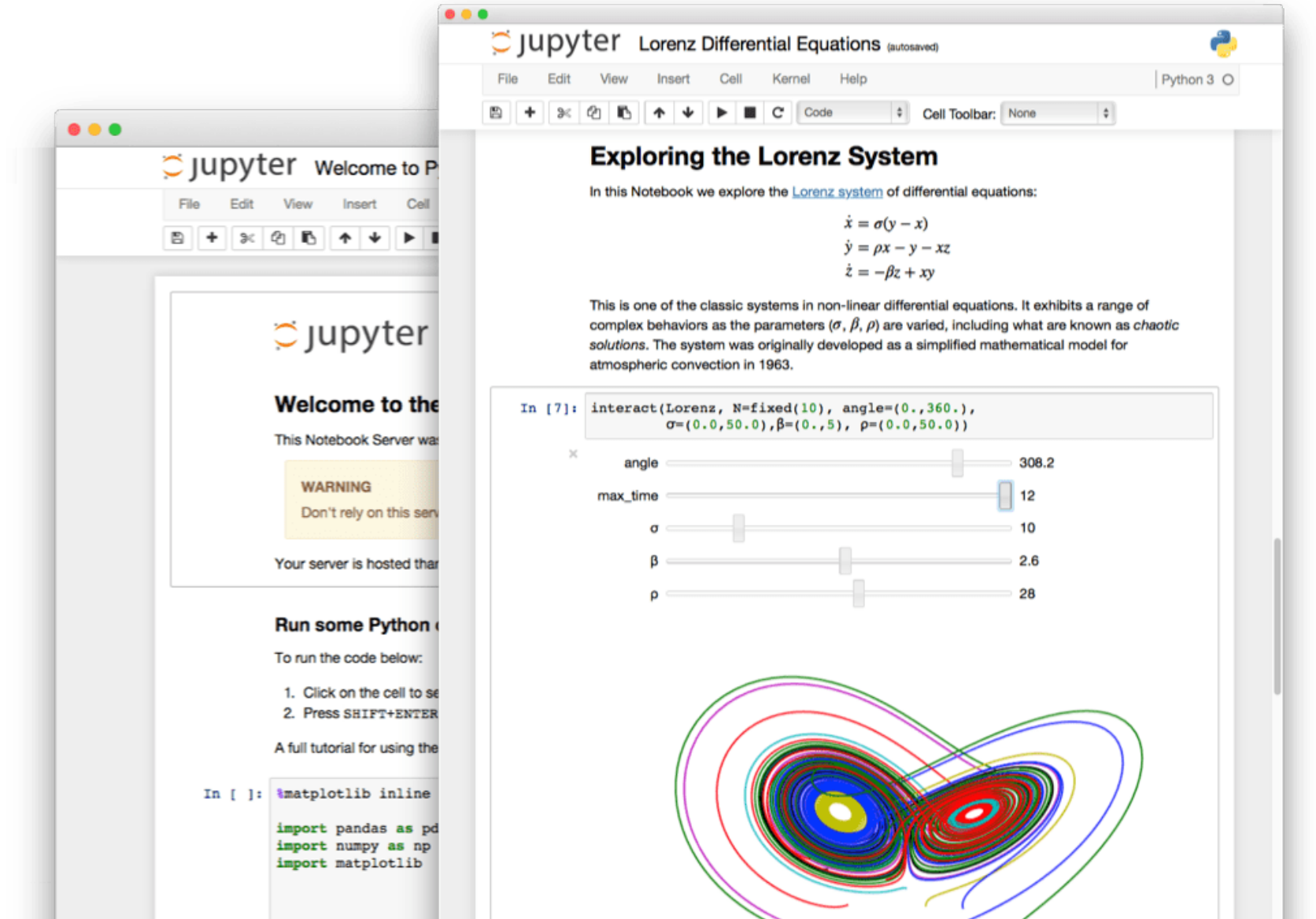
- 우선, 질문. 각자 개발환경을 뭘로 쓰시나요?
 - Pycharm? Atom? Vs code? Spyder? Jupyter notebook? Vim????

최소한의 rule

- 우선, 질문. 각자 개발환경을 뭘로 쓰시나요?
 - Pycharm? Atom? Vs code? Spyder? Jupyter notebook? Vim????
- Jupyter notebook 쓰시다. 안 써봤으면 이번에 한 번 써봅시다.
 - 코드 공유하는 측면에서 좋아요. 자세한 설명은 다음 슬라이드 스크린샷으로 대체합니다.
 - 모든 노트북의 첫 셀에는 docstring을 마크다운으로 작성합니다.
 - Jupyter notebook의 확장자는 .ipynb인데, .py로 쉽게 변환할 수 있어요.
 - (마크다운으로 설명 => 코드) 를 반복하여 작성하여 코드 공유를 편하게 합니다.
 - 잘 모르면 일단 `pip install jupyter` or `conda install jupyter`
- 일단 Dropbox로 파일 공유합니다.
 - 그냥 쓰시다
- 코드공유를 위해 Git과 github을 쓸까? 고민해봤지만.
 - 나중에 쓰시다. 지금은 쉽게 시작하는 게 중요한 것 같아요.

Jupyter notebook img

- 설명 없습니다 호호



Veeeeeeeeery far future

- Wil van der aalst를 프로세스 마이닝 분야에서 몰아냅니다
- BPI challenge 2018 winner가 됩니다.
 - 이공십팔 어감이 좋아요
- 책을 만들고 저작권 부자가 됩니다.
 - Primer on Process mining 보다는 잘할 수 있음, 이거 XX임.
 - Jupyter notebook으로 만들면, markdown 파일로 변환이 쉽고, 나중에 pdf 형식으로 만들기 쉬워요 (물론 안 해봄)
- 오픈소스 프로젝트 기여로 reputation을 획득합니다
 - "이새끼들아, 내가 왕년에 이런 것도 했어"를 할 수 있습니다
- 학과 내 동아리로 만들고 외주를 받아 돈을 벌니다

마지막.

- **한 달 동안 빠시게 하고 안할 바에야 1년 동안 천천히 합시다.**
 - 지속가능한 취미 생활 좋아요.
- **그리고 존대말합시다.**
 - 어차피 서로 존중안할거, 존중하는 척이라도.
- **서로 바쁜 와중에 하는 것이므로 시간을 지킵시다.**
 - 만약, 안될 경우 미리미리 공지하여 일정 조율
- **관심있는 사람 있으면 좀 더 데려와도 좋을 것 같습니다.**
 - 이거 어차피 취미 생활입니다
- **발표 순서를 정합시다 (승훈 => 정은 => 지택)**
 - Process Mining 수업 때 만든 플러그인 소개하는 식으로 해도 좋을 것 같습니다.