

# NLG in Practice

Yaji Sripada

# Scope

- Limited to data-to-text NLG
- Only high level ideas here, details in the Open labs
  - Discussions in the open labs seem to work!

# Topics Already Discussed in the lab

- What is corpus analysis? And why is it important?
- Corpus analysis vs corpus writing
- NLG vs Templates
- Single Pipeline vs multiple pipelines
- What is the task of microplanning?
- Why understanding microplanning is hard?
- Why plan sentences first?
- Why not plan documents first?
- SimpleNLG's way of building/constructing sentences
- Why NLG community is obsessed with REG?

# Talk Structure

- Big Ideas
  - Challenges
  - Frustration
  - Opportunity
- NLG and Business Process
- Conclusion

Big Ideas

# Challenge 1

- Making clients understand the concept of NLG is a challenge
- Perceptions vary widely/wildly
  - Overly optimistic or overly pessimistic
  - Templated solutions create a mental block

# Challenge 2

- Finding a killer use case is a challenge
- Use cases promoted during early client engagement often fail

# Challenge 3

- Presenting a good case for added value of NLG is a challenge
- Industry needs NLG case studies with proven track record of ROI



# Challenge 4

- Integrating NLG to client IT environment is the biggest challenge
- Increasing number of organizations have complex stack of data related technologies
  - Internally developed
  - Third party
- It is never easy to find simple solutions to integration problems
  - E.g. MySQL vs SQLite in the Open labs

# Frustration

- NLG is a black box
  - Little appreciation of NLG related innovations
  - Contrast this with interest on innovations in predictive analytics
  - Templates vs NLG is the only technical issue of merit
  - Document planning is also interesting – but there are no interesting and generic solutions here (other than Schemas)

# Opportunity

- Unprecedented demand for NLG
  - The regular supply of start-ups proves this

# NLG and Business Process

# NLG

- Science
- Technology
- Methodology

# Business Process

- Marketing and Sales
  - Science heavy!
- Development
  - Methodology heavy!!
- Deployment, operation and maintenance
  - Technology heavy!!!
- Quality Assurance
  - ???

# Marketing

- How to market NLG?
  - Before and after NLG
- What specific problem does NLG solve?
- Most nasty questions about NLG often rise in this context
- Competition
  - NLG vs templates
  - NLG vs InfoVis
  - NLG1 vs NLG2

# Development

- How to optimize methodology for lowering development costs?
- A few ideas, nothing tested
  - Language resources and development tools
  - Skilled workforce



# Deployment, operation and maintenance

- Highly modular and highly configurable technology is the ideal
- At the same time configuration should be easy and intuitive!

# Quality Assurance

- Manual proof-reading
- Tool driven quality checkers

# NLG Skills in practice

- Single most important skill
  - Ability to analyse corpus rapidly
  - Define user stories that reflect the insights from the corpus analysis
- Of course five star programmers are always in great demand!
- Data science knowledge is a big bonus!!

# Conclusion

- Several non-NLG challenges to overcome
- Unprecedented demand for NLG
- Science of NLG is important in practice
  - May not be at expected stages of business process
- Code monkeys are not what industry needs
- These are early days for NLG in practice, watch the space for big and sudden shifts in practice!