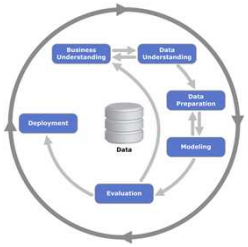
SAS Institute

Natural Language Processing

(Semantics, Morphology, Syntax, Linguistics)

Richard Frederick, PMP (rfrederick.pmp@gmail.com; 214-755-7035 talk or text; www.linkedin.com/in/rfrederick)



CBDA CERTIFIED IIBA

ISO International Organization for Standardization

ANSI American National Standards Institute

sas

OMG OBJECT MANAGEMENT GROUP

INCOSE International Council on Systems Engineering

IEEE

PMI Project Management Institute

Scaled Agile Framework

IIBA International Institute of Business Analysis

UNIFIED MODELING LANGUAGE

W3COMPUTING A DEVELOPER REFERENCE WEBSITE

Requirements Development Management (RDM) for

Business Data Analytics

(The Language of Data)

Richard Frederick, PMP (rfrederick.pmp@gmail.com; 214-755-7035 talk or text; www.linkedin.com/in/rfrederick)

Requirements Agenda

Business Data Analytics "The Language of Data"

01-FOUNDATION

- **LAB-Introductions (Name, Job Title, Objectives)**
- Dashboards and Robots (Data Mining & Machine Learning)
- Requirements and Testing (Four Quadrants)
- What are Structured Language Requirements? (Structured English and Structured Query Language)
- Why Should You Care? (Primary Source of Project Problems)
- How Do They Work? (Discreet Intellectual Property Inventory)
- Types of Requirements (Product, Project, DATA)
- Natural Language Processing (Morphology, Semantics, Syntax and Linguistics)
- OMG-SBVR (Semantics of Business Vocabulary & Rules)
- IEEE-EARS (Easy Approach to Requirements Syntax)
- INCOSE (Rules for Writing Requirements) & QVscribe
- Waterfall and Agile (Assembly Methods)

02-ELICIT

- **LAB-Vision/Scope (Seek to Understand)**

Elicitation Techniques:

- Document Analysis (Low Hanging Fruit)
- Interface Analysis (Navigation & Functionality)
- Benchmarking (Actual Data)
- Brainstorming (Every Idea is a Good Idea until it becomes a Bad Idea)
- Prototyping (Minimum Viable Product)
- Reverse Engineering (Begin with the End in Mind)
- Interview (Thinking Questions)
- Workshop (Group Interviews)
- Observation (What do you See?)
- Survey Questionnaire (Paper equals proof)

03-ANALYZE

- What are Models? (Pictures of Language)
- **LAB-The Language of Modeling (GIVEN pre WHEN process THEN output-result)**
- Types of Models (Context-Structure, Usage, Data Behavior, Process Flow)
- Context-Structure (Vision, Roadmap, Scope WBS)
- Usage (EPIC, UseCase, UserStory, Feature)
- Data Behavior (ERD, JOIN-Denormalization, Star Schema, Dimensional OLAP, Dashboard, Intelligence)
- Data Behavior (Data Dictionary, DataFlow, Data Structure Instance, Data Element Attribute, Data Store)
- Data Behavior (Process Logic, Business Rules)
- Process Flow (Swimlane)

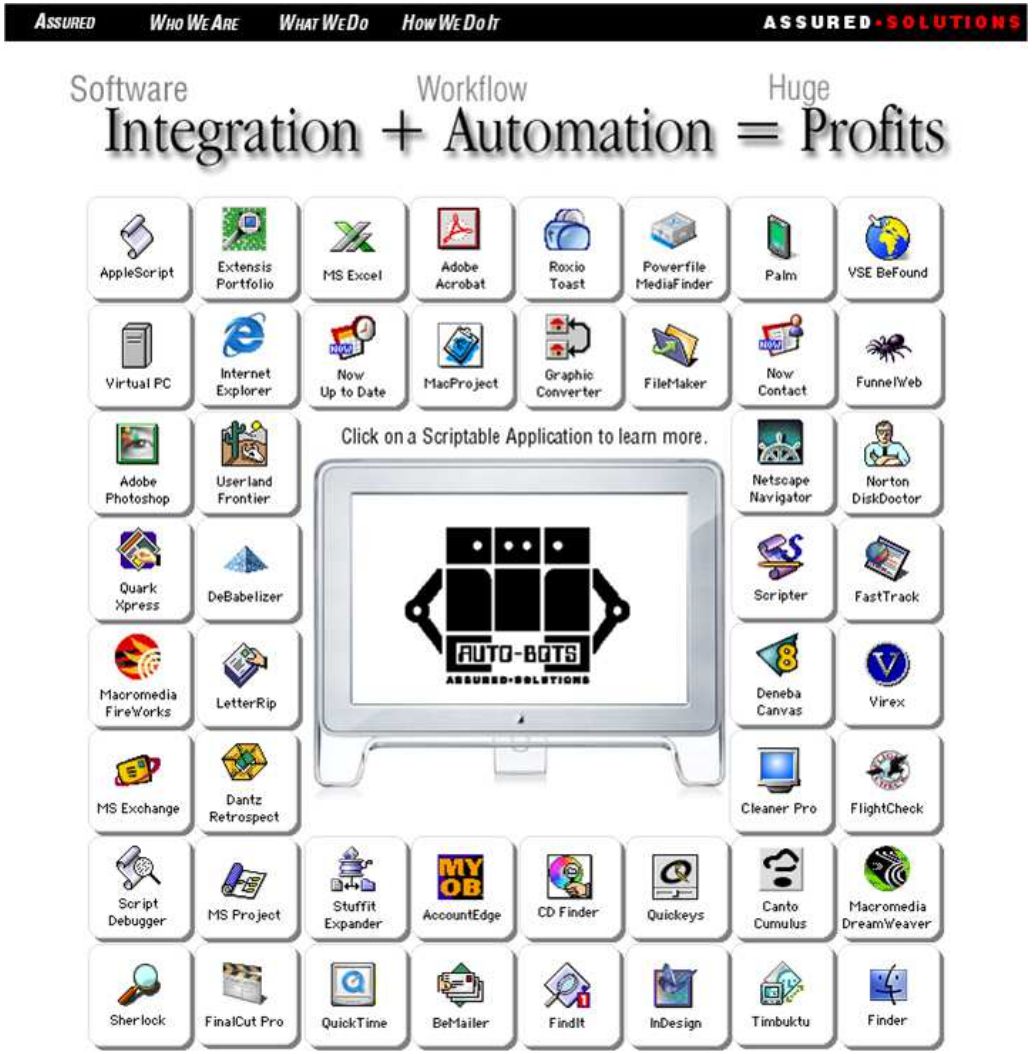
04-DOCUMENT

- Categorization, Organization, Documentation, Integration, Automation
- Making Documents Easy to Read (Fonts & Navigation)
- Document Types (BRD, TRD)
- **LAB-Business Requirement Document (Concept of Operation)**
- Technical Requirement Document (System Specification)

05-VALIDATE

- Validation thru Triangulation (Prep Drills)
- Traceability (Project Unique Identifier)
- Requirements Baseline (ROM Estimate, Planning Estimate, Definitive Estimate)
- **LAB-Estimating Story Points (Complexity and Risk)**
- Lessons Learned (Course Wrap-Up)

Assured Solutions

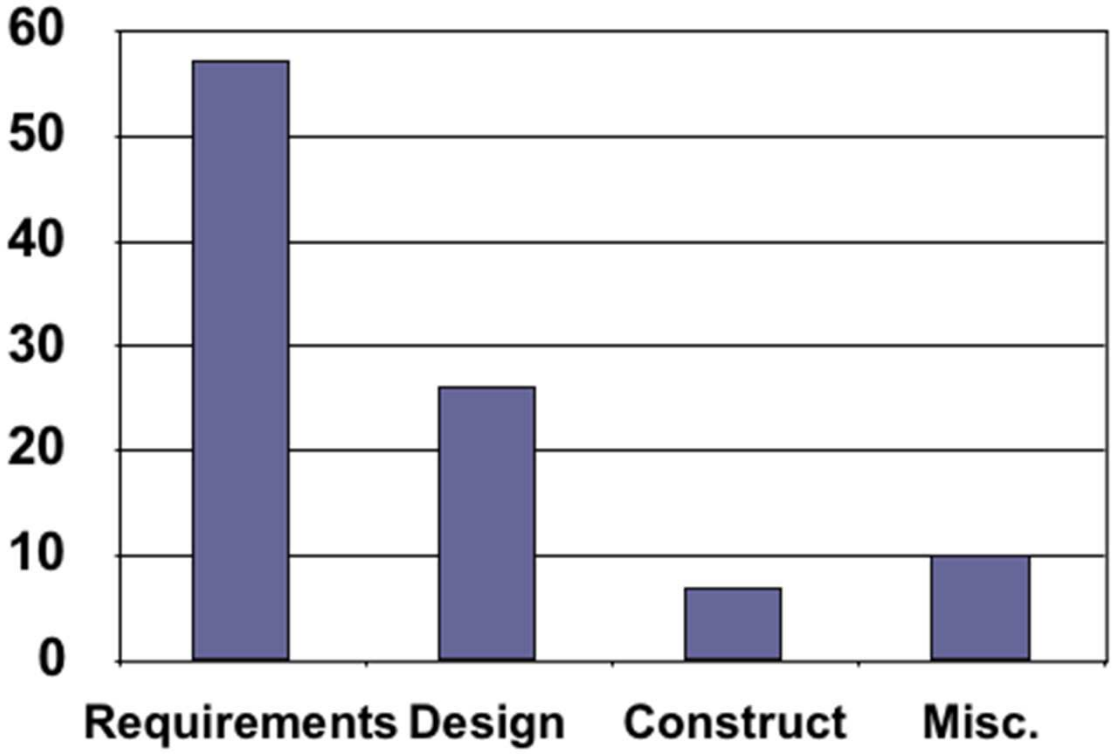


Why Should You Care About Requirements?

(Primary Source of Project Problems)

Requirements:

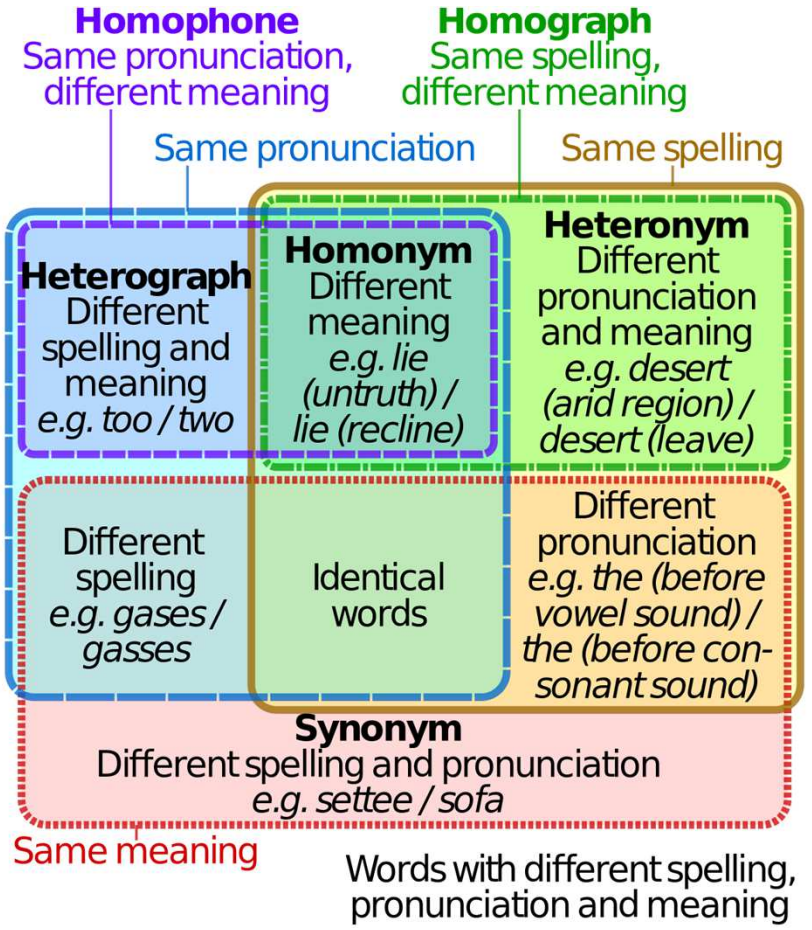
The Primary Source of Errors found at Acceptance Testing



Percent source of errors found at acceptance testing - James Martin

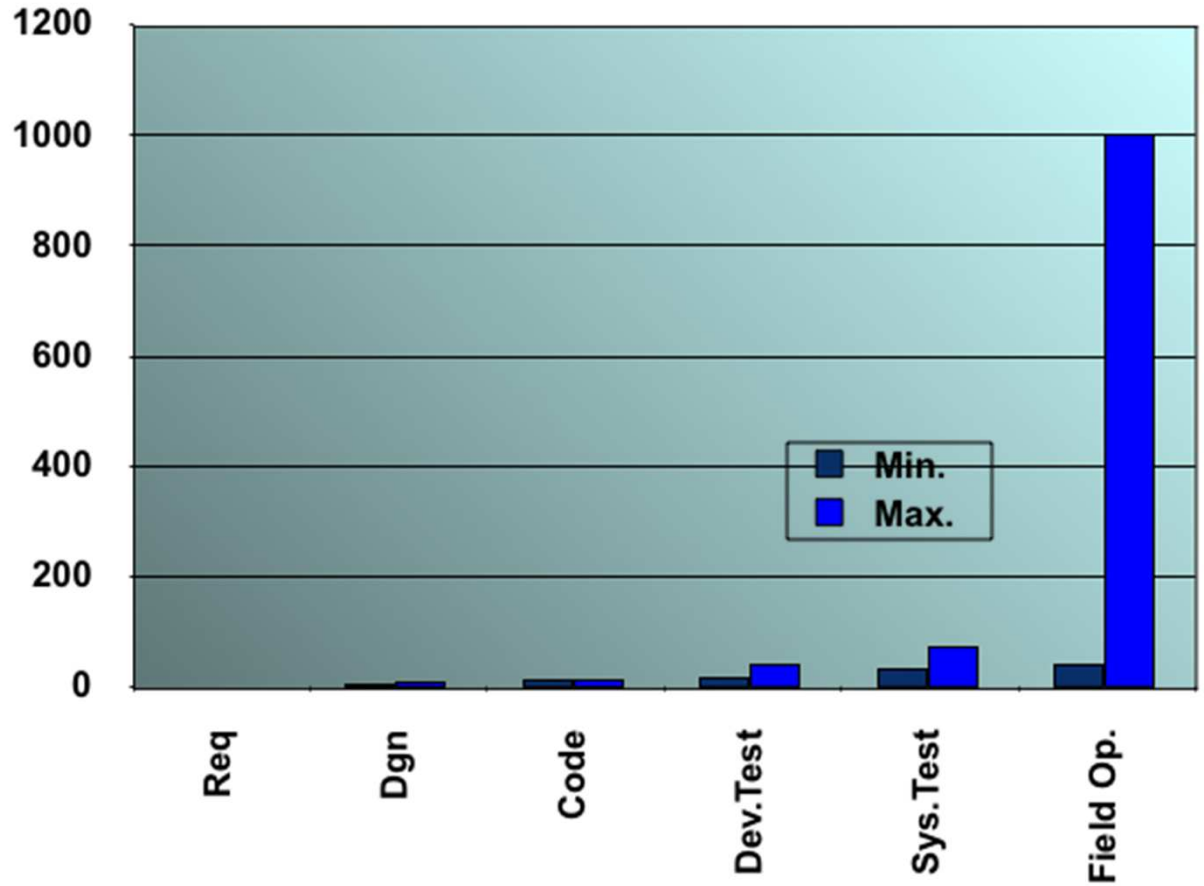
Requirements:
The Problem with Natural Language

- ***Natural Language "Nuanced Ambiguity" in Documentation introduces avoidable Business Risks***
  - **These business risks are very costly, even catastrophic, to the organization involved**
- The Challenge: *"To Elicit and Abstract Structured Language from Stakeholder Business Requirements"*
- Synonyms, Homonyms, Heterographs, Homophones, Heteronyms and Homographs



Requirements:

The Cost to Fix Defects at the Requirements Stage is Zero



SAS Institute

Natural Language Processing

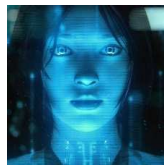
(Semantics, Morphology, Syntax, Linguistics)

Overview "The Building Blocks of Language"

- **Semantics:** *relates to the meaning of words and statements*
- **Morphology:** *the study of the structure and form of words*
- **Syntax:** *the study of how words and phrases form sentences*
- **Linguistics:** *the structured science of text*





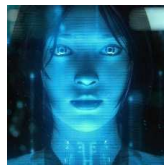
Semantics:

Relates to the meaning of words and statements

- The meaning of it all
- Approaches to meaning
  - Semantic networks
  - Deductive logic
  - Rule-based systems
- **Useful for classification**





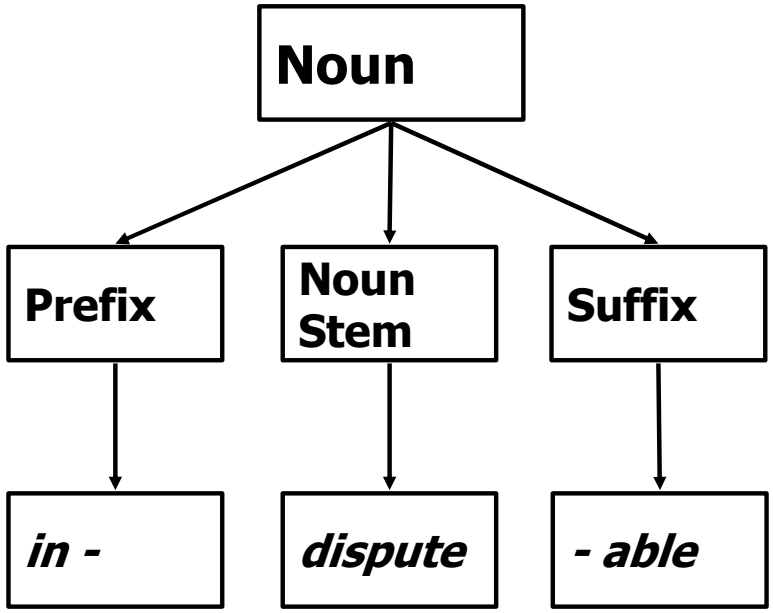



OMG-SBVR: *Semantics of Business Vocabulary & Rules*

Morphology:

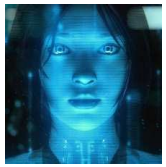
the study of the structure and form of words

- Understanding words
  - Stems
  - Affixes
    - Prefix
    - Suffix
  - Inflectional elements
- Reducing complexity of analysis
- Reduces complexity of representation
- Supports text mining





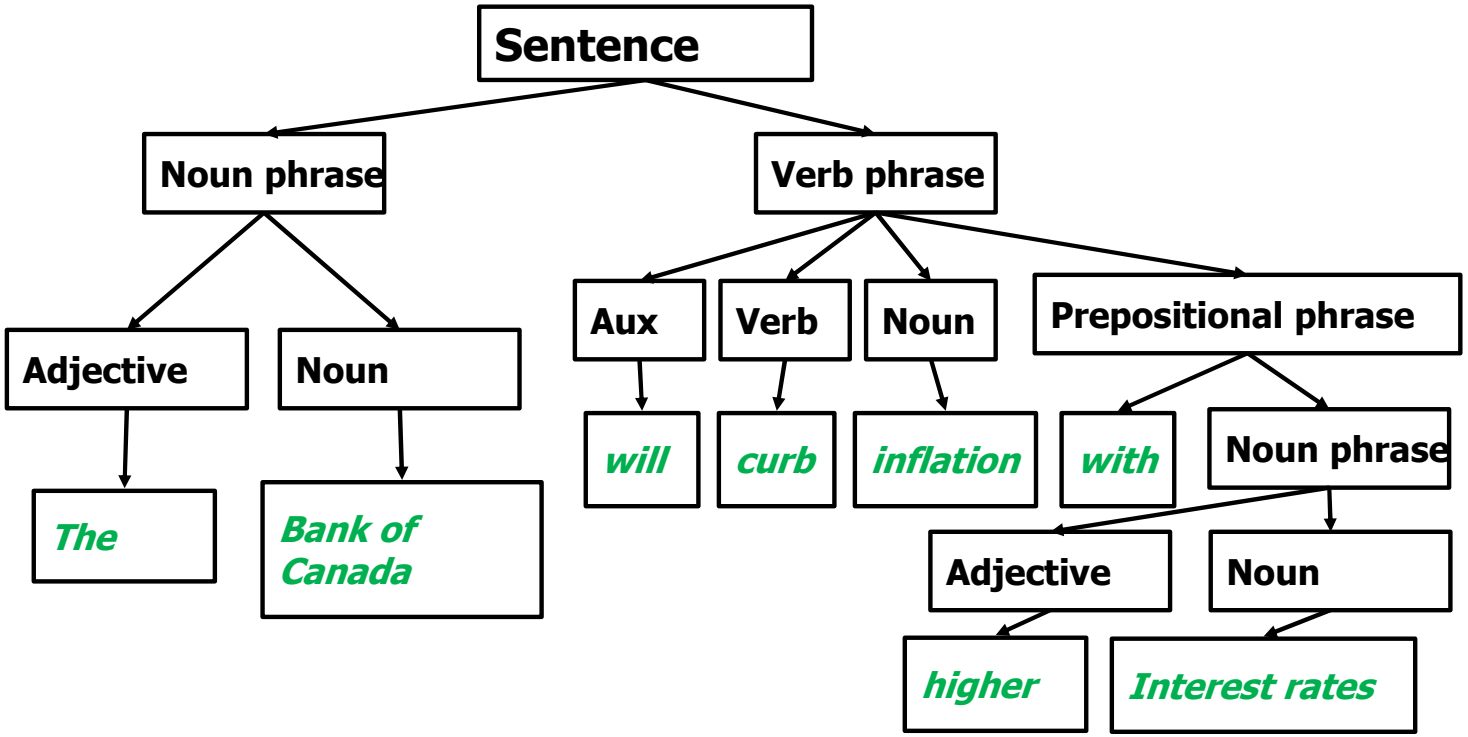


Syntax:

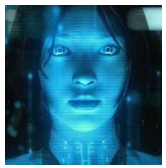
the study of how words and phrases form sentences

"The Bank of Canada will curb inflation with higher interest rates"







IEEE-EARS: *Easy Approach to Requirements Syntax*

Linguistics:

The Structured Science of Text

Text is unstructured, ambiguous, and language dependent.

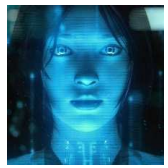
The Linguistic Approach:

- Does not treat a document as a "collection of words"
- **Removes ambiguity by extracting structured concepts**

Concepts are the **DNA** of text.





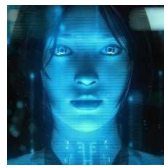



Linguistics: Part-of-Speech Tagging

A: Adjective	B: Adverb	C: Preposition
D: Determiner	N: Noun	V: Verb
O: Coordination	P: Participle	S: Stop Word



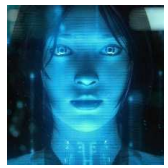


Linguistics: Concept Extraction





Step 1: Part-of-Speech Tagging

Using	a	tool	like	LexiQuest	Mine	is	a	great
V	P	N	A	N	N	V	P	A

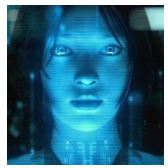
idea	for	any	organization	that	is	interested	in	maintaining
N	P	A	N	P	V	V	P	V

information	on	competitive	intelligence.
N	P	N	N

Linguistics: Concept Extraction





Step 2: Matching to Known Patterns

This:

V P N A N N V P A N PA N P V V P V N PN N

Looks Most Like:

N C D N N

(32 Known patterns for English)

Linguistics: Concept Extraction

Step 3: Reading the Sentence and Extracting the Concept

> *"Using a tool like LexiQuest Mine is a great idea for any organization that is interested in maintaining information on competitive intelligence."*

And extracts the concept:

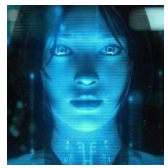
Competitive Intelligence

Concepts are:

- Noun based
- Can be longer than one word





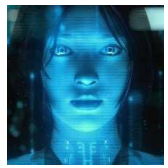



In Summary "Train the Trainer"

- **Semantics:** *relates to the meaning of words and statements*
- **Morphology:** *the study of the structure and form of words*
- **Syntax:** *the study of how words and phrases form sentences*
- **Linguistics:** *the structured science of text*





Let’s stay in contact:

Richard Frederick, PMP

214-755-7035 (text or talk)

Rfrederick.pmp@gmail.com

www.linkedin.com/in/rfrederick

https://meetings.hubspot.com/rfrederick-pmp