Extracting Cybersecurity related entities, terms and concepts from text.

RAVENDAR LAL

ADVISER: DR. TIM FININ

M.S. THESIS DEFENSE

EBIQUITY RESEARCH LAB

UMBC

So… what's happening here?

An Excerpt:

"Adobe has released security updates for Adobe Reader and Acrobat X (10.1.3) and earlier versions for Windows and Macintosh. These updates resolve a buffer overflow vulnerability that could lead to remote code execution……"

So… what's happening here?

An Excerpt:

"Adobe has released security updates for Adobe Reader (SOFTWARE) and Acrobat X (10.1.3) (SOFTWARE) and earlier versions (NER_Modifier) for Windows (OS) and Macintosh (OS). These updates resolve a buffer overflow (MEANS) vulnerability in AcrobatReader.exe (FILE_NAME) that could lead to remote code execution (CONSEQUENCES) ……"

Preview

I. Introduction and Problem

II. My Contribution

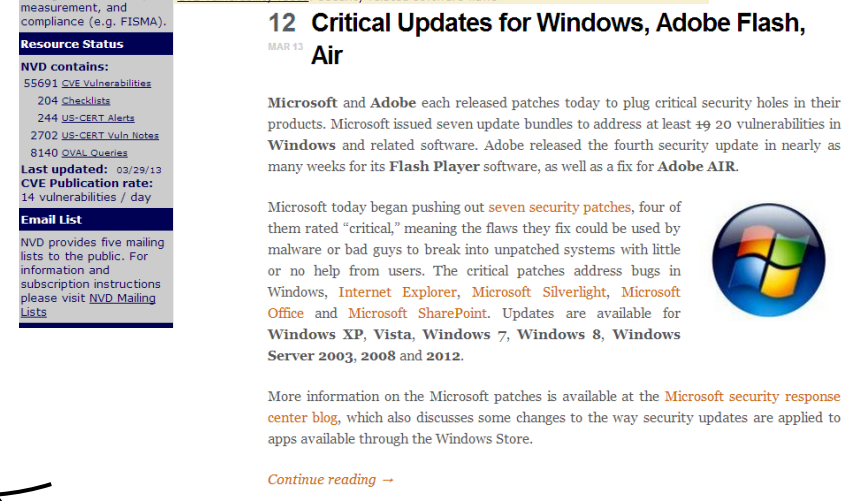
III. Results and Evaluation

IV. Future work and Conclusion

Introduction

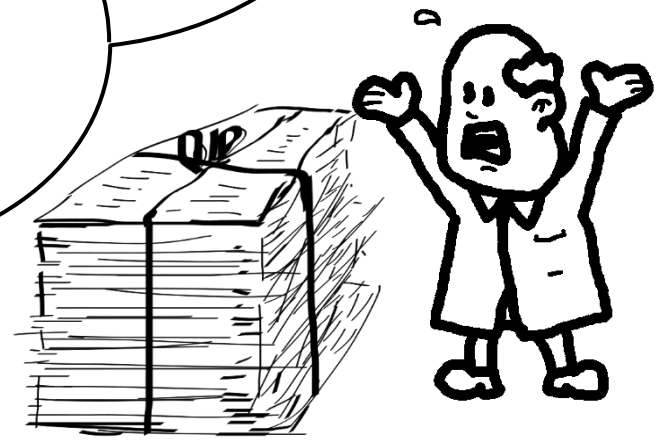
- Cyber Security Threat
- Attacks ranging Government to large internet population
- Zero Day Attacks
  - YAJ0
- Software Vulnerabilities

Too many data sources…



twitter

Web



Big Picture?

Better auto processing of:

- threats with minimal human effort
- concepts and entities from live data streams
- relations between the entities

Questions like:

Which software is being talked about?

What kind of attack they are mentioning?

Who and how it is being attacked?

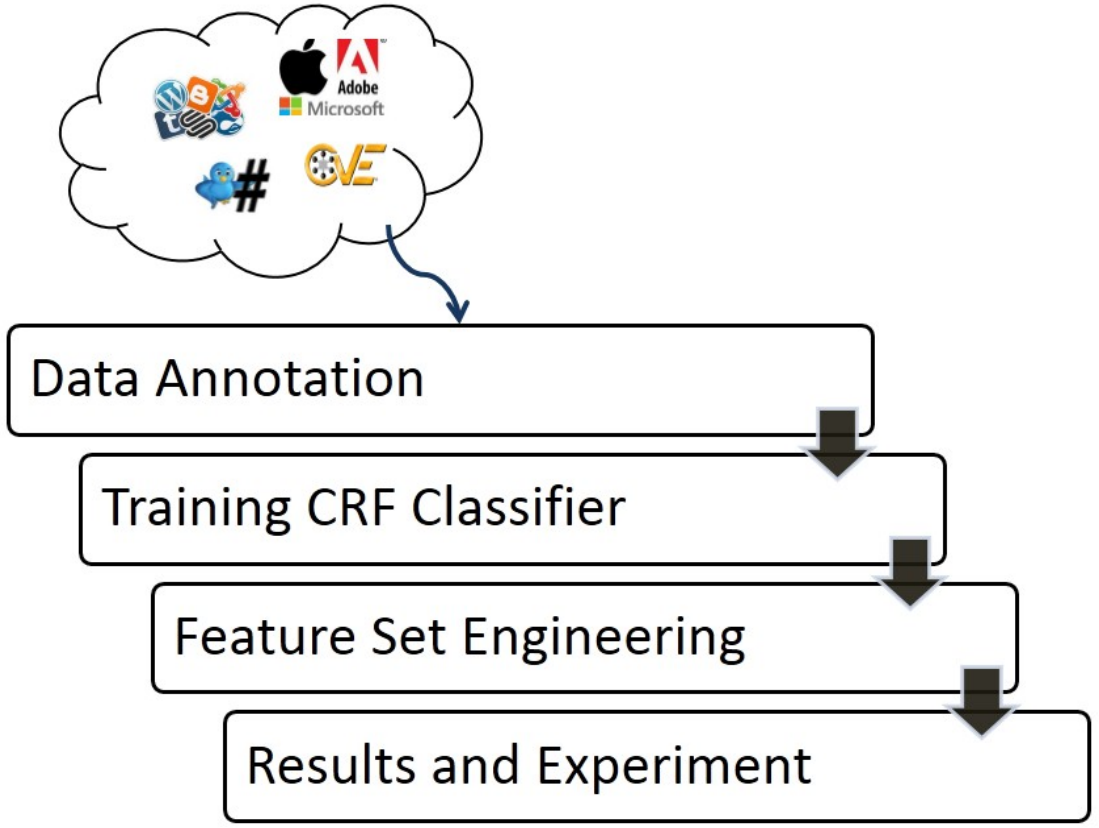
My Mission Statement

"Develop a system that can identify mentions of Cybersecurity-related entities (e.g., software applications) and concepts (e.g., buffer overflow) to assist in extracting Cybersecurity data from text."

My Contribution

- ✓ Dataset Collection - Corpus of Cybersecurity related text
- ✓ Manual Annotation of collected data
- ✓ Interface between Stanford NER and annotation tool
- ✓ Cybersecurity Entity and Concept Spotter

Training Methodology



Dataset

- Blogs (50 Various Blog posts)
  - Major contributor (Krebsonsecurity.com)
- Common Vulnerability and Enumerations (300 CVEs)
  - Best for some classes specially like Means and Consequences
- Official Security Bulletins (100 Documents)
  - Adobe
  - Microsoft

Common Vulnerabilities and Exposures (CVE)

- Maintained by MITRE (mitre.org)

- Example of a CVE Summary:

"The Firefox sandbox in Adobe Flash Player before 10.3.183.67 and 11.x before 11.6.602.171 on Windows and Mac OS X, and before 10.3.183.67 and 11.x before 11.2.202.273 on Linux, does not properly restrict privileges, which makes it easier for remote attackers to execute arbitrary code via crafted SWF content, as exploited."

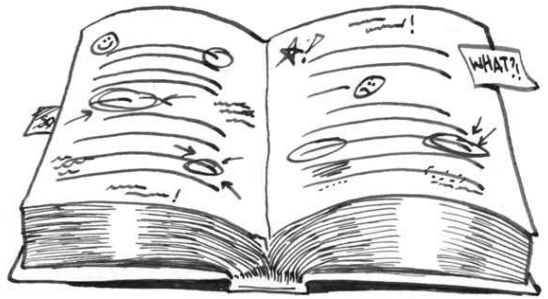
Security Bulletins

Excerpt from MS-Bulletin

"This security update resolves a privately reported vulnerability in Visual Studio Team Foundation Server. The vulnerability could allow elevation of privilege if a user clicks a specially crafted link in an email message or browses to a webpage that is used to exploit the vulnerability. In all cases, however, an attacker would have no way to force users to perform these actions. Instead, an attacker would have to convince users to visit a website, typically by getting them to click a link in an email message or Instant Messenger message that takes them to the attacker's website……"

Annotation Classes

1. Software (e.g., Microsoft .Net Framework 3.5)
   a. Operating_System (e.g., Windows XP SP2)
2. Network_Terms (e.g., Firewall)
3. Attack (From the work of of Jeffery Undercoffer)
   a. Means –Methods to conduct an attack(e.g., Buffer overflow)
   b. Consequences – final result of an attack (e.g., Denial of service)
4. File_Name (e.g., msll_32.dll)
5. Hardware (e.g., IBM Mainframe B152)
6. NER_Modifier (e.g., Acrobat Reader X **and earlier versions**)
7. Other_Technical_Terms (None of the above. E.g., HTML)

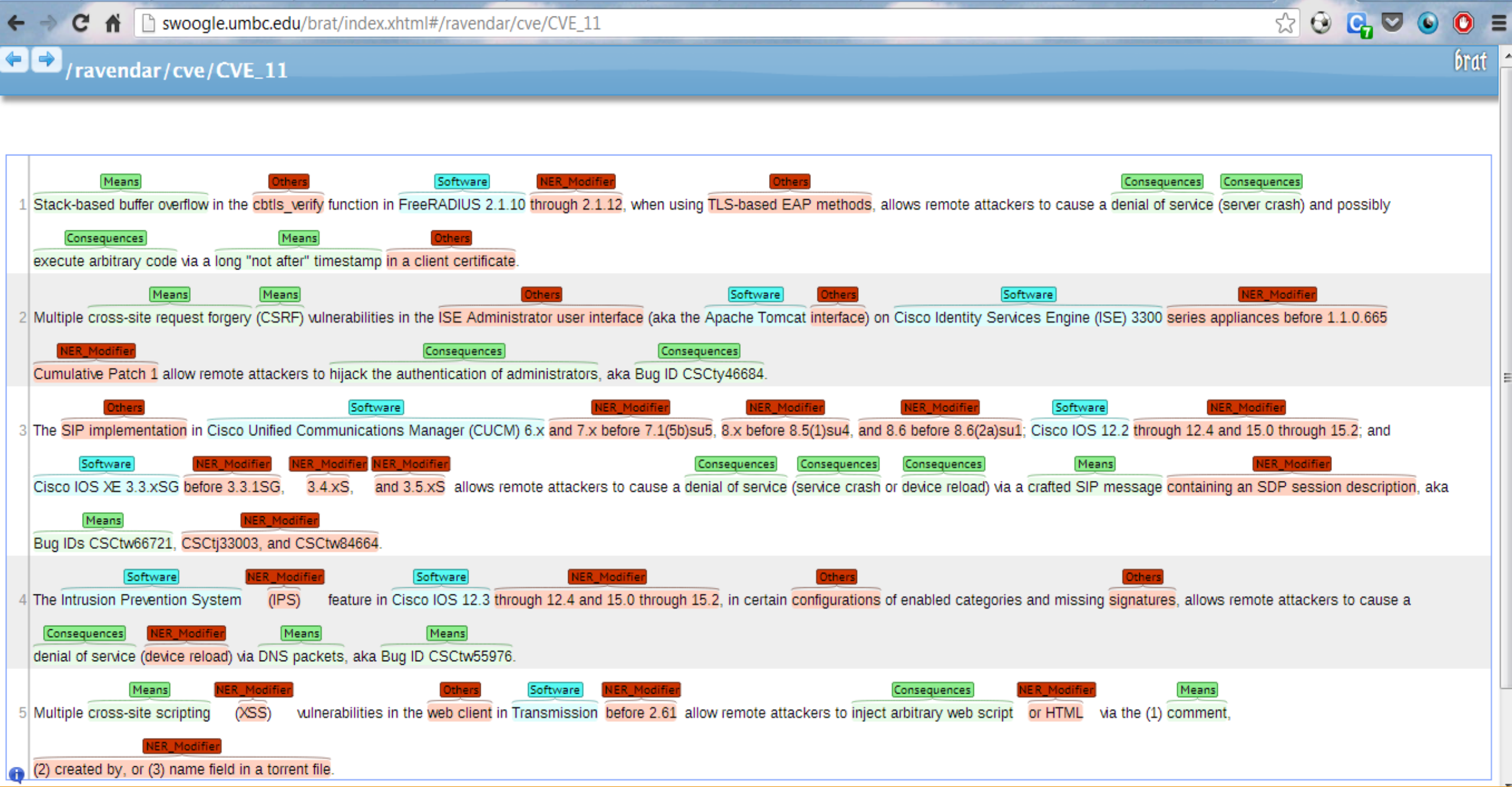


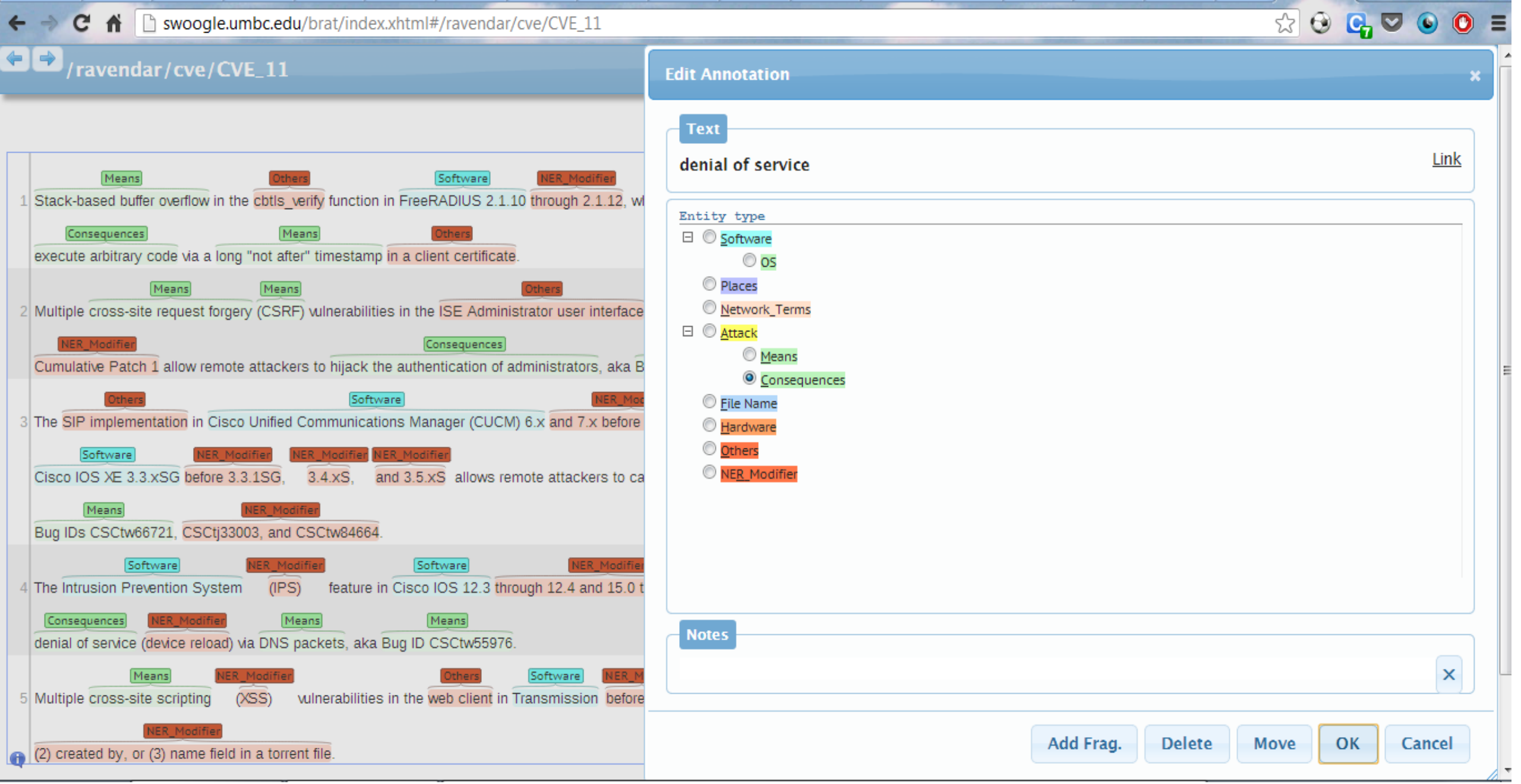
Annotation Task

- Most time consuming and painstaking part
- Asked CS grad students to help me in this regard
- 12 Annotators
- Provided training and guidelines
- Why Mturk Failed?
  - User selection
  - Domain expert knowledge required

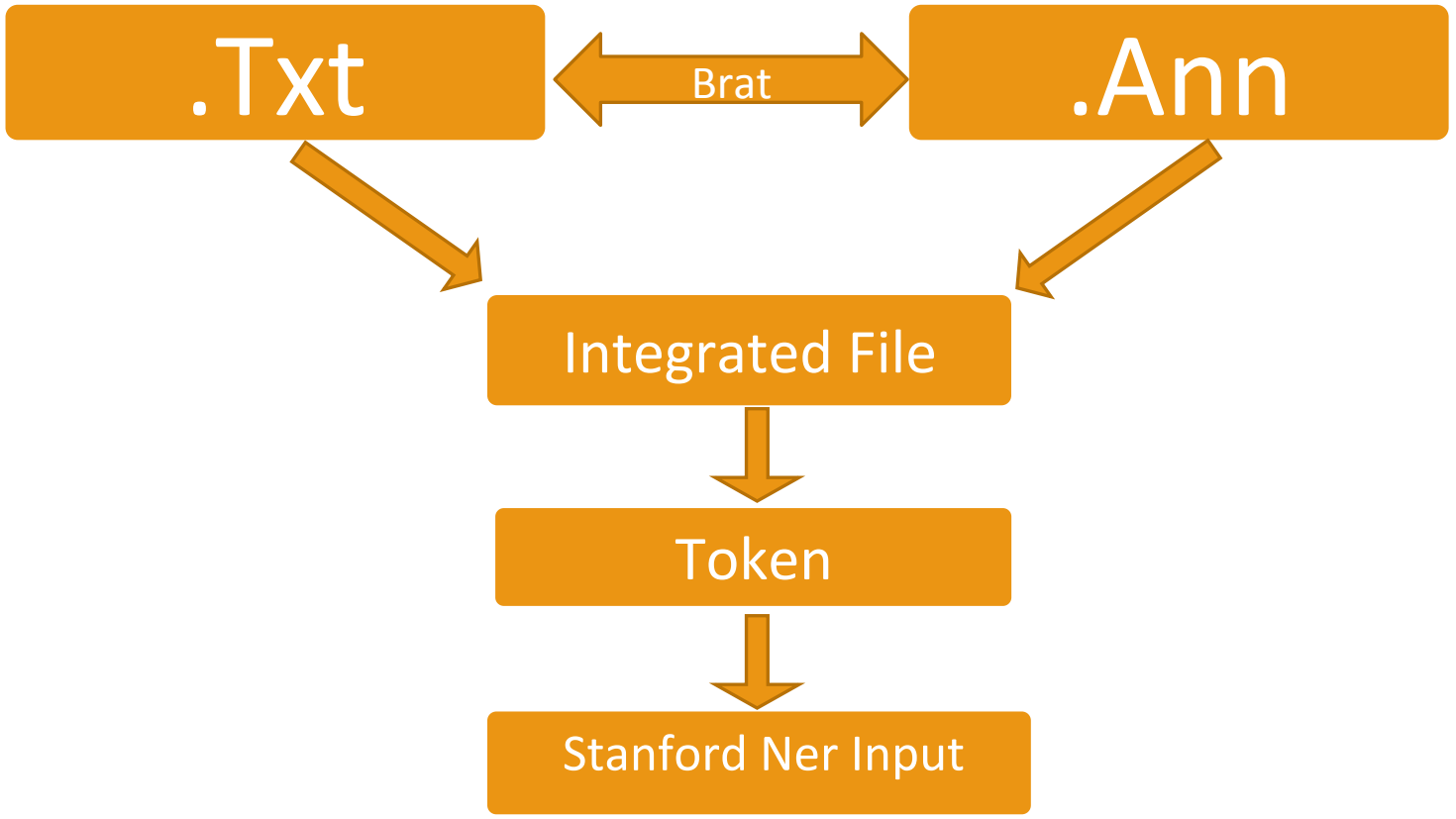
Annotation Task (Brat)

- Brat (http://brat.nlplab.org/index.html)
- Easy to setup server and data
- Currently online at: http://swoogle.umbc.edu/brat/
- Easy to annotate
  - Keyboard Shortcuts
  - Highly interactive User Interface


swoogle.umbc.edu/brat/index.xhtml#/ravendar/cve/CVE_11
/ravendar/cve/CVE_11
brat
1 Stack-based buffer overflow in the cbtls_verify function in FreeRADIUS 2.1.10 through 2.1.12, when using TLS-based EAP methods, allows remote attackers to cause a denial of service (server crash) and possibly execute arbitrary code via a long "not after" timestamp in a client certificate.
2 Multiple cross-site request forgery (CSRF) vulnerabilities in the ISE Administrator user interface (aka the Apache Tomcat interface) on Cisco Identity Services Engine (ISE) 3300 series appliances before 1.1.0.665 Cumulative Patch 1 allow remote attackers to hijack the authentication of administrators, aka Bug ID CSCty46684.
3 The SIP implementation in Cisco Unified Communications Manager (CUCM) 6.x and 7.x before 7.1(5b)su5, 8.x before 8.5(1)su4, and 8.6 before 8.6(2a)su1; Cisco IOS 12.2 through 12.4 and 15.0 through 15.2; and Cisco IOS XE 3.3.xSG before 3.3.1SG, 3.4.xS, and 3.5.xS allows remote attackers to cause a denial of service (service crash or device reload) via a crafted SIP message containing an SDP session description, aka Bug IDs CSCtw66721, CSCtj33003, and CSCtw84664.
4 The Intrusion Prevention System (IPS) feature in Cisco IOS 12.3 through 12.4 and 15.0 through 15.2, in certain configurations of enabled categories and missing signatures, allows remote attackers to cause a denial of service (device reload) via DNS packets, aka Bug ID CSCtw55976.
5 Multiple cross-site scripting (XSS) vulnerabilities in the web client in Transmission before 2.61 allow remote attackers to inject arbitrary web script or HTML via the (1) comment, (2) created by, or (3) name field in a torrent file.


swoogle.umbc.edu/brat/index.xhtml#/ravendar/cve/CVE_11
/ravendar/cve/CVE_11
1 Means Others Software NER_Modifier
Stack-based buffer overflow in the cbtls_verify function in FreeRADIUS 2.1.10 through 2.1.12, wl
Consequences Means Others
execute arbitrary code via a long "not after" timestamp in a client certificate.
2 Means Means Others
Multiple cross-site request forgery (CSRF) vulnerabilities in the ISE Administrator user interface
NER_Modifier Consequences
Cumulative Patch 1 allow remote attackers to hijack the authentication of administrators, aka B
3 Others Software NER_Mod
The SIP implementation in Cisco Unified Communications Manager (CUCM) 6.x and 7.x before
Software NER_Modifier NER_Modifier NER_Modifier
Cisco IOS XE 3.3.xSG before 3.3.1SG, 3.4.xS, and 3.5.xS allows remote attackers to ca
Means NER_Modifier
Bug IDs CSCtw66721, CSCtj33003, and CSCtw84664.
4 Software NER_Modifier Software NER_Modifier
The Intrusion Prevention System (IPS) feature in Cisco IOS 12.3 through 12.4 and 15.0 t
Consequences NER_Modifier Means Means
denial of service (device reload) via DNS packets, aka Bug ID CSCtw55976.
5 Means NER_Modifier Others Software NER_M
Multiple cross-site scripting (XSS) vulnerabilities in the web client in Transmission before
NER_Modifier
(2) created by, or (3) name field in a torrent file.
Edit Annotation
Text
denial of service
Link
Entity type
Software
OS
Places
Network_Terms
Attack
Means
Consequences
File Name
Hardware
Others
NER_Modifier
Notes
Add Frag.
Delete
Move
OK
Cancel

Annotations To Stanford NER (Interface)



Brat System

1. Txt File: This security update resolves a privately reported vulnerability in the implementation of SSL and TLS in Microsoft Windows.

2. Ann File: T1 Network_Terms 90 93 SSL

T2 Network_Terms 98 101 TLS

T3 Operating_System 105 122 Microsoft Windows

3. Int File: This security update resolves a privately reported vulnerability in the implementation of **SSL#NETWORK** and **TLS#NETWORK** in **Microsoft@Windows#OPERATINGSYSTEM**

4. Token File

This

security

update

….

SSL#NETWORK

Privately

TLS#NETWORK

Vulnerability

…..

Microsoft@Windows#OPERATINGSYSTEM

.

5. Stanford NER Input File

This O

security O

a O

privately O

……

TLS NETWORK,

Vulnerability O

in O

…..

Microsoft OPERATINGSYSTEM,

Windows OPERATINGSYSTEM,

Stanford NER

- Part of Stanford NLP Toolkit
  - Stanford NER was introduced in 2005

- Provides java based general implementation of Conditional Random Field Sequence models
- Provides well-engineered feature extractors for Named Entity Recognition
- Provides a lot of options for Feature selection

Training New Models From Stanford NER

- Features
  - edu.stanford.nlp.sequences.NERFeatureFactory
    - Easiest option: just add new features here
    - Lots of built in stuff
  - Specifying features
    - **edu.stanford.nlp.sequences.SeqClassifierFlags**
    - Stores global flags
    - Initialized from Properties file

Features

- Big Problem: Too Little Documentation but great online support

- Most Critical Part: Understanding the data
  - How entities are described?
  - what are the words that we want to extract and any special property of (or surrounding) the word?

- Initial selection is very important

Features that we used…

- useTaggySequences=true (i.e. HMM of classes)
- useNGrams=true (i.e. make substrings of word and see if there is any significant pattern that can be noticed from training data for those "Class" of an entity).
- usePrev= true (i.e. Use the pair of (previous word, class) during training data and make features on that)
- useNext = true (i.e. Use the pair of (next word, Class))
- maxNGramLeng=6 (max size means it can make smaller than that too)
- noMidNGrams=true (n-grams that only contain beginning or end of the word)
- useWordPairs = true (Features for (pw, w, c) and (w, nw, c))
- gazette= OperatingSystem.txt, Softwares.txt
- useGazettes = true
  - cleanGazette=true
  - SloppyGazette = false

Results and Evaluation

5-Fold Cross Validation Results

Strategy for training and testing Systems:

- Divided annotated data in 5 chunks (20% each)
- Trained the classifier with 5 various combinations of chunks
- Training: Testing ratio is (80:20)
- Evaluation score is Based on precision and recall model

  Precision = TP / (TP+ FP)

  Recall = TP / (TP + FN)

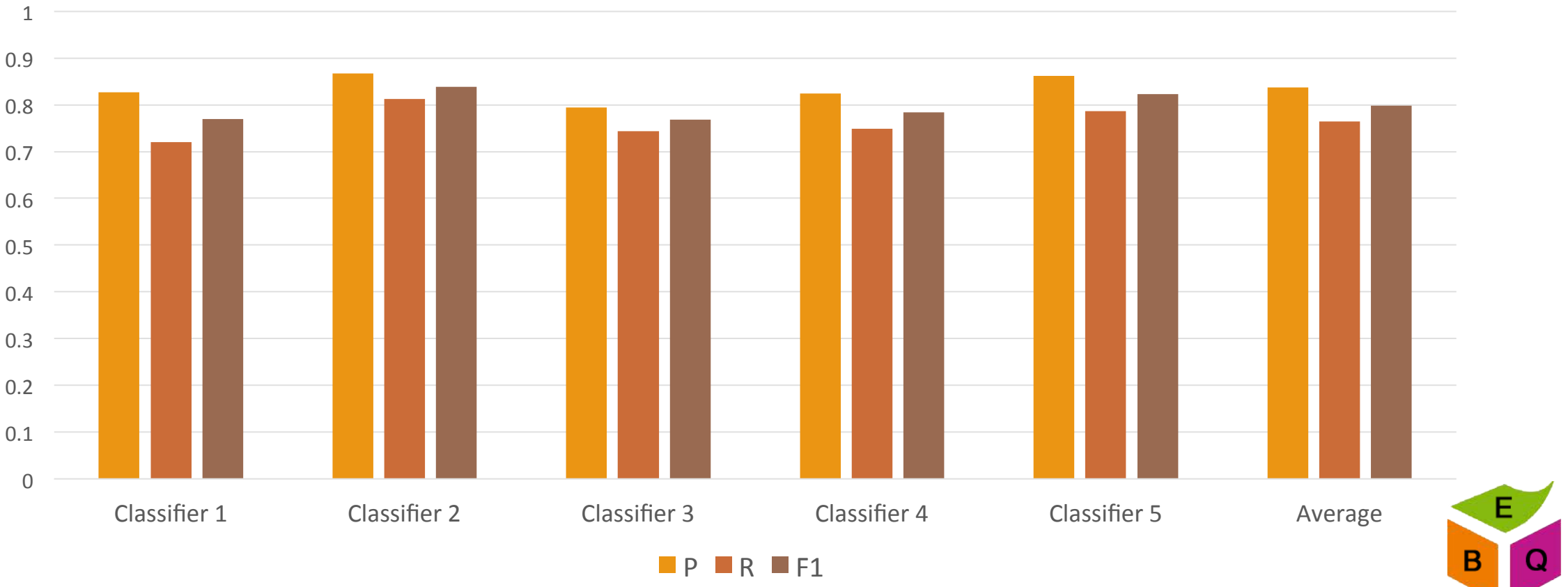
  F1 = 2 (Precision x Recall) / (Precision + Recall)

Data Corpus

- 300 CVEs
- 50 Adobe Security Bulletins
- 50 Microsoft Security Bulletins
- 50 Blog Posts
- Processing speed at around 200 words/sec

Task	Tokens	Entities
Training	30,000	3,800
Testing	9,000	1,200

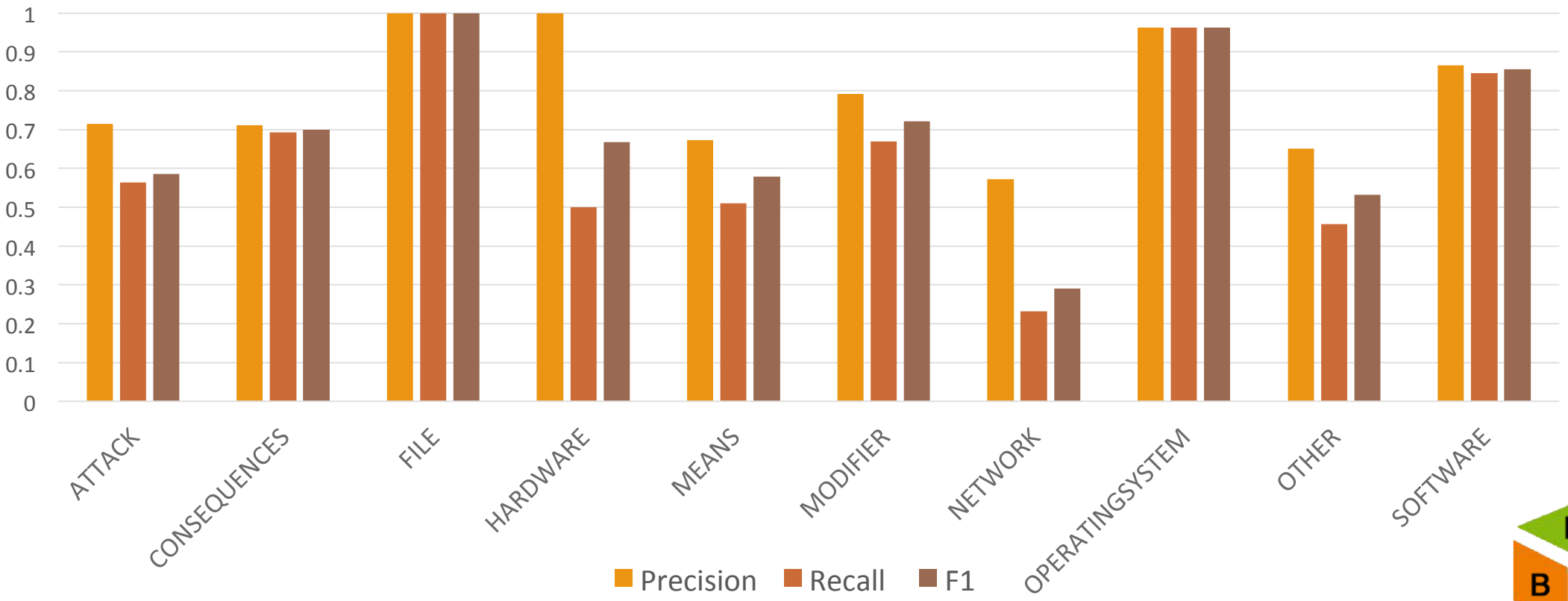
5-Fold Cross Validation Results



Data for previous graph

Classifiers	P	R	F1	TP	FP	FN
Classifier 1	0.8273	0.72	0.7699	594	124	231
Classifier 2	0.8668	0.8127	0.8389	729	112	168
Classifier 3	0.794	0.7442	0.7683	605	157	208
Classifier 4	0.8239	0.7484	0.7844	702	150	236
Classifier 5	0.8625	0.7862	0.8226	809	129	220
Average	0.836536	0.76388272	0.798560316	3439	672	1063

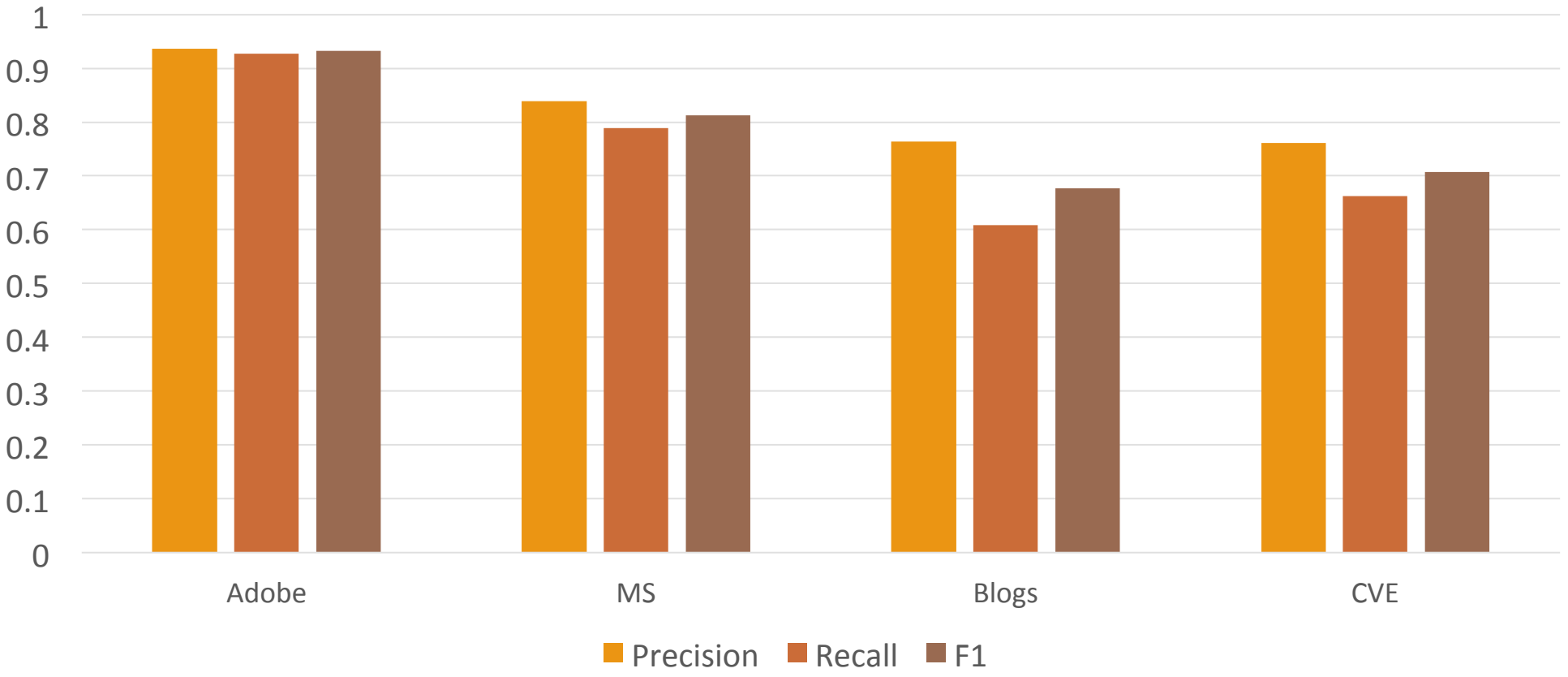
Average (Class-wise)



Average (Class-wise)

Entity	TP	FP	FN
ATTACK	30	14	27
CONSEQUENCES	299	123	135
FILE	52	0	0
HARDWARE	3	0	2
MEANS	185	94	177
MODIFIER	320	79	147
NETWORK	14	15	45
OPERATINGSYSTEM	920	34	36
OTHER	167	89	230
SOFTWARE	1449	224	268
Total	3439	672	1063

Average (Dataset-wise)



Inter-Annotator Agreement

- To see agreement on various classes between users
- Class level agreement – helps to identify those classes which are confusing
- Same pattern was found on trained classifiers
- Experimented on 10% of total dataset for main classes

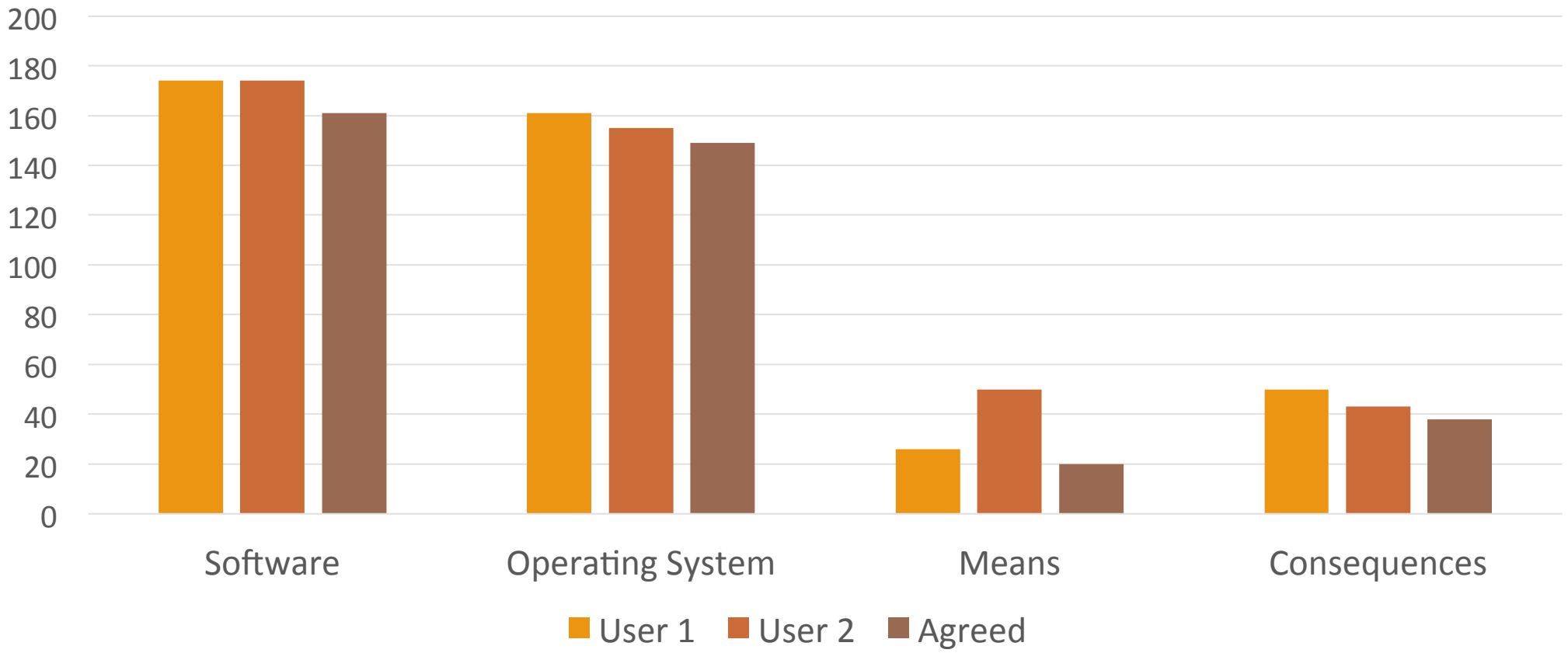
Some of the problems in text

"Cross-site scripting (XSS) vulnerability in show_bug.cgi in **Bugzilla before 3.6.13, 3.7.x and 4.0.x before 4.0.10, 4.1.x and 4.2.x before 4.2.5, and 4.3.x and 4.4.x before 4.4rc2** allows remote attackers to inject arbitrary web script or HTML via the id parameter……."

Some of the problems in text

"Cross-site scripting (XSS) vulnerability in show_bug.cgi in **Bugzilla before 3.6.13, 3.7.x and 4.0.x before 4.0.10, 4.1.x and 4.2.x before 4.2.5, and 4.3.x and 4.4.x before 4.4rc2** allows remote attackers to inject arbitrary web script or HTML via the id parameter that can result in delayed service response or even denial of service……."

Inter-Annotator Agreement (classes)



Other Systems for extracting named entities and concept terms

- OpenCalais
  - Designed for people, places, organization, facts and events
  - Very Poor Performance on Means and Consequences
  - Unable to capture version numbers

- DBPedia Spotlight
  - Identify only those concepts that are present in Wikipedia
  - Poor performance on list on means, consequences, file name and even software with low confidence score

OpenCalais

Example:

"Adobe Reader and Acrobat 9.x before 9.5.4, allow remote attackers to execute arbitrary code or cause a denial of service (memory corruption) via a crafted PDF document, as exploited in the wild in February 2013".

DBpedia Spotlight

Example:

"Buffer Overflow in the Broker Service in Adobe Flash Player 10.x and earlier versions on Windows XP SP2 and Mac OS X on Linux, allows attackers to execute arbitrary code via unspecified vectors."

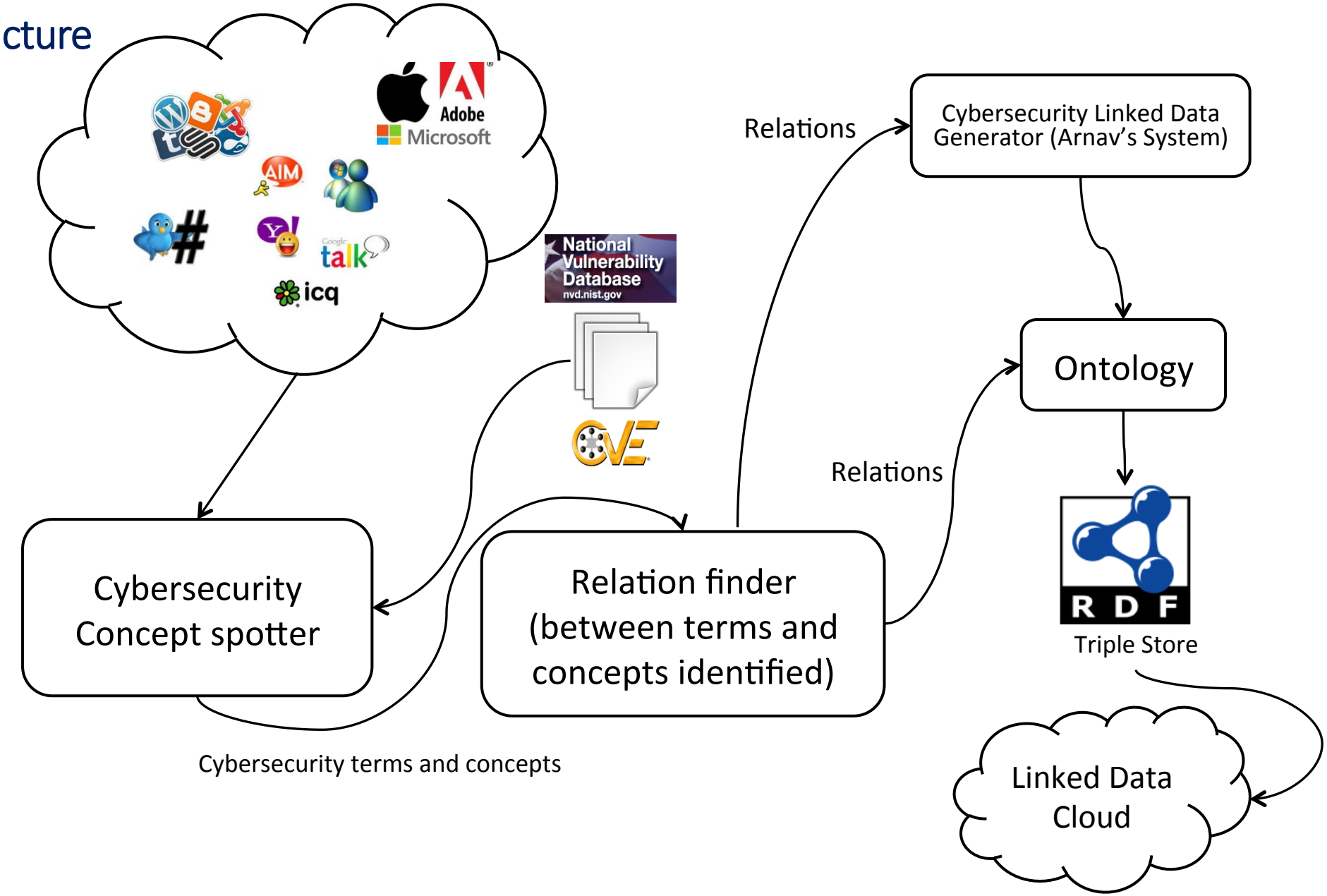
Challenges

- Consequences vs. Means
  - Is it a consequence? Or Is it a means? Or May be none?
- Data Noise (i.e., “(Parenthesis)”)
- Domain Experts for annotations
- Stanford NER documentation

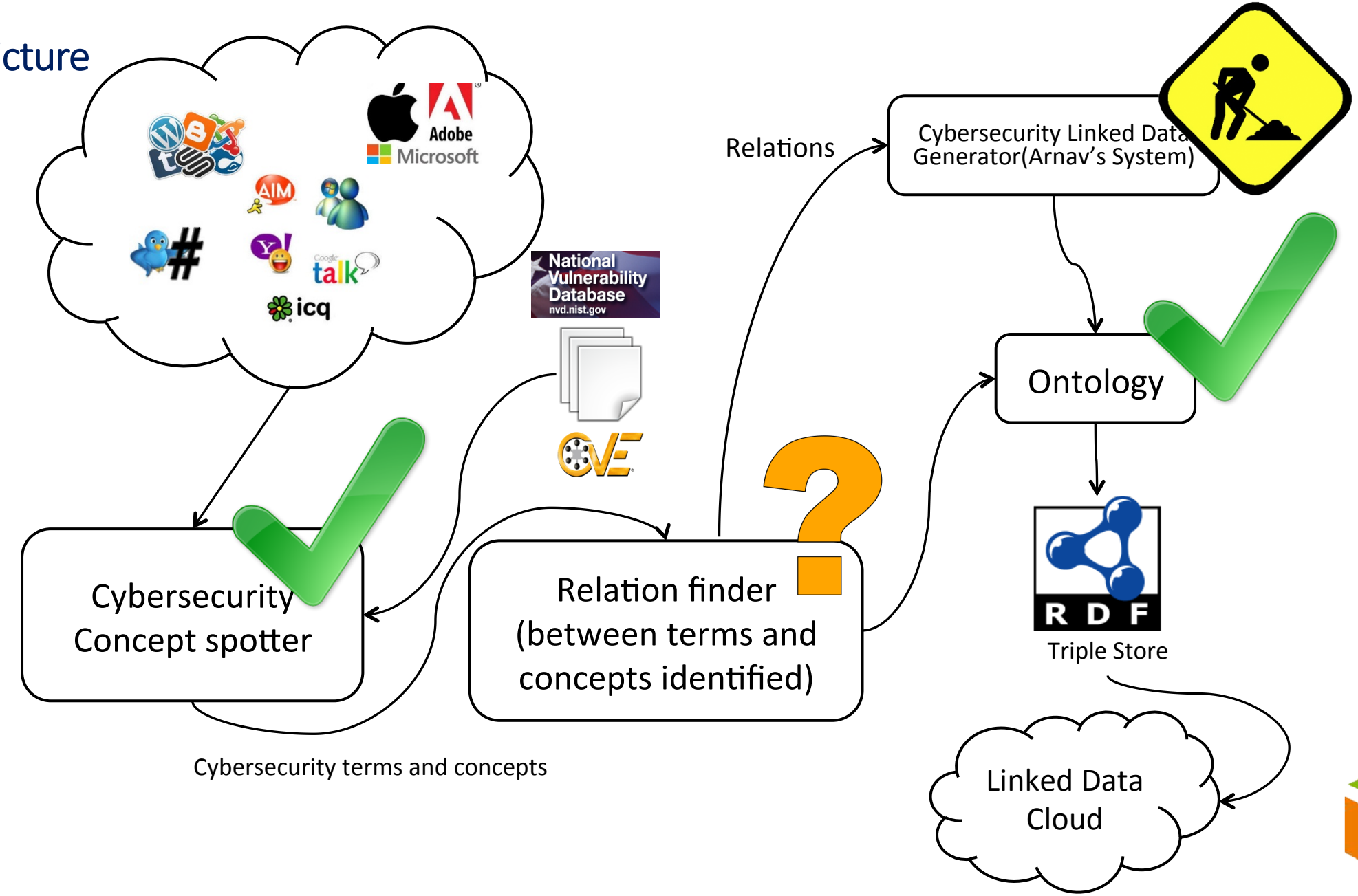
Conclusion

- There is a lot of information on Internet on software vulnerabilities
- Need to build automated service to process data streams
- Early detection, early prevention
- This Cybersecurity Entity and Concept spotter is the first step and can be used for multiple applications requiring processing of technical texts
- We believe this work can be used as foundation for analyzing Cyber security web text

Back to Big Picture



Back to Big Picture



Questions?

Thank You.

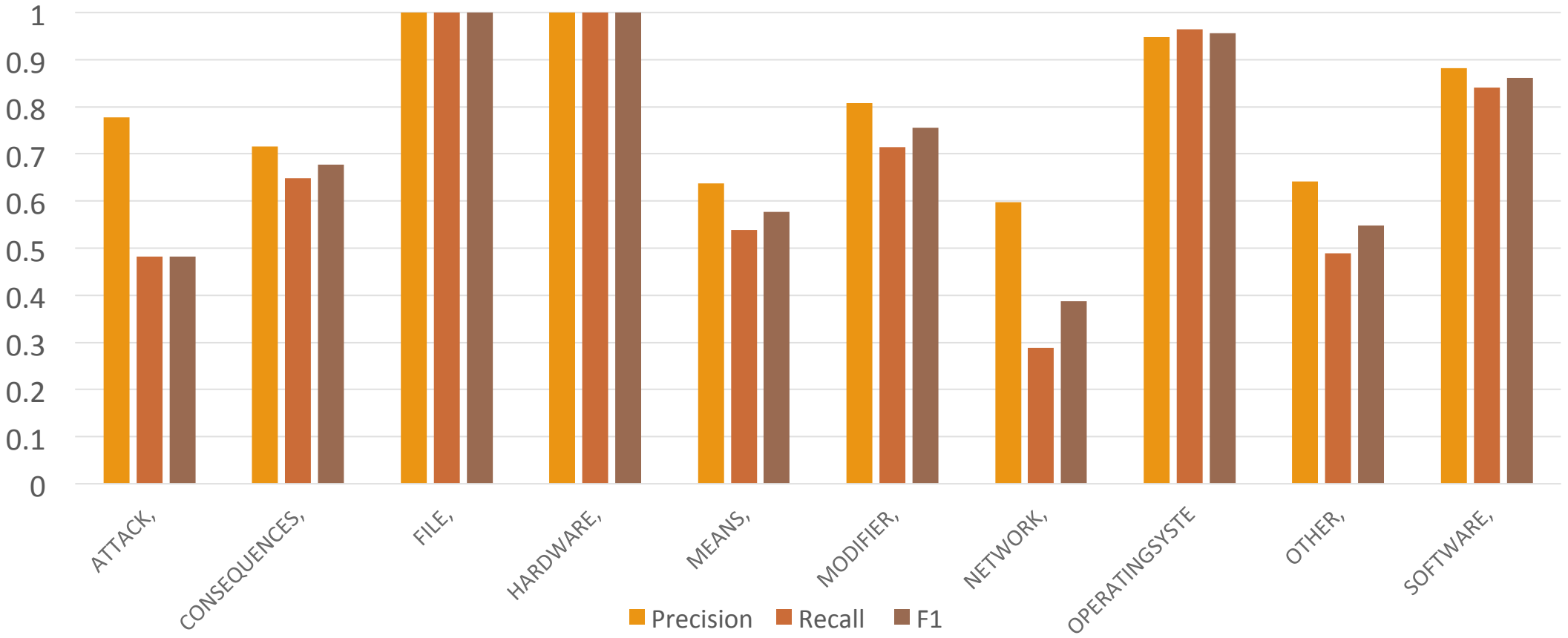
Back Up Slides

3-Fold Cross Validation Results

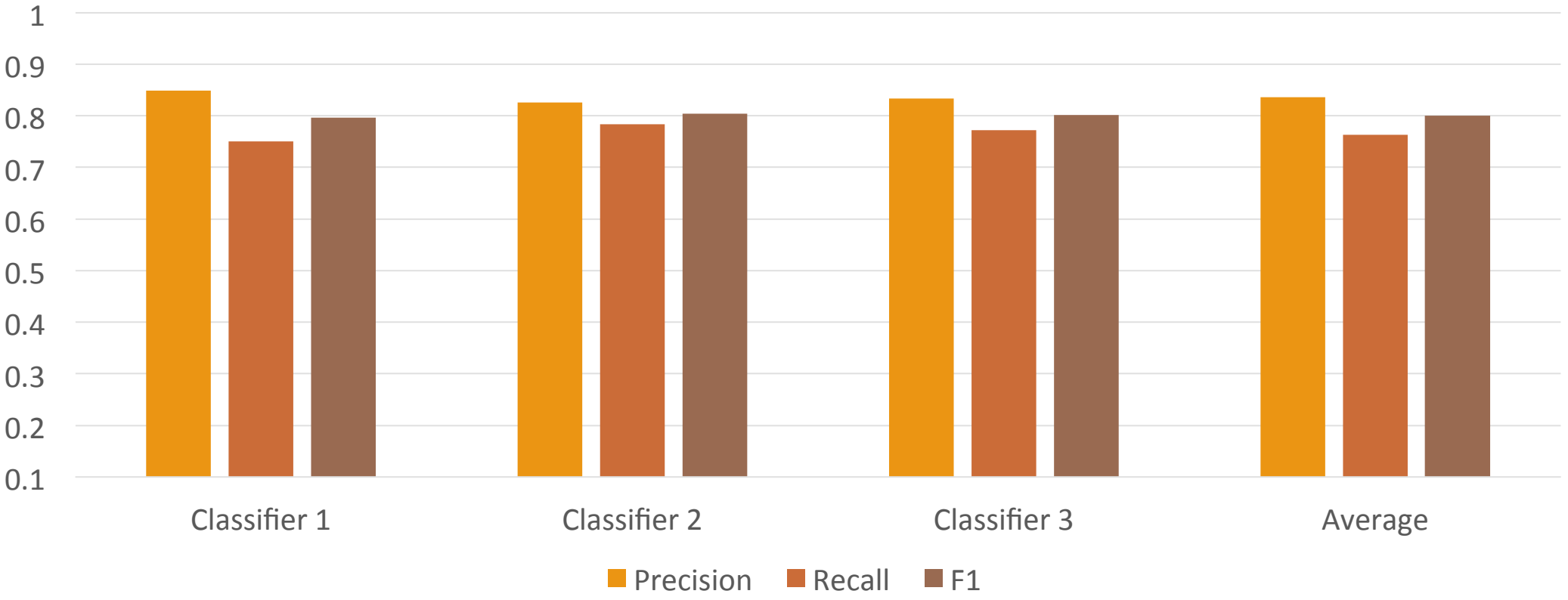
Strategy for training and testing Systems:

- Divided annotated data in 3 chunks (33% each)
- Trained 3 classifiers with various combinations of chunks
- Trained on 2/3rd and Tested on 1/3rd
- Evaluation score is Based on precision and Recall model

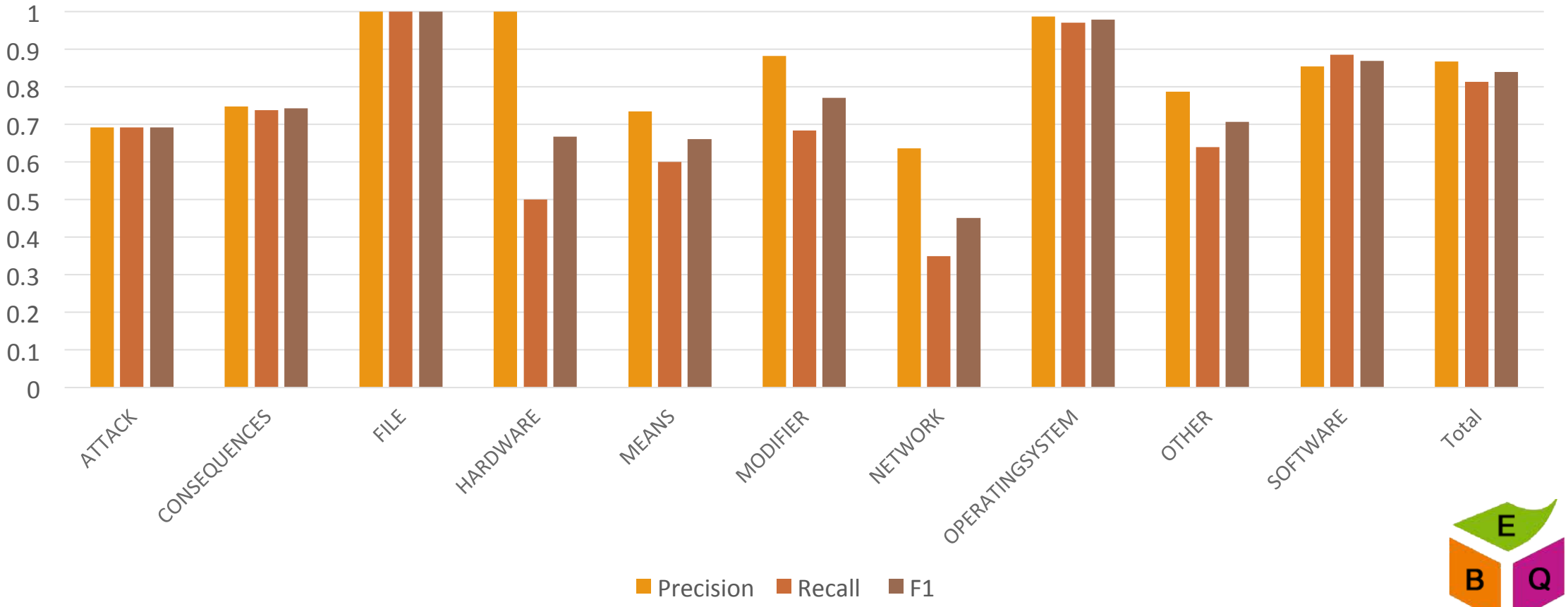
Average (Class-wise) for 3-Fold



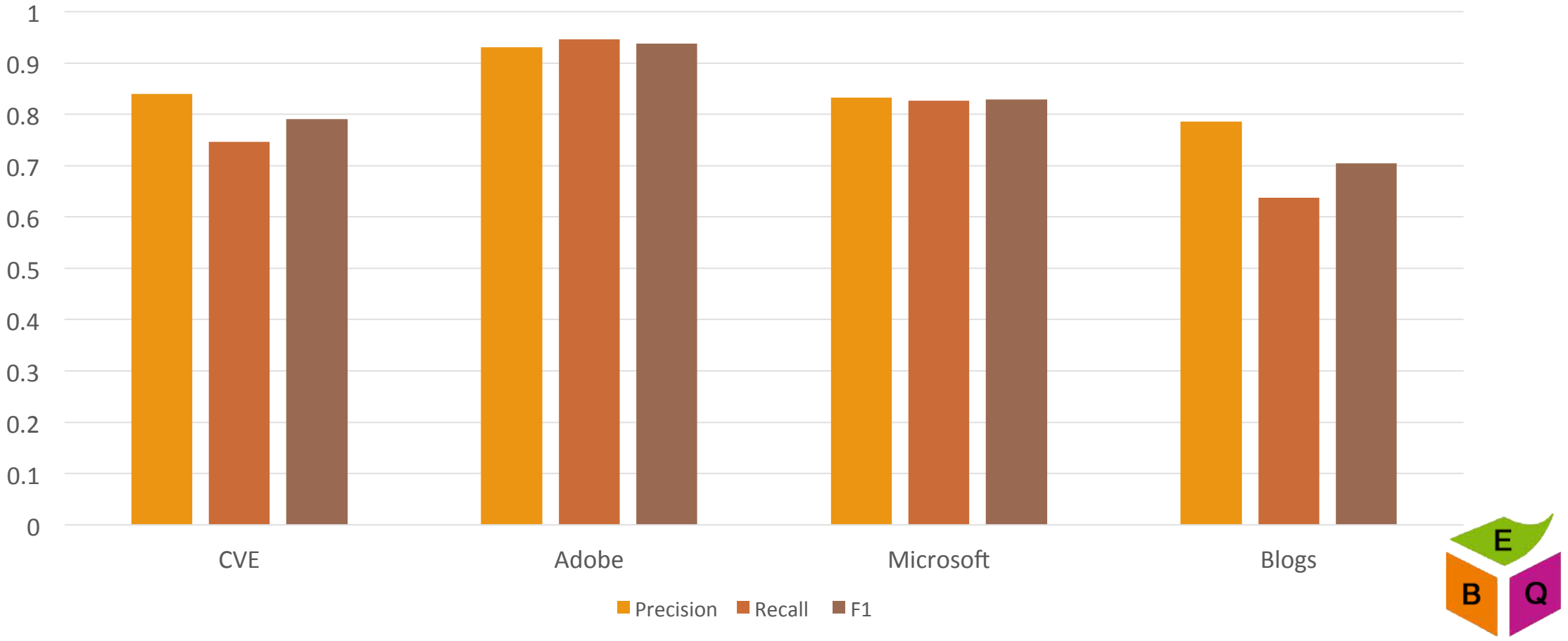
Average (Dataset-wise) for 3-Fold



Classifier 2



Classifier 2 (Data-wise)



Classifier 2 (In detail)

Entity	TP	FP	FN
ATTACK,	4	1	6
CONSEQUENCES,	59	31	27
FILE,	9	0	0
MEANS,	45	27	44
MODIFIER,	46	27	21
NETWORK,	1	1	7
OPERATINGSYSTEM,	156	8	4
OTHER,	20	12	40
SOFTWARE,	265	50	50
Totals	**605**	**157**	**208**

Classifier 2 (Adobe)

Entity	P	R	F1	TP	FP	FN
ATTACK,	0.7273	0.7273	0.7273	8	3	3
CONSEQUENCES,	0.5556	0.3846	0.4545	5	4	8
FILE,	1	1	1	2	0	0
OPERATINGSYSTEM,	1	1	1	128	0	0
OTHER,	0.6667	0.6667	0.6667	2	1	1
SOFTWARE,	0.9167	1	0.9565	121	11	0
Totals	0.9301	0.9466	0.9383	266	20	15

Classifier 2 (CVE)

Entity	P	R	F1	TP	FP	FN
ATTACK,	1	0.5	0.6667	1	0	1
CONSEQUENCES,	0.7407	0.8163	0.7767	40	14	9
FILE,	1	1	1	7	0	0
HARDWARE,	1	0.5	0.6667	2	0	2
MEANS,	0.7778	0.5714	0.6588	28	8	21
MODIFIER,	0.8816	0.6907	0.7746	67	9	30
NETWORK,	0.3333	0.3333	0.3333	1	2	2
OPERATINGSYSTEM,	1	0.9348	0.9663	43	0	3
OTHER,	0.8333	0.625	0.7143	35	7	21
SOFTWARE,	0.8182	0.8438	0.8308	81	18	15
Totals	0.8402	0.7457	0.7902	305	58	104

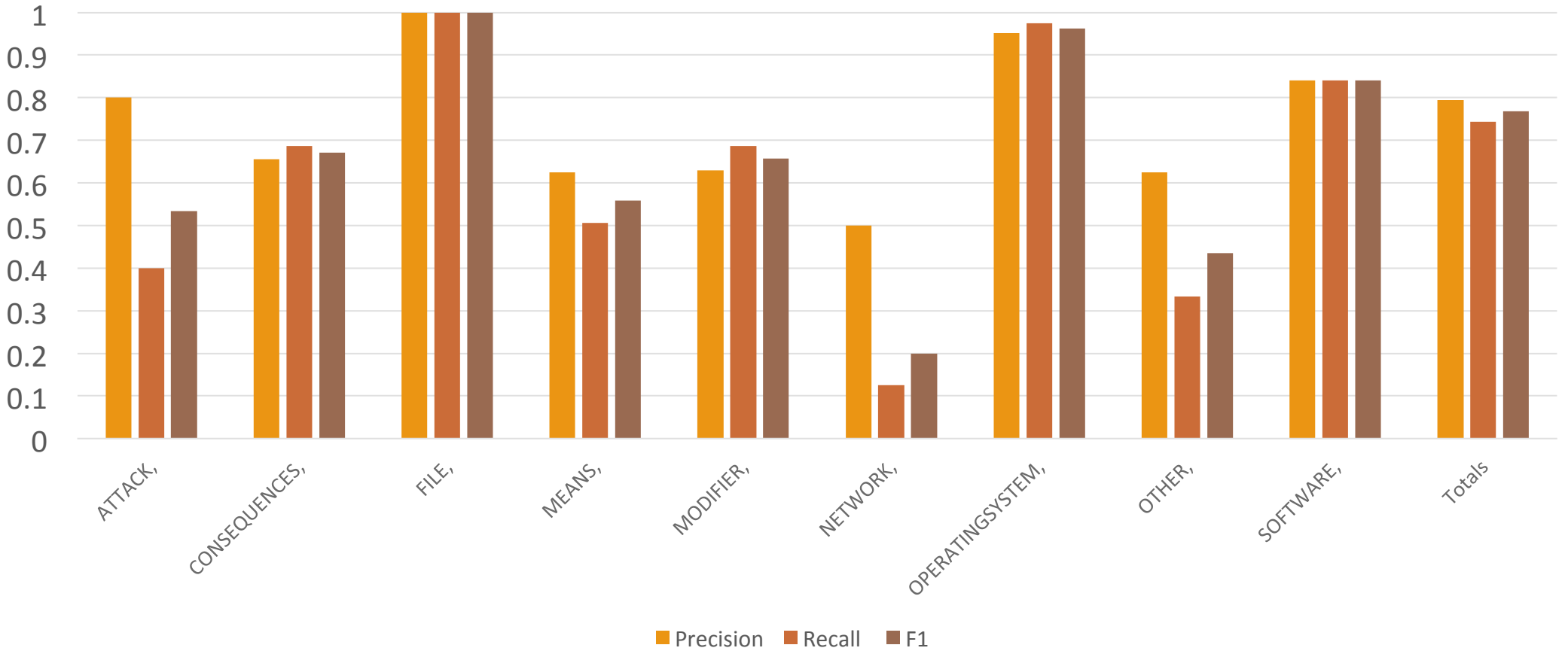
Classifier 2 (Microsoft)

Entity	P	R	F1	TP	FP	FN
CONSEQUENCES,	0.875	0.875	0.875	14	2	2
MEANS,	0.7273	0.8889	0.8	8	3	1
NETWORK,	0.8571	0.4	0.5455	6	1	9
OPERATINGSYSTEM,	0.9744	0.9268	0.95	38	1	3
OTHER,	0.7333	0.7333	0.7333	11	4	4
SOFTWARE,	0.7708	0.881	0.8222	37	11	5
Totals	0.8321	0.8261	0.8291	114	23	24

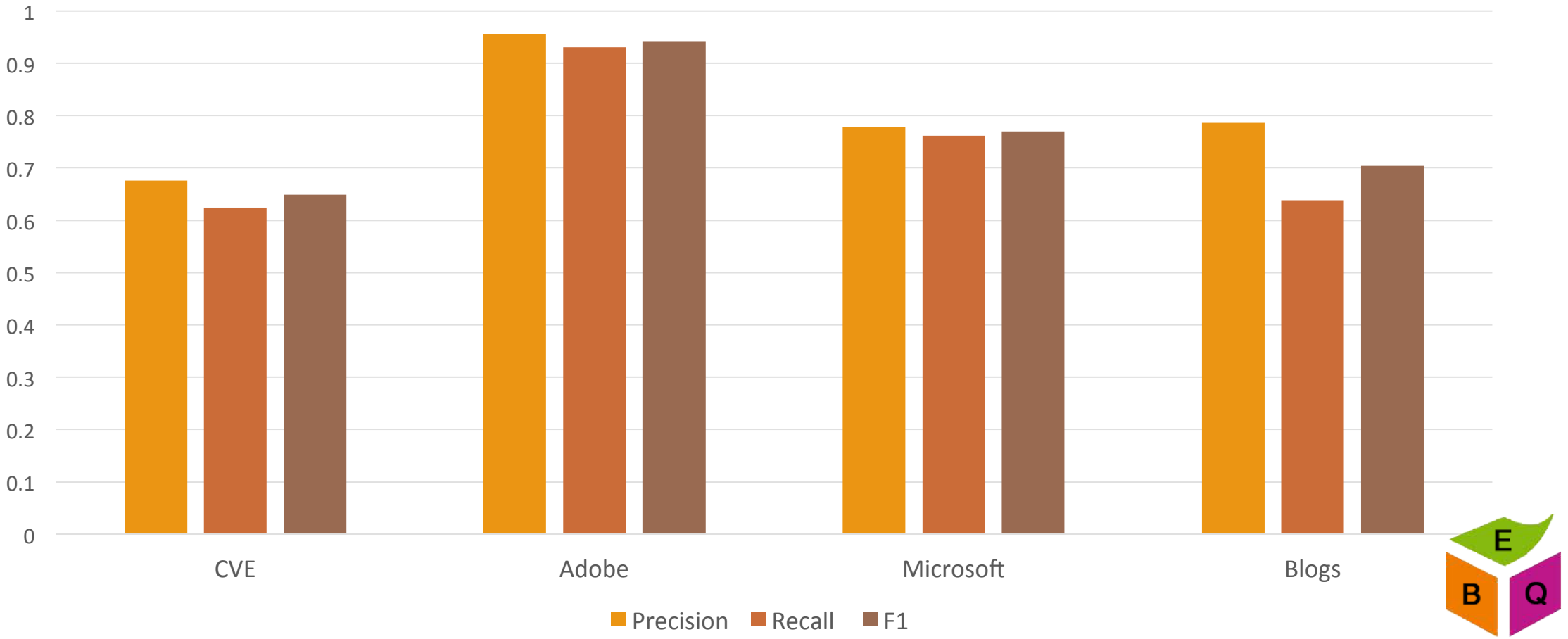
Classifier 2 (Blogs)

Entity	P	R	F1	TP	FP	FN
MEANS,	0	0	0	0	1	2
NETWORK,	0	0	0	0	1	2
OPERATINGSYSTEM,	0.875	0.9333	0.9032	14	2	1
OTHER,	0	0	0	0	1	1
SOFTWARE,	0.8108	0.6667	0.7317	30	7	15
Totals	0.7857	0.6377	0.704	44	12	25

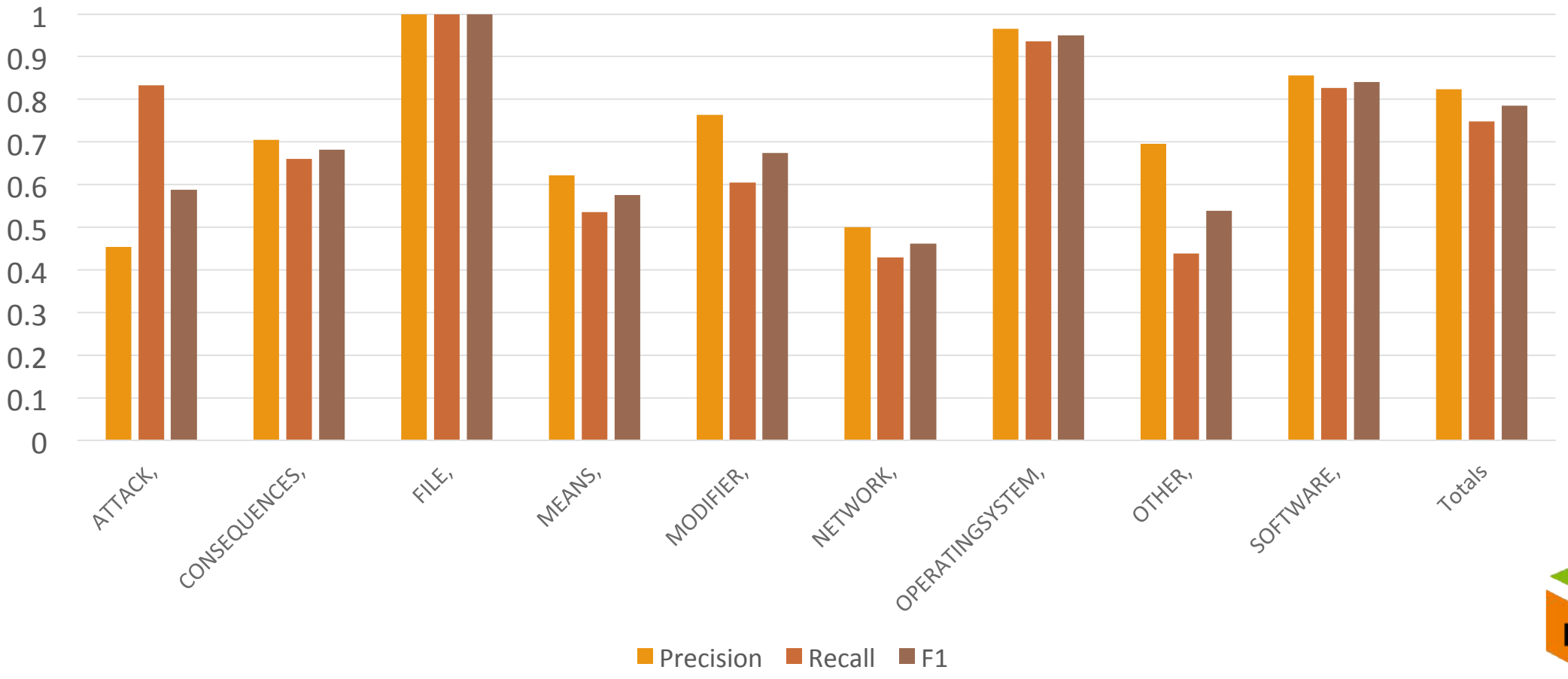
Classifier 2



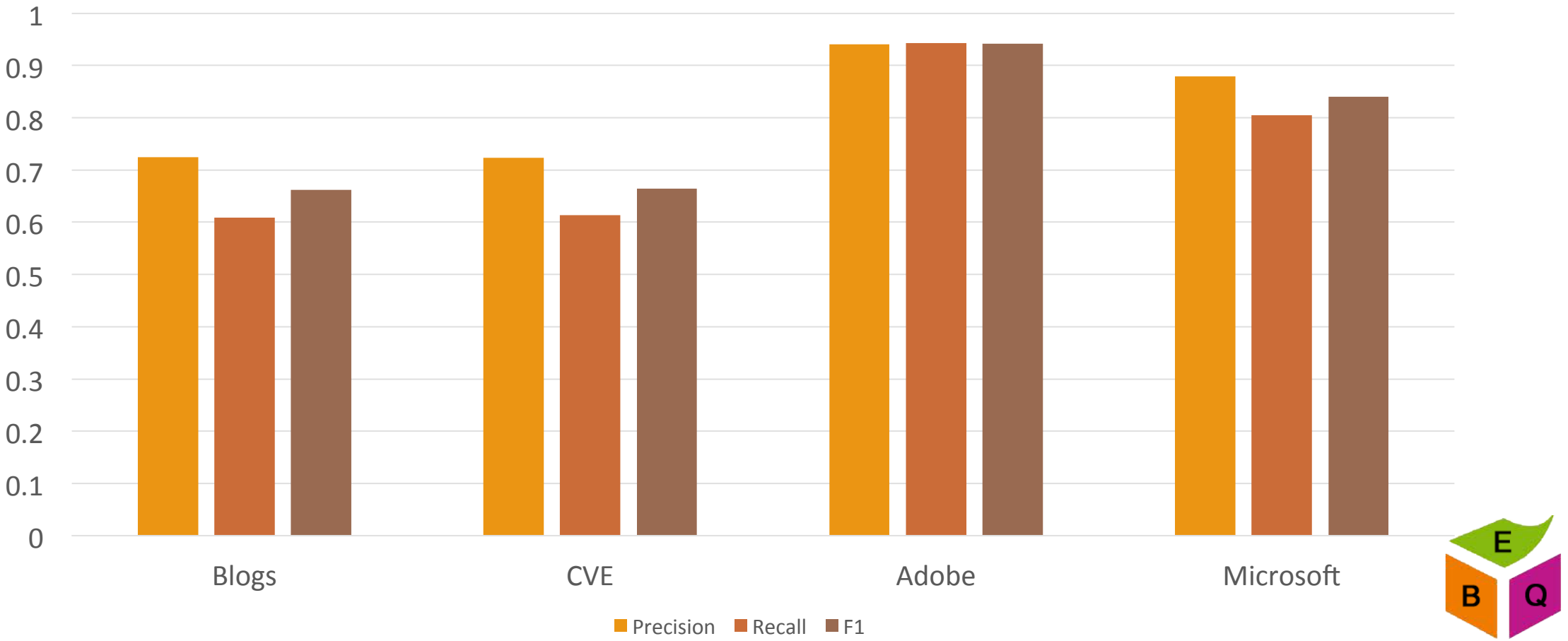
Classifier 2 (Data-wise)



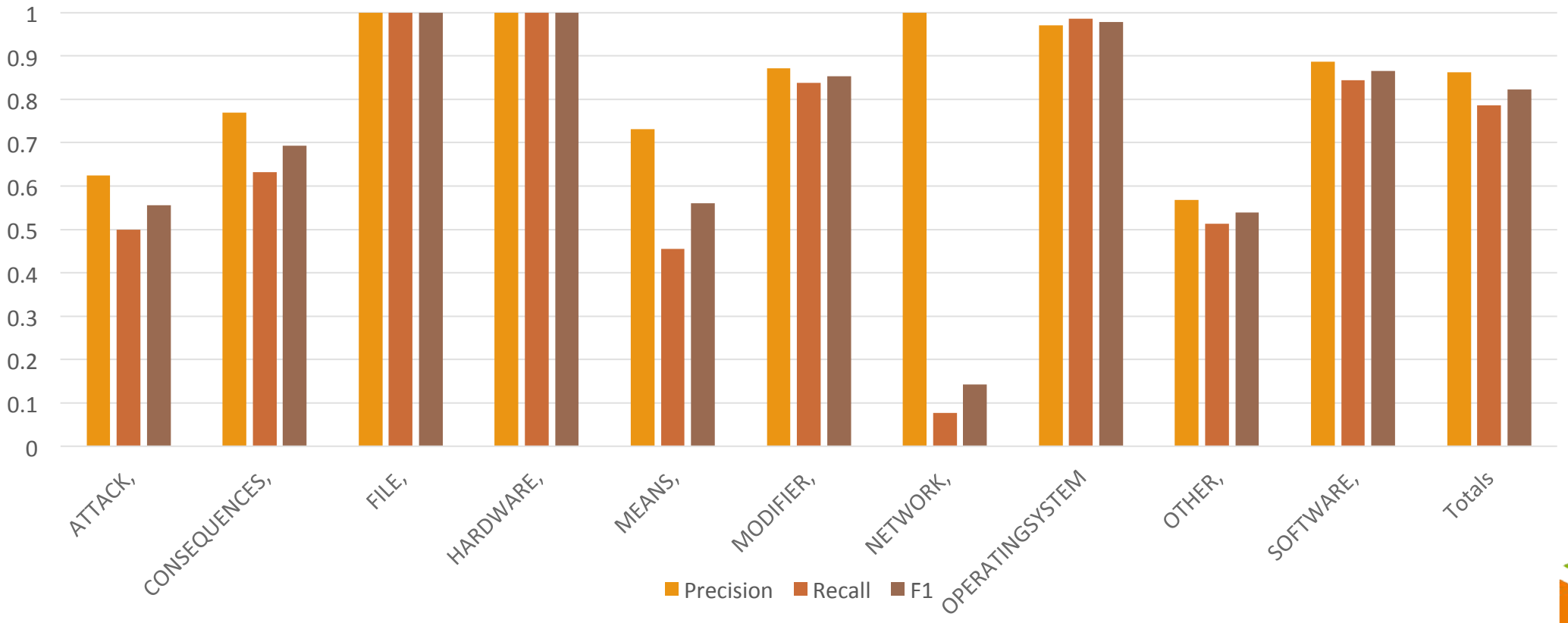
Classifier 3



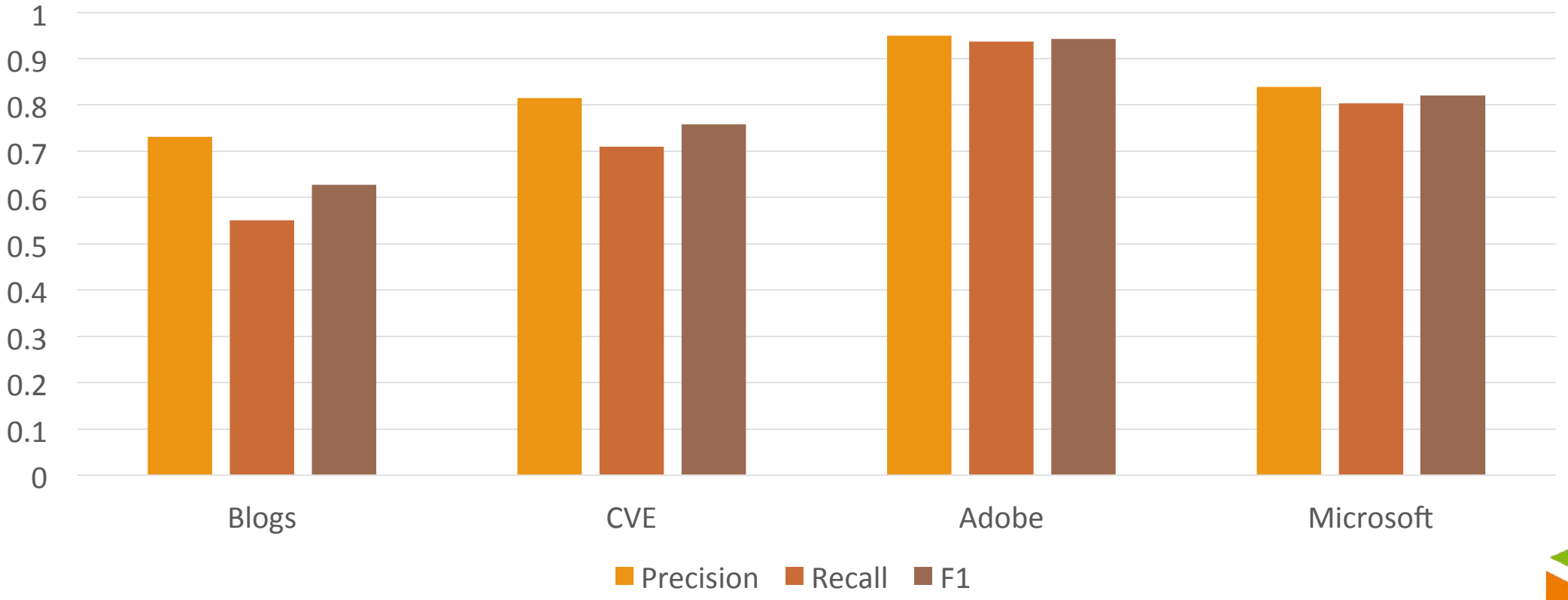
Classifier 3 (Data-wise)



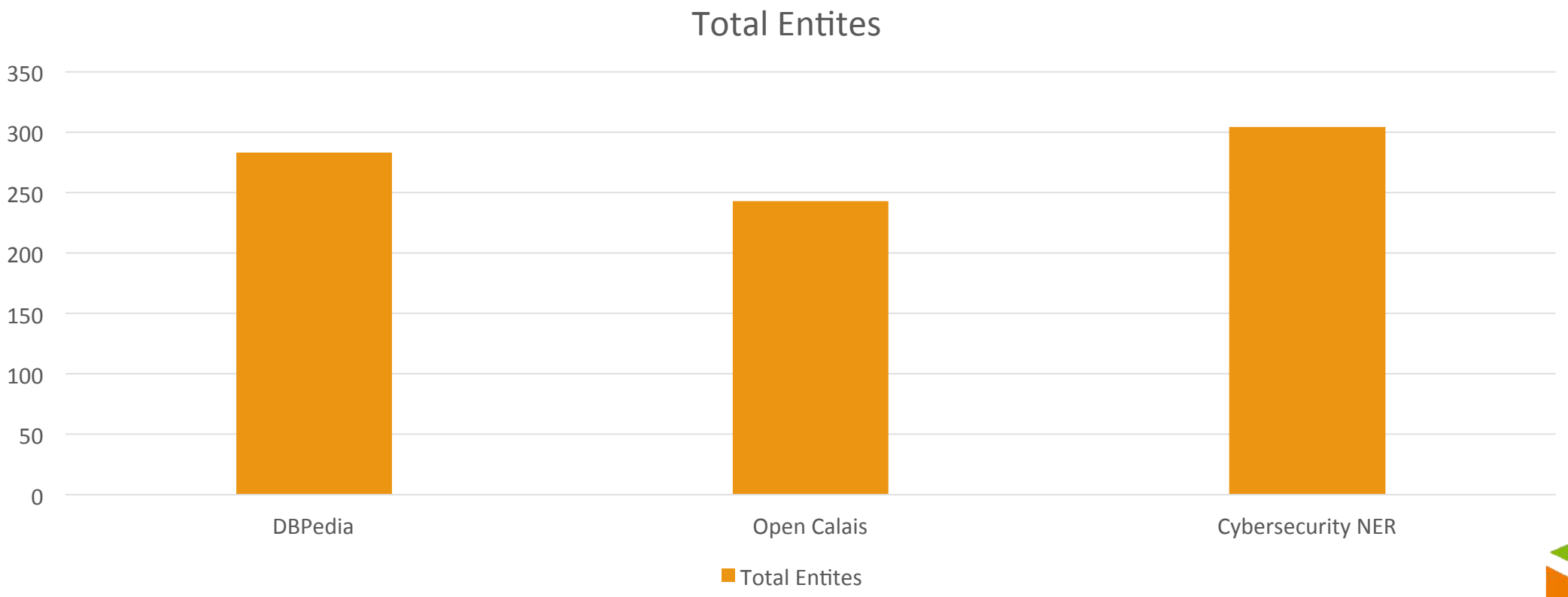
Classifier 4



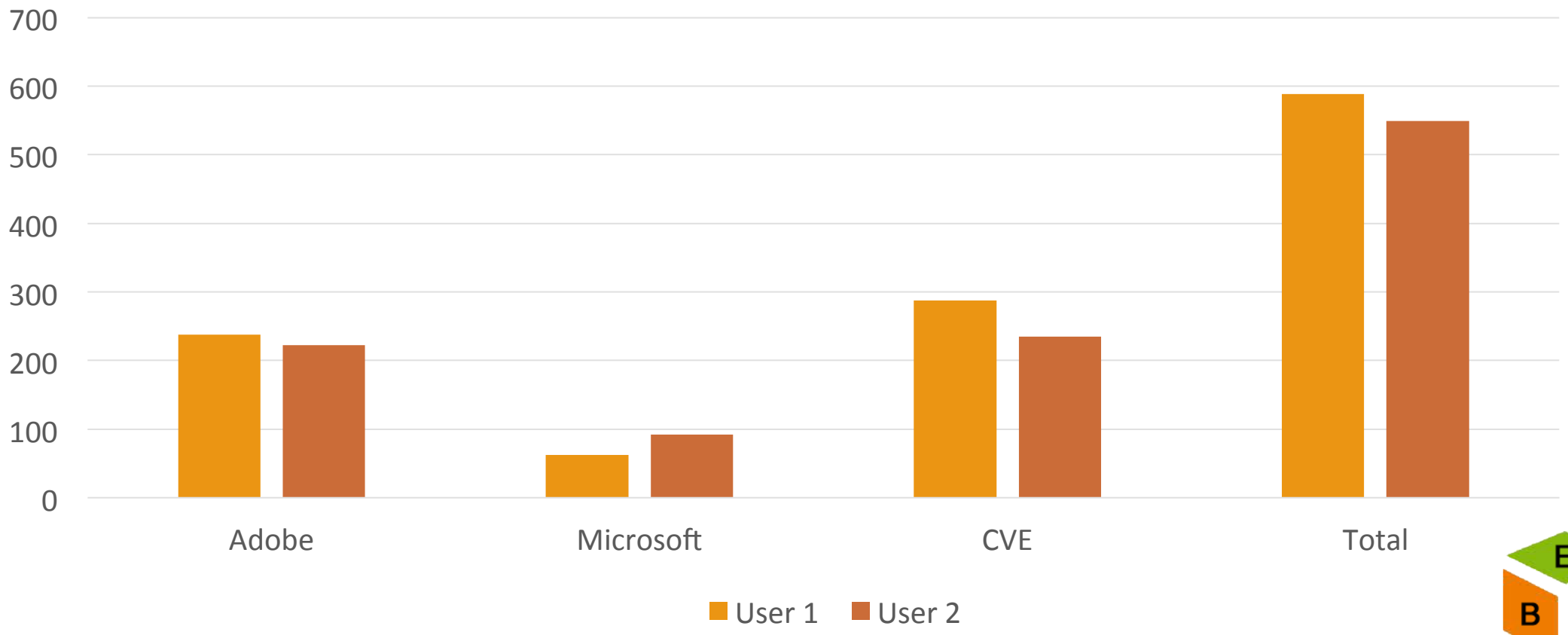
Classifier 4 (Data-wise)



In Numbers


Total Entites
350
300
250
200
150
100
50
0
DBPedia
Open Calais
Cybersecurity NER
Total Entites

Inter-Annotator Agreement (Dataset)



Why Mturk failed?

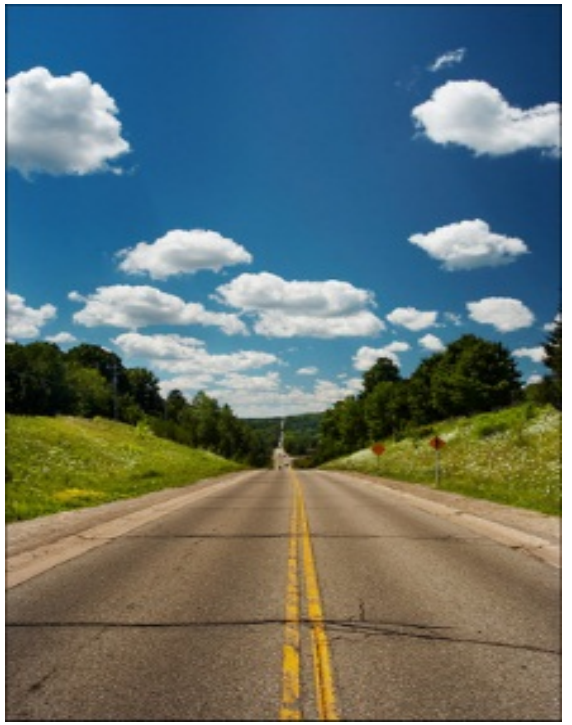
- User selection
- Domain expert knowledge
- Small Experiment

Training Approach: CRF

- Conditional Random Field (CRF) was introduced in 2001
- Stanford NER provides java based general implementation of CRF model
- Based on the idea of Sequence model (HMM)

Future Work

- Part of Speech Tagging:
  - This will help in improvement of results
  - This will help in finding relationships between entities
- Summarization and Inference:
  - Summarizing piece of text to infer



Web

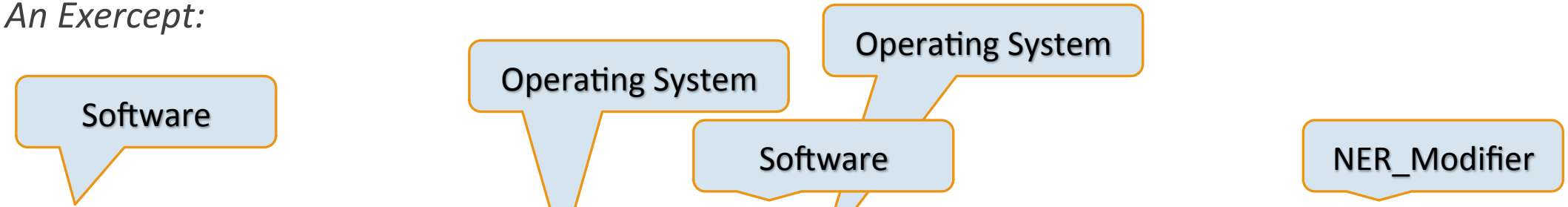

CYBER
ATTACK
INTERNATIONAL
CRIME
INTERNET
SERVICE
UNAUTHORIZED
COMPUTER
WARFARE
NETWORKS
INDIVIDUALS
CRIMINAL
THEFT
INFORMATION
CONTENT
BETWEEN
HACKING
CYBERSPACE
SERVICES
ACTIVITIES
TARGET
DATA
CASES
SENDING
MAIL
NEEDED
ACCESS
WEBSITES
INCLUDING
PORNOGRAPHY
ILLEGAL
CHILD
ALTERING
SYSTEMS
COMMUNICATIONS
TERRORISM
NETWORK
ROOMS
OFFENSIVE
CRIMES
SECURITY
HACKERS
GROUPS
USED
DRUG
FRAUD
INVESTIGATORS
USING

Motivation

- Various Information sources
  - Blogs
  - Security Bulletins
  - CVEs (i.e. 55691)
  - Chat rooms and Forums
  - Twitter
  - Network scanner logs (i.e. Nessus)

So… what's happening here?

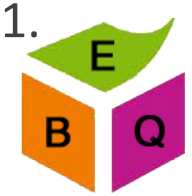
An Exercept:



"Adobe has released security updates for Adobe Reader and Acrobat X (10.1.3) and earlier versions (NER_Modifier) for Windows and Macintosh. These updates resolve a buffer overflow vulnerability in AcrobatReader.exe (FILE_NAME) that could lead to remote code execution (CONSEQUENCES) ……"

References

- Mulwad, Varish, et al. "Extracting Information about Security Vulnerabilities from Web Text." *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on*. Vol. 3. IEEE, 2011.
- Finkel, Jenny Rose, Trond Grenager, and Christopher Manning. "Incorporating non-local information into information extraction systems by gibbs sampling."*Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2005.
- More, Sumit, et al. "A Knowledge-Based Approach To Intrusion Detection Modeling." *Security and Privacy Workshops (SPW), 2012 IEEE Symposium on*. IEEE, 2012.
- Mendes, Pablo N., et al. "DBpedia spotlight: shedding light on the web of documents." *Proceedings of the 7th International Conference on Semantic Systems*. ACM, 2011.



References (2)

- Jiao, Dazhi, and David J. Wild. "Extraction of CYP chemical interactions from biomedical literature using natural language processing methods." *Journal of chemical information and modeling* 49.2 (2009): 263-269.
- Finin, Tim, et al. "Annotating named entities in Twitter data with crowdsourcing." *Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's Mechanical Turk*. Association for Computational Linguistics, 2010.
- http://brat.nlplab.org/index.html
- http://nlp.stanford.edu/software/CRF-NER.shtml

Relevant work

- Mulwad, Varish, et al. "Extracting Information about Security Vulnerabilities from Web Text." *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2011 IEEE/WIC/ACM International Conference on*. Vol. 3. IEEE, 2011.
-