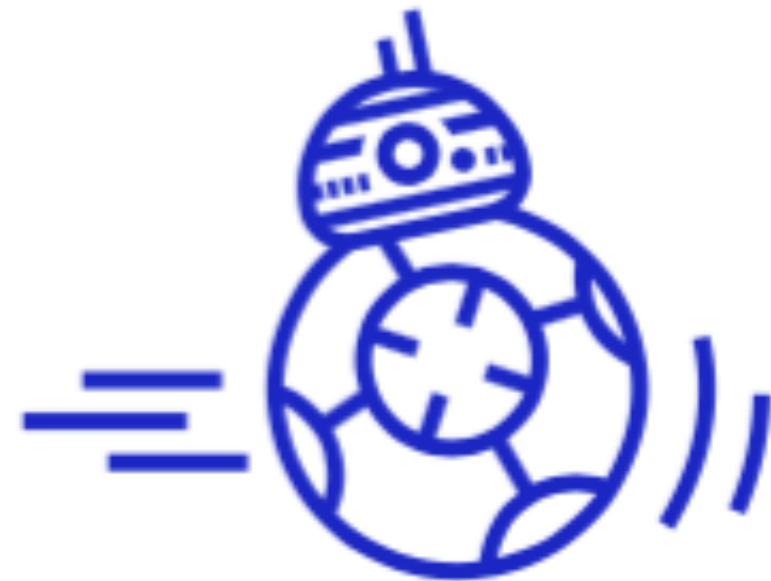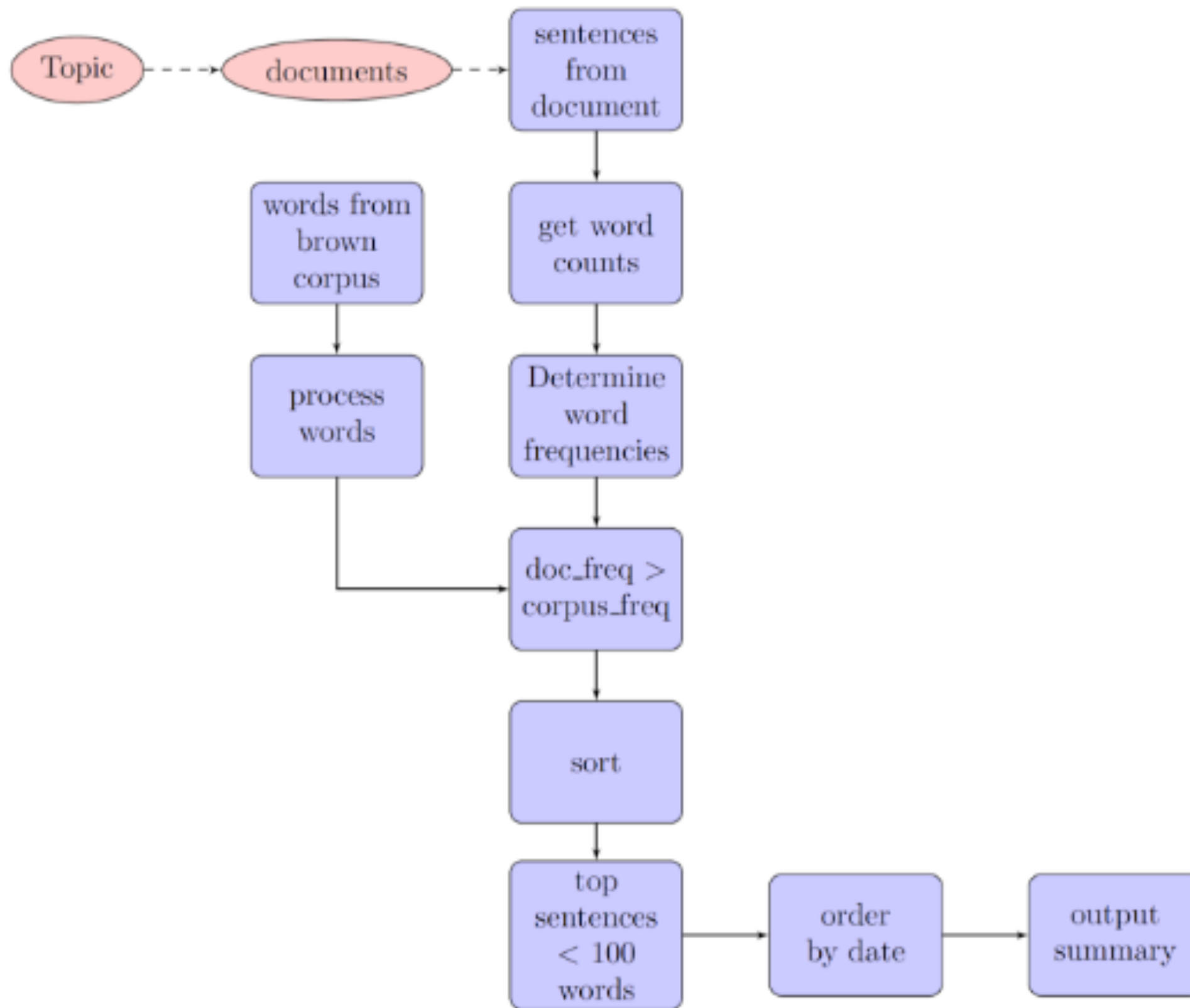# Deliverables 3

Matt Calderwood

Kirk LaBuda
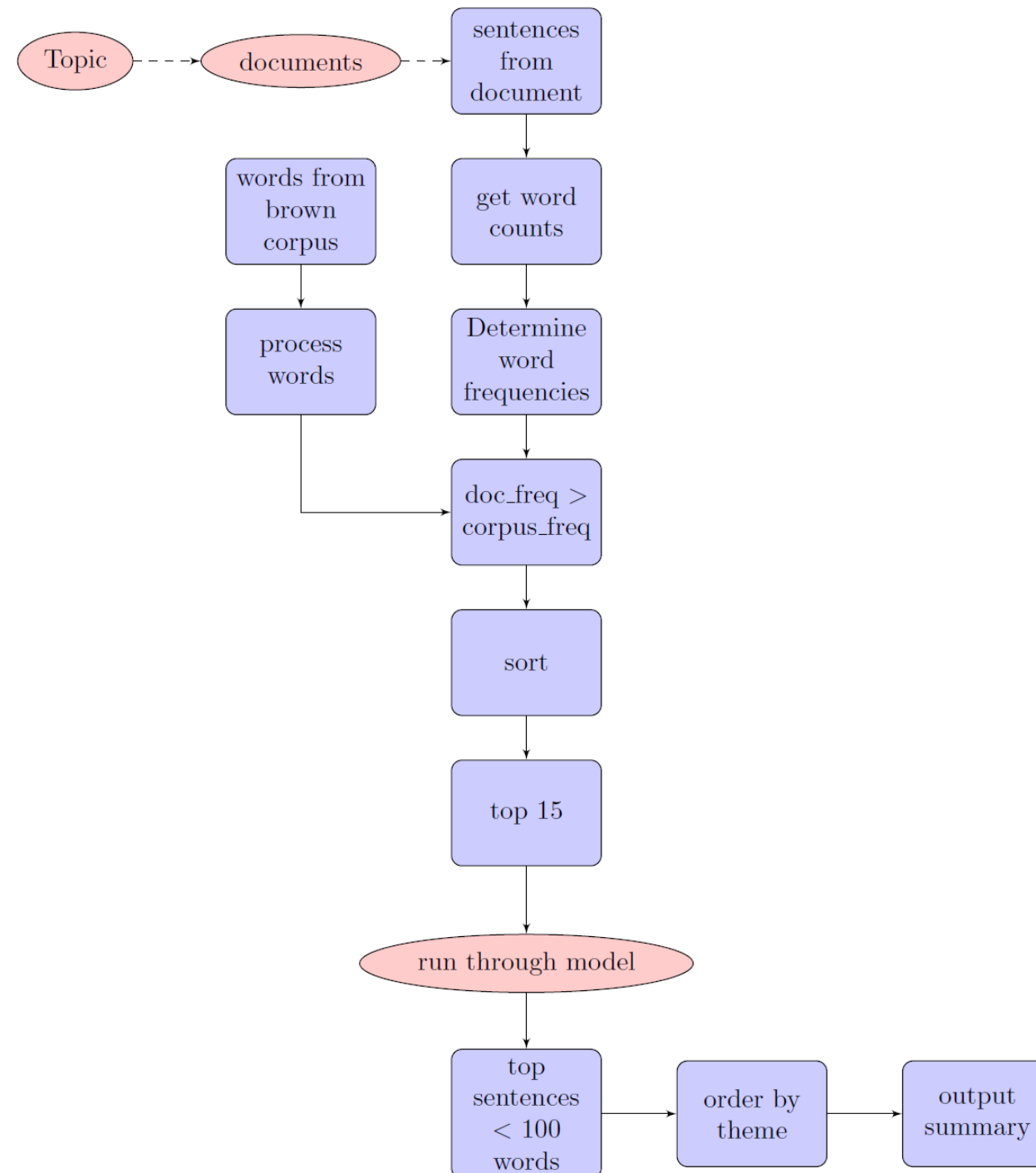
Nick Monaco
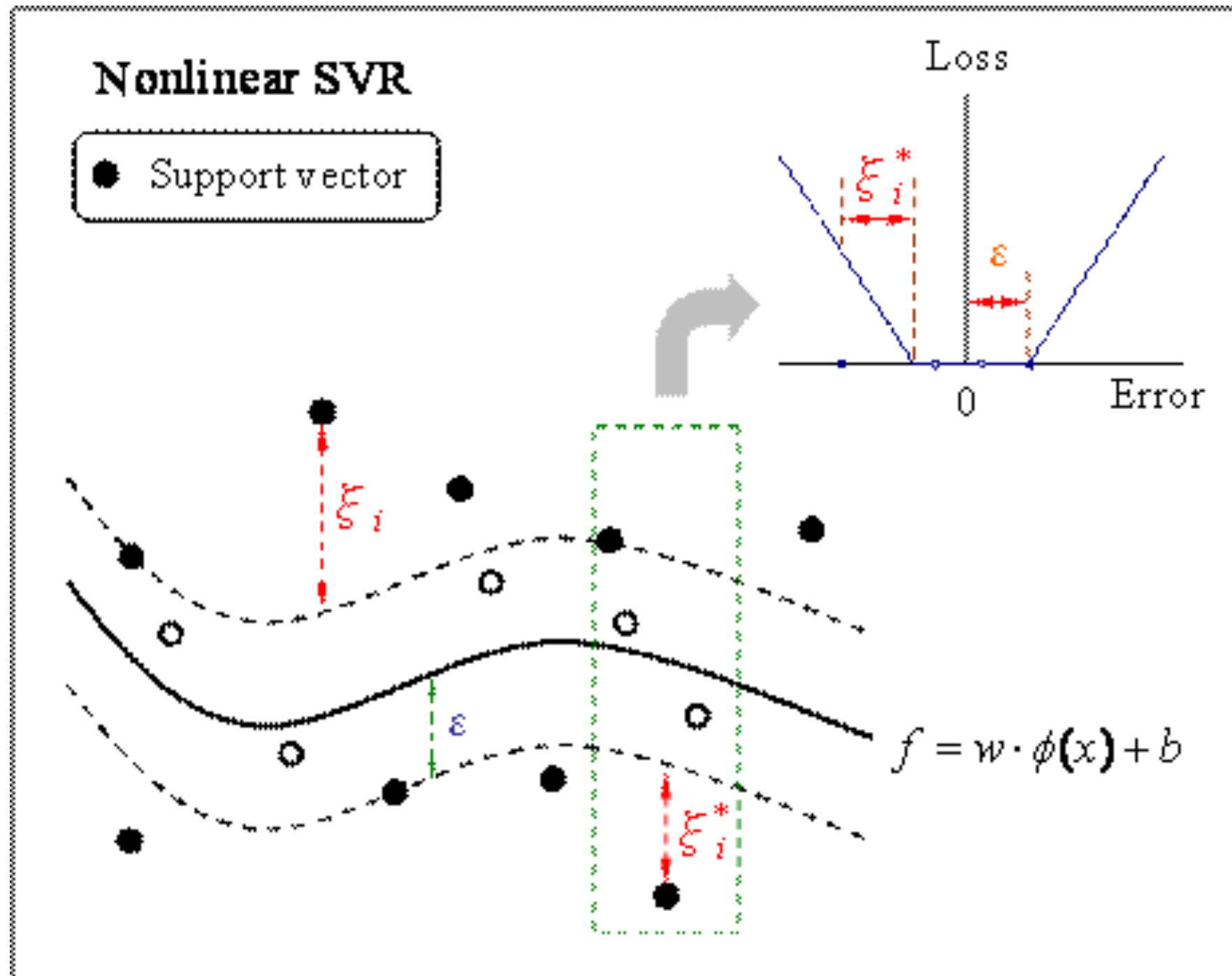
D2 System Architecture Diagram

Updated diagram for D3

System Changes

- Content Selection - incorporated new machine learning approach with support vector regression. Currently using *linear kernel.*

- Plot summaries in 3D space - use hyperplane to predict ROUGE score and choose summary with best probable ROUGE score
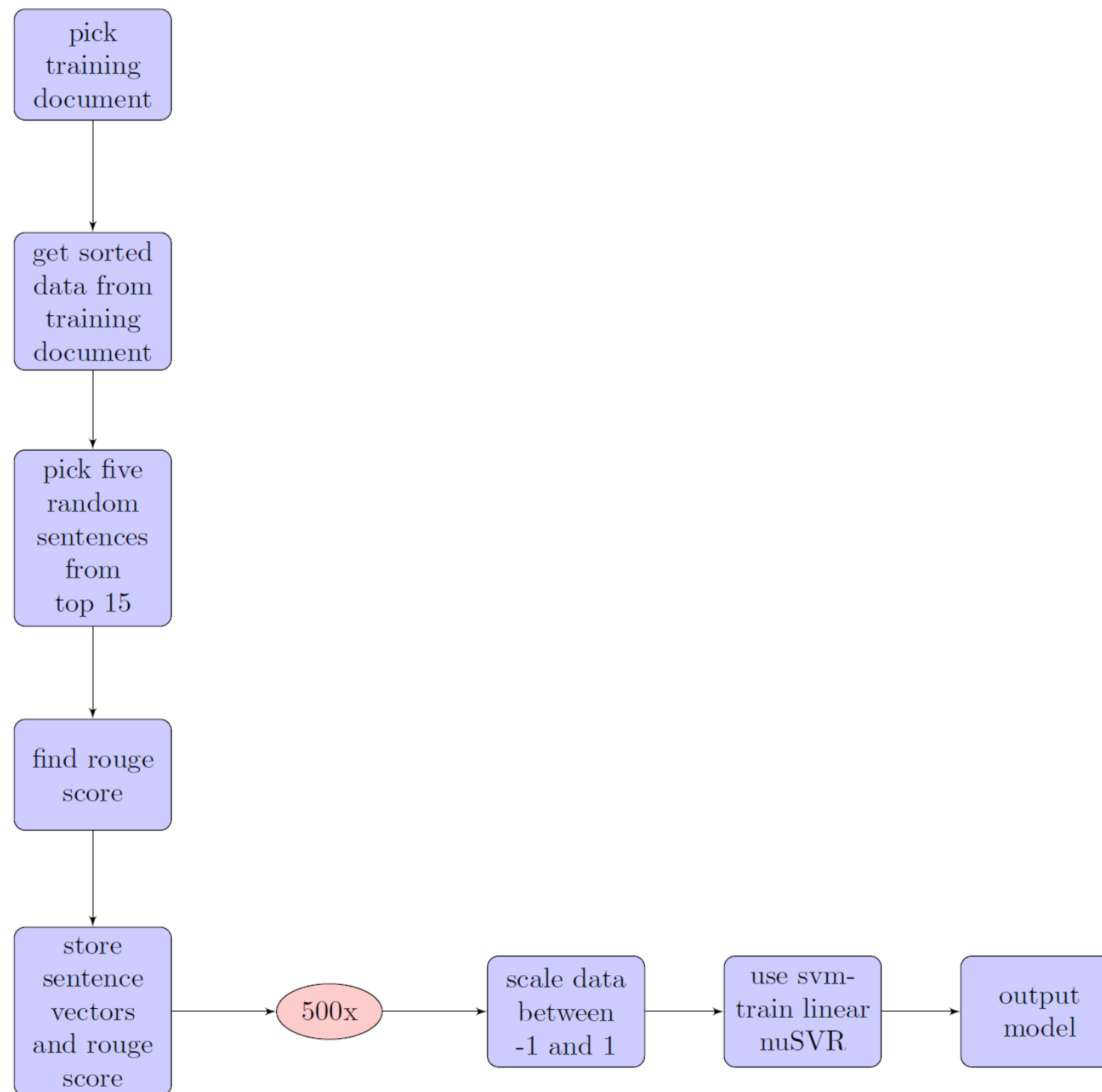
Support Vector Regression Diagram

"

Suppose we are given training data $\{(x_1, y_1),...,(x_\ell, y_\ell)\} \subset X \times R$, where X denotes the space of the input patterns (e.g. $X = R^d$). In ε-SV regression, our goal is to find a function $f(x)$ that has at most ε deviation from the actually obtained targets y_i for all the training data, and at the same time is as flat as possible. …we do not care about errors as long as they are less than ε…

Explanation of SVR, Alex J. Smola

Support Vector Regression - Training Diagram
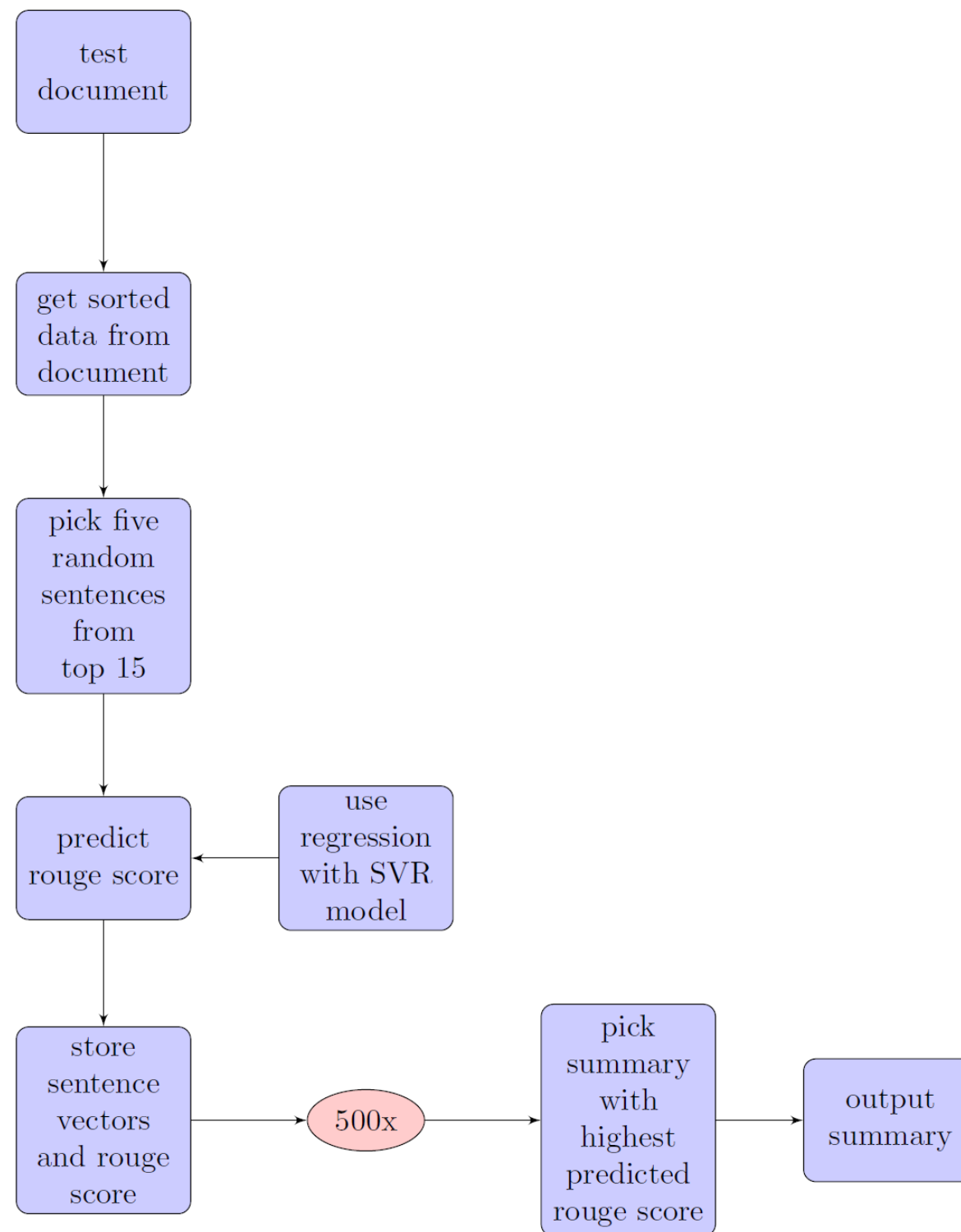
Support Vector Regression - Testing Diagram

System Changes (cont.)

- Information Ordering - Order summary sentences by theme. Shallow approach.

- Content Realization - no changes

Successes

- Machine Learning/SVR- approach seems promising.

- Info Ordering - shallow approach seems reasonable, has yielded some good results.

Issues

- <u>Machine learning approach</u> - still experimenting with different kernel fns for SVR. Planning to use more features.

- <u>Info Ordering</u> - multiple sentences occasionally registering as one sentence -skews results.

- <u>Content Realization</u> - could use sentence compression - some summaries contain long sentences.

- <u>Runtime-</u> room for optimization with runtime.

<u>Good:</u>

```
D1019-A.M.100.D.1
Six masked assailants have taken more than 20 people, including
Americans and Europeans, hostage on a remote resort island in eastern Malaysia,
police said Monday. Prior to the rebel announcement, Col. Ernesto de Guzman,
chief of staff of the Philippine military's Southern Command, said officials were
trying to determine whether there was a connection with the hostage situation on the
southern Philippine island of Basilan, about 275 miles northeast of Semporna,
the port from which tourists take boats for Sipadan.
```

<u>(#meh):</u>

```
D1030-A.M.100.F.1
And, more importantly, are they safe? The company introduced Cholestin at the end of 1996,
after nearly three years of study. Current co-host Robin Hartl joined ``Hometime'' in July 1992.
Obviously we are not breaking any laws. He pleaded guilty to using a telephone in the trafficking
of drugs and was sentenced to five years' probation.
```

Qualitative summary examples

ROUGE results

	D2	D3
ROUGE-1 Average_R	0.10987	**0.12951**
ROUGE-2 Average_R	0.01891	**0.02431**
ROUGE-3 Average_R	0.00502	**0.00612**
ROUGE-4 Average_R	0.00129	**0.00198**

Works Cited

- Chih-Chung Chang and Chih-Jen Lin, LIBSVM : a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, 2:27:1--27:27, 2011. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

- Smola, Alex J., and Bernhard Schölkopf. A Tutorial on Support Vector Regression ∗ (n.d.): n. pag. Http://alex.smola.org/papers/2003/SmoSch03b.pdf. 30 Sept. 2003. Web. 16 May 2016.

- Yu, Pao-Shan, Shien-Tsung Chen, and I-Fan Chang. "Support Vector Regression for Real-time Flood Stage Forecasting." Journal of Hydrology, 328 (3–4), Pp. 704–716, Sept. 2006. Web. 16 May 2016.