# Secret Words: Revealing Semantic Patterns in Passwords through Visualization

Jeffrey Hickson ■ Julie Thorpe & Christopher Collins ■ Faculty of Business & IT and Science, UOIT ■ Ontario, Canada

Jeffrey.Hickson@mycampus.uoit.ca ■ Christopher.Collins@uoit.ca ■ Julie.Thorpe@uoit.ca

## Research Goals

Our goal is to reveal semantic patterns in the way people choose passwords. The reason we are doing this is because if common semantic patterns exist in passwords, this could be used by attackers to crack passwords much faster. Our goal is to discover these patterns before they are exploited for malicious means, and update password rules and recommendations accordingly.

## Lexical Analysis

While analyzing the results of the program we saw that a lot of the words found seem to have a positive sentiment. For instance the word “love” appears about 584000 times in the password database, or 1.3% of all found words.

## Password Processing

We store the passwords in the database, and then request them when we start the mining process. We then compare the passwords with the dictionary of our choice, collecting the words from the dictionary that are in the password. This list of words is then passed into an algorithm that computes the combination of the words that achieves the longest coverage of the password in the fewest words. The results are then written back into the database. The process is outlined below in Figure 1.

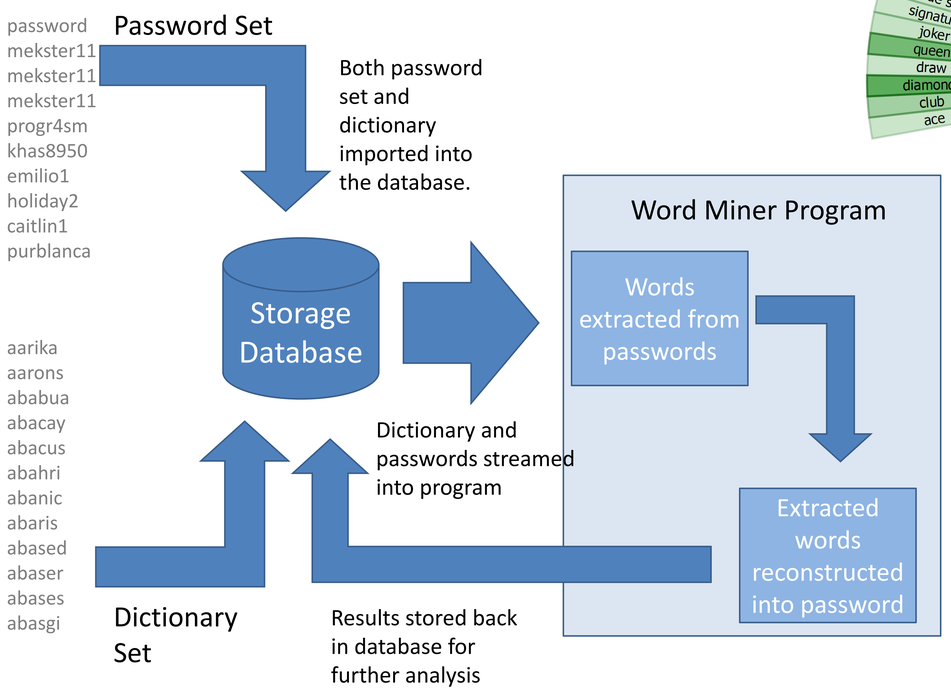


Fig 1: Word Miner Process

## Common Words

10 most common words and how frequently they occur. See Figure 3 for top 100 words from the RockYou dataset in a Wordle diagram, sized by occurrence.

| | | | |
|---|---|---|---|
| Love | 1.33% | As | 0.38% |
| Me | 0.57% | An | 0.37% |
| Baby | 0.51% | Girl | 0.36% |
| You | 0.49% | La | 0.34% |
| In | 0.39% | Is | 0.32% |

## DocuBurst

DocuBurst is an existing platform for visualization that allows us to detect higher-level semantic patterns in words. For example, Figure 2 shows that materials are a popular category in the words we extracted from passwords.

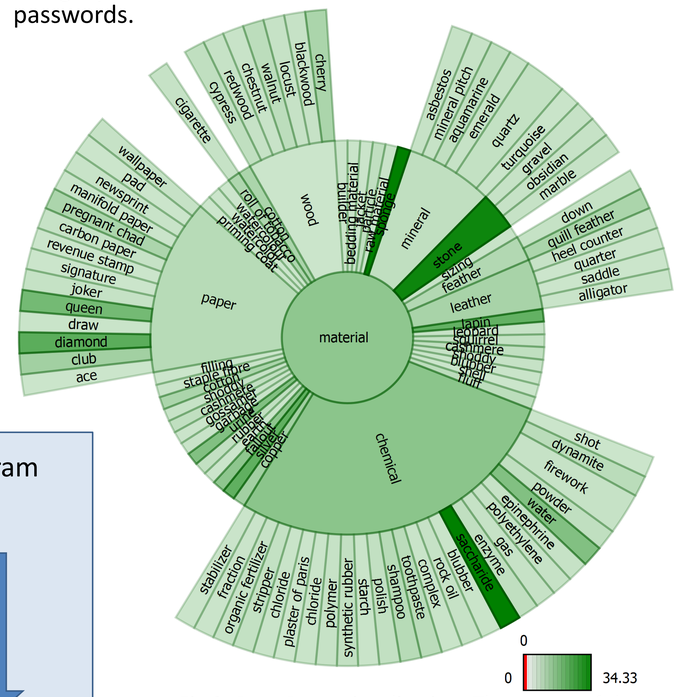


Fig 2: DocuBurst visualization. The center of the diagram represents the common category of all the words displayed.


Fig 3: Wordle tag cloud visualization. Words with a ‘*’ character have been censored.

## Patterns in Passwords

Some of the patterns we have found thus far within the sets of passwords we analyzed have so far been interesting. Some of the common patterns we have found come straight from the list of typical “don’t use one of these” passwords, such as: “password”, and “trustno{one,1}”, as well as some newer patterns which haven’t been noticed before like: “Ilove<noun>”, and some surprisingly common words such as animal names, very popularly “butterfly”, and emotions are also very popular.

## Security Implications

- We expect that our results will define a new type of password dictionary, which can be effectively used to detect weak choices while a user is creating/changing their password.
- Our results can also be used to define rules and recommendations to help users create stronger passwords.

## Ongoing Research

In the future we plan to further extend the program by adding a sentiment ranking of all the found words, and for each password; analyzing the use of other dictionaries; and analyzing other released password datasets. We also plan to generate password dictionaries, password rules, and password recommendations based on our findings.

## Acknowledgements

Dr. Collins and Dr. Thorpe would like to thank NSERC fro providing Discovery Grant support, through which this project was funded.

UOIT CHALLENGE INNOVATE CONNECT

NSERC CRSNG