

पासवर्ड रिकवरी के लिए आपना अद्वितीय शब्दकोश(डिक्शनरी) बनाना

© 2013 पास्केप सॉफ्टवेर (हिन्दी अनुवाद : धीरेन कुमार)
पास्केप सॉफ्टवेर (हिन्दी अनुवाद : धीरेन कुमार)

1. सार	3
2. कस्टम वर्डलिस्ट बनाना	5
2.1 अंतरराष्ट्रीय शब्दकोश बनाना	6
2.2 ~ 1 मिलियन लोकप्रिय इंटरनेट डोमेन की सूची बनाना।	7
2.3 यूजर्स के नामों के साथ एक शब्दकोश बनाना	7
Index	0

सार

1 सार

पहला पासवर्ड रिकवरी टूल आने के बाद से पुल के नीचे बहुत सारा पानी बह गया है, लेकिन एक [डिक्शनरी अटैक](#) (या वर्डलिस्ट रिकवरी) अभी भी ऐसे अधिकांश कार्यक्रमों में उपयोग किए जाने वाले पासवर्ड हैकिंग के मूलभूत तरीकों में से एक है। हालाँकि, यदि 1988 में त्वरित प्रसार के लिए आवश्यक पहले [व्यापक रूप से ज्ञात कंप्यूटर वर्म](#) 400 सबसे अधिक उपयोग किए जाने वाले शब्द थे, तो वर्तमान में पासवर्ड रिकवरी यूपिलिटी के लिए बड़े और अधिक जटिल शब्दकोशों की आवश्यकता होती है।

कस्टम वर्डलिस्ट बनाना

2 कस्टम वर्डलिस्ट बनाना

आप एक प्रोफेशनल रूप से लिखित पासवर्ड रिकवरी प्रोग्राम को कैसे पहचान सकते हैं? यह वास्तव में बहुत सरल है। यदि प्रोग्राम इंस्टॉलेशन किट या वेबसाइट पर अद्वितीय डिक्शनरी हैं, तो आप निश्चित रूप से सुनिश्चित हो सकते हैं कि यह एक प्रोफेशनल प्रोग्राम है। शौकिया प्रोग्रामर आमतौर पर नए शब्दकोश बनाने की जहमत नहीं उठाते हैं और बस मौजूदा का उपयोग करते हैं; विशेष रूप से, क्योंकि उनमें से बहुत सारे इंटरनेट पर हैं। सबसे आश्चर्यजनक बात यह है कि अधिकांश बड़ी डाउनलोड करने योग्य वर्डलिस्ट में बहुत अधिक अनुपयोगी कचरा होता है और उनके निर्माता उन्हें साफ करने में समय नहीं लगाते हैं, बस सभी मौजूदा सामान और फाइलों को एक विशाल ढेर में मिला देते हैं। परिणामस्वरूप, आपको HTML टेग्स के साथ १०-२० GB ट्रेश, अस्पष्ट टेक्स्ट के लंबे तार, और यहां तक कि अन्य शब्दकोशों के कॉपीराइट, जो गंभीर प्रोजेक्ट्स में अनुपयोगी हो जाते हैं।

गुणवत्ता, थीम वाली शब्द सूची बनाना एक जटिल और धन्यवाद रहित कार्य है - नई फाइलें बनाने में सारा जीवन लग जाता है। इस लेख में हम यह दिखाने की कोशिश करेंगे कि कैसे आप अपने स्वयं के अनोखे शब्दकोश बनाते समय कुछ नियमित संचालन को प्रभावी ढंग से स्वचालित कर सकते हैं। ऐसा करने के लिए आपको थोड़ी कल्पना और [शब्दकोश प्रसंस्करण के लिए युटिलिटीज के एक सेट](#) की आवश्यकता होगी। हम छोटी शुरुआत करेंगे और कई लोकप्रिय भाषाओं से अपना सेट एकत्र करने का प्रयास करेंगे। प्रत्येक शब्दकोश में सर्वाधिक लोकप्रिय शब्द होंगे। दूसरे भाग में हम शीघ्र ही 1,000,000 सर्वाधिक लोकप्रिय इंटरनेट डोमेन की एक सूची प्रस्तुत करेंगे। अंत में, तीसरे भाग में, हम एक विशाल (1.5 जीबी से अधिक) बनाएंगे, वास्तव में सबसे बड़ा जात, अद्वितीय शब्दकोश जिसमें आउटपुट के रूप में यूजर के नामों के 124 मिलियन से अधिक रिकॉर्ड होंगे। तो, आइये देखते हैं।

2.1 अंतरराष्ट्रीय शब्दकोश बनाना

हाल ही में, Android KitKat डिक्शनरी डेटाबेस का नवीनतम वर्जन प्रकाशित किया गया था। इन डेटाबेस का उपयोग शब्द ऑटो-कम्प्लेशन के लिए किया जाता है। उदाहरण के लिए, अंग्रेजी डेटाबेस में 150 हजार से अधिक शब्द हैं। हम उनका उपयोग अपनी खुद की अंग्रेजी और अंतरराष्ट्रीय शब्द-सूचियां बनाने के लिए कर सकते हैं। प्रत्येक डेटाबेस इस तरह के रिकॉर्ड के साथ एक कम्प्रेस्ड फ़ाइल है:

```
word=days,f=152,flags=,originalFreq=152
word=developed,f=152,flags=,originalFreq=152
word=east,f=152,flags=,originalFreq=152
word=election,f=152,flags=,originalFreq=152
word=formed,f=152,flags=,originalFreq=152
```

एक विशेष फ़िल्टर का उपयोग करके बिना जरूरी जानकारी के प्रत्येक रिकॉर्ड को साफ करना आसान है। आइए [डिक्शनरी क्रिएशन टूल](#) खोलें, अनज़िप्ड डेटाबेस फ़ाइल के लिए पाथ निर्दिष्ट करें, और विशेष नियमों के आधार पर फ़िल्टर विकल्प सेट करें। आइए नियम क्षेत्र में निम्नलिखित टेक्स्ट दर्ज करें:

```
=0 =1w=2o=3r=4d=5=e,E=
```

'Next' बटन पर क्लिक करें और शब्दकोश तैयार रखें। आइए विस्तार से समीक्षा करें कि यहां क्या हुआ। पहला भाग (**=0 =1w=2o=3r=4d=5=**) उन पंक्तियों को खारिज कर देता है जो हमें आवश्यक वर्ण अनुक्रम से शुरू नहीं होती हैं, यानी 'word=' के साथ। फिर 'e' कमांड कैरेक्टर सबस्ट्रिंग को फ़र्स्ट पोजीशन से फ़र्स्ट कॉमा में ले जाता है। फिर 'E=' बराबर चिह्न के बाद जो कुछ भी है, यानी हमें जिस शब्द की जरूरत है उसे ले लेता है। जैसा कि आप देख सकते

हैं, यह सब काफी सरल है। एक बार जब आप इसे बाकी फाइलों के लिए कर लेते हैं, तो आपके पास दो दर्जन से अधिक अंतरराष्ट्रीय शब्दकोश उपयोग के लिए तैयार हो जाते हैं।

2.2 ~ 1 मिलियन लोकप्रिय इंटरनेट डोमेन की सूची बनाना।

आर्काइव पासवर्ड रिकवर करने में, अक्सर ऐसे पासवर्ड होते हैं जो URL एड्रेस या डोमेन के नाम की तरह दिखते हैं। सबसे लोकप्रिय डोमेन नामों के साथ एक शब्दकोश क्यों न बनाएं; आखिर नेट पर ऐसा कोई शब्दकोश नहीं है? कहते ही काम हो जाना। हमें केवल 1 मिलियन लोकप्रिय डोमेन के [22 MB डेटाबेस](#) की आवश्यकता है। डेटाबेस की प्रत्येक पंक्ति में लोकप्रियता सूची में स्थिति और अल्पविराम द्वारा अलग किया गया एक URL पता होता है। सरलतम कमांड 'E' दें और हमें पदों के बिना डोमेन की एक साफ सूची मिलती है। यदि हमें [देश-कोड शीर्ष-स्तरीय डोमेन](#) (.org, .com, आदि) की आवश्यकता नहीं है, तो हम 'E,e.' के लिए फ़िल्टर में थोड़ा फेरबदल कर सकते हैं। उस स्थिति में, प्रसंस्करण के बाद हमें जो फ़ाइल मिलती है, उसे डुप्लिकेट से साफ करना होगा, क्योंकि, उदाहरण के लिए, स्रोत लाइनें google.com, google.com.hk, google.fr, आदि बस google बन जाएंगी। हमें लगभग 11 MB या 900,000 से अधिक नाम (देश के डोमेन के बिना) मिलेंगे। आप हमारी वेबसाइट से एक [मौजूदा शब्दकोश डाउनलोड कर सकते हैं](#) जिसमें 1.1 मिलियन से अधिक रिकॉर्ड 14 से 3.५ 3.5 MB तक कंप्रेस्ड हैं।

2.3 यूजर्स के नामों के साथ एक शब्दकोश बनाना

Adobe वेबसाइट को 2013 के अंत में हैक कर लिया गया था, जिसके परिणामस्वरूप रजिस्टर्ड यूजर्स के ई-मेल पते और एन्क्रिप्टेड पासवर्ड के साथ 150 मिलियन से अधिक एन्क्रिप्टेड रिकॉर्ड नेट पर लीक हो गए - बेचारा पुराना Adobe।

इस तथ्य के बावजूद कि एन्क्रिप्शन एल्गोरिथम काफी पुराना था (कंपनी के अनुसार 3DES), यह यूजर्स के लिए काफी प्रोफेशनल और सौभाग्य से उपयोग किया गया था, उनके पासवर्ड अभी तक डिकोड नहीं किए गए हैं। आइए ई-मेल पतों के पहले भाग को निकालकर उपयोक्ता नामों का एक शब्दकोश बनाने का प्रयास करें। वास्तव में, ऐसे शब्दकोश प्रोग्राम में पासवर्ड रिकवरी के लिए बहुत अच्छे हैं जो पासवर्ड म्यूटेशन के विशेष नियमों का उपयोग करते हैं।

लीक पासवर्ड डेटाबेस 9 GB से अधिक है और इसमें ऐसे रिकॉर्ड हैं जो लगभग इस तरह दिखते हैं:

```
103239615-|-|techlaf@yahoo.com-|-|getc/eE3AqY=-|-|normal|--
103239616-|-|sepro777@yahoo.co.jp-|-|QBk2p9UC6XA=-|-|--
103239617-|-|jack_galante@yahoo.com-|-|wiNNHCMhG64=-|-|coche|--
```

सबसे पहले, हमें उस टेक्स्ट को काट देना चाहिए जिसकी हमें आवश्यकता नहीं है और केवल ई-मेल पता डेटाबेस छोड़ दें। आइए डिक्वेशनरी जनरेशन यूटिलिटी लॉन्च करें और फ़िल्टर के रूप में निम्न कमांड दर्ज करें:

```
%1|E|E|e||e@
```

यह क्या करता है विस्तार से:

% 1 | - उन पंक्तियों को छोड़ दें जिनमें '|' नहीं है प्रतीक

E|E|e | - ई-मेल पते को छोड़कर सभी फ़िल्ड हटाएं

|| - पहले और अंतिम प्रतीक को काट दें (प्रत्येक ई-मेल पते के आरंभ में और अंत में एक '-' चिह्न होता है)।

e@ - '@' के बाद सब कुछ हटा दें

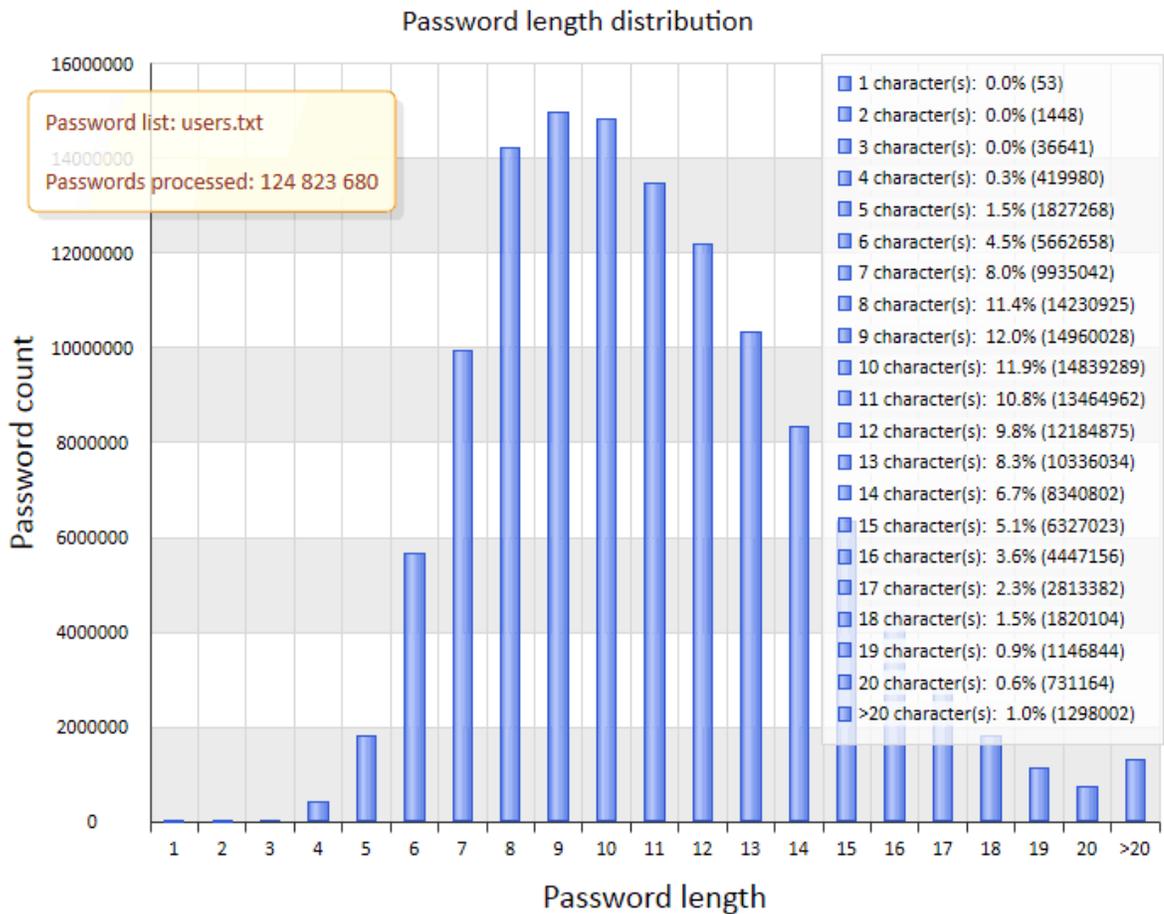
एक उदाहरण के रूप में, आइए देखें कि इनपुट शब्द पर प्रत्येक नियम लागू होने के बाद क्या होगा:

इनपुट शब्द: 103239615-|-|techlaf@yahoo.com-|-|getc/eE3AqY=-|-|normal|--

इनपुट नियम: %1|E|E|e|[e@

नियम	परिणाम आउटपुट
%1	103239615- - techlaf@yahoo.com- getc/eE3AqY=- normal --
E	-- techlaf@yahoo.com- getc/eE3AqY=- normal --
E	-techlaf@yahoo.com- getc/eE3AqY=- normal --
e	-techlaf@yahoo.com-
	-techlaf@yahoo.com
[techlaf@yahoo.com
e@	techlaf

एक बार इनपुट वर्डलिस्ट के लिए कमांड निष्पादित हो जाने के बाद, हमें यूजरनेम की एक सूची मिलती है जिसे डुप्लिकेट से छुटकारा पाने के लिए हमें सॉर्ट करना होता है। डुप्लिकेट को हटाने के बाद भी सूची काफी प्रभावशाली है - 124 मिलियन से अधिक शब्द। अंतिम शब्दकोश के कुछ दिलचस्प आंकड़ों के लिए नीचे देखें।



Character set exclusivity (all passwords)

