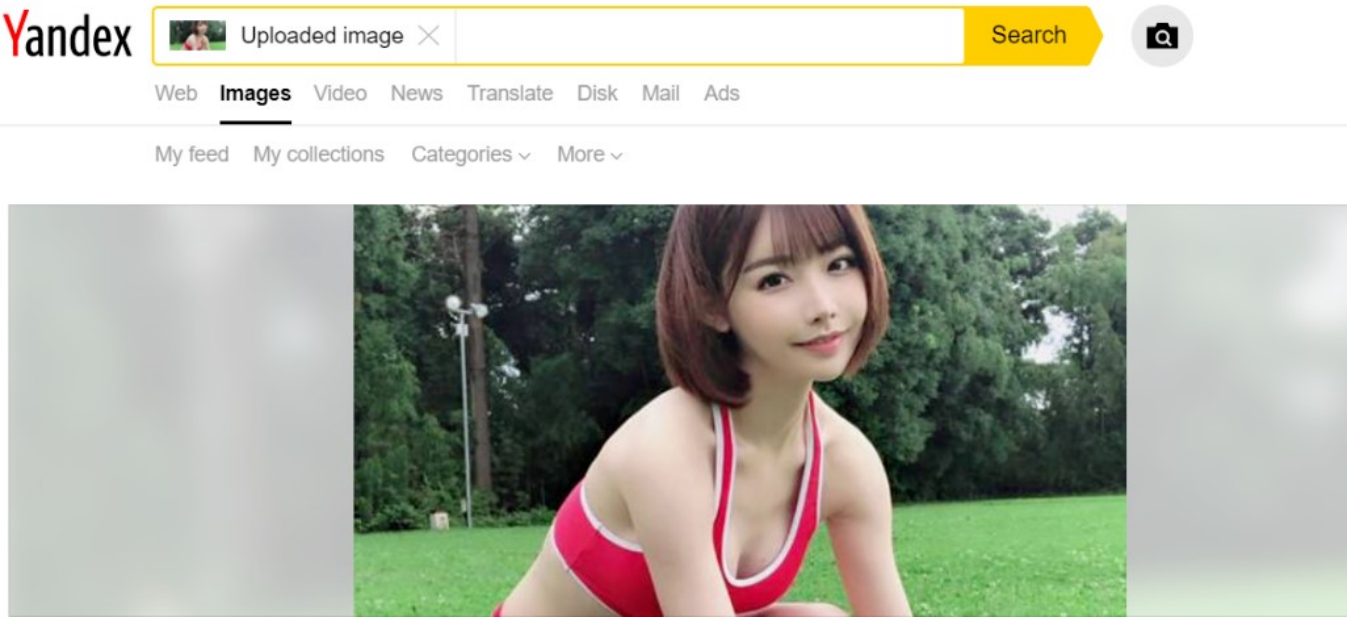
俄版百度Yandex44.7GB源代码被前员工泄露！

老司机们的这一生，应该会见到很多日本小姐姐。但许多时候，仅凭一张截图或者海报，是很难找到她们的作品的。在这种情况下，Yandex的以图识图功能，往往能发挥令人意想不到的作用。

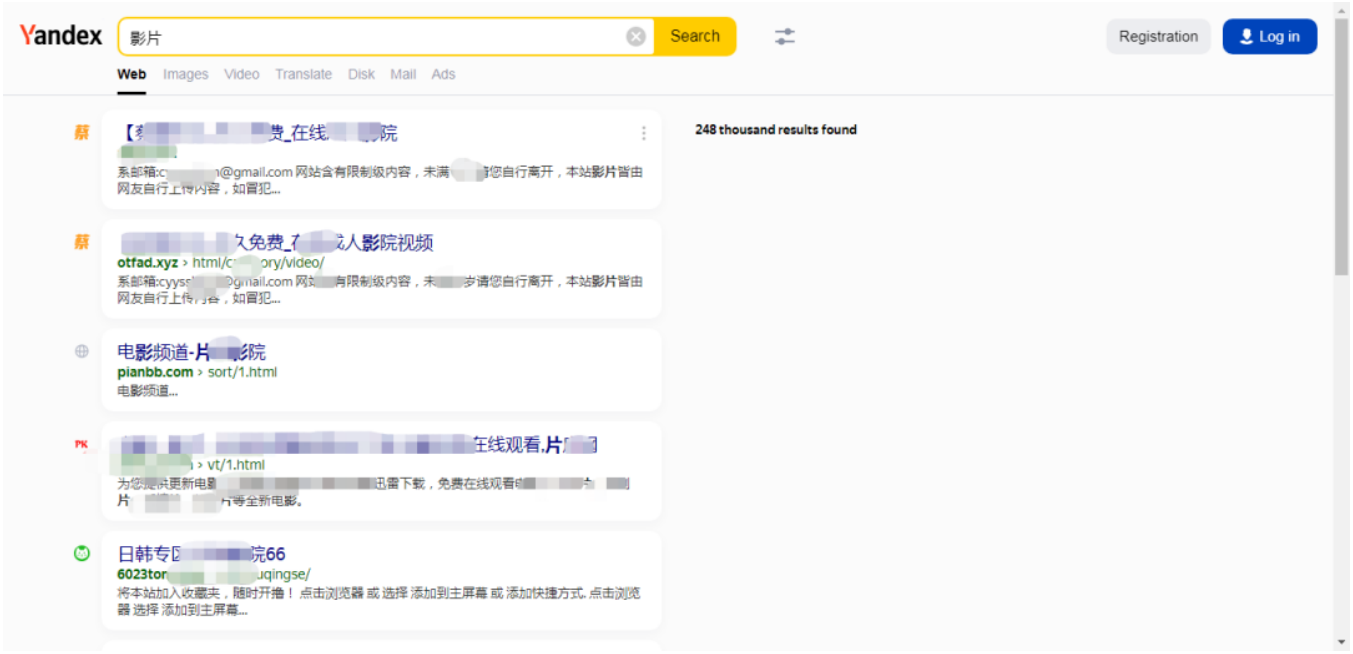


可以说，它是ok酱用过的搜索引擎里，搜图功能最强大的一个。

另外，身为一款来自俄罗斯的搜索引擎，用它搜索中文也会有惊喜。

不少中国网友都将Yandex视为珍宝，隔三差五搜一搜。

不怕搜不到，就怕营养跟不上。



话又说回来，Yandex在俄罗斯本土，也一直比谷歌更受欢迎。

除了搜索业务外，Yandex旗下还有几十种不同应用。

邮箱、网购、打车、外卖、地图、翻译.....服务范围涵盖几乎所有行业。

在俄罗斯，它等同于百度+淘宝+美团+滴滴+网易云+支付宝等等，重要性不言而喻。

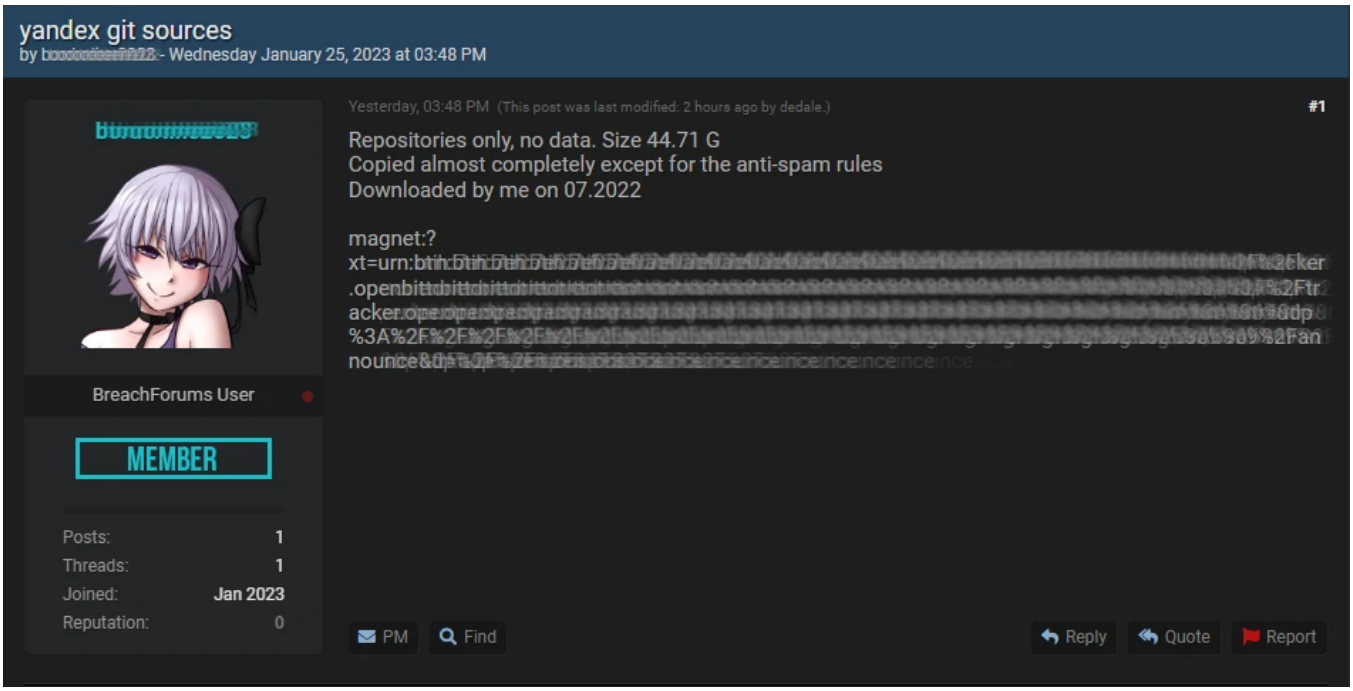
然而，就在最近，这家俄罗斯第一大科技巨头，却遇上了大麻烦。

1月25日，一条磁力链接出现在黑客论坛上，很快在全网炸开了锅。

泄露者宣称，这是“ Yandex git sources ”， 该源码库包含了Yandex除反垃圾邮件规则之外的全部源代码。

包括搜索引擎和索引爬虫，地图、AI助手、打车服务、广告服务、邮件、网盘、云服务和支付服务等等产品。

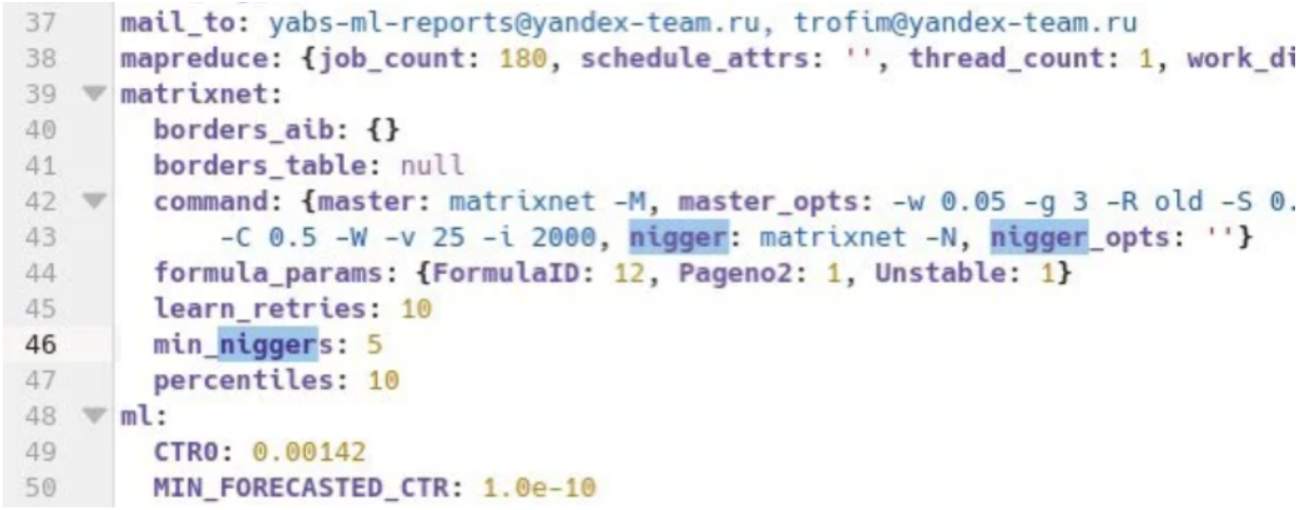
泄露者还表示，源代码是在去年7月窃取到的，文件总容量为44.7GB。



这一大瓜曝出来后，引起了全球网友的围观，相关词条迅速登上谷歌、推特的热搜。

各路大神将源代码扒了个底朝天，在这个过程中，Yandex的一些内部“潜规则”也被翻了出来。

最让人吃惊的是，有网友发现，Yandex的开发团队在编写代码时竟然会使用种族歧视性词语。



```
mail_to: yabs-ml-reports@yandex-team.ru, trofim@yandex-team.ru
mapreduce: {job_count: 180, schedule_attrs: '', thread_count: 1, work_di
matrixnet:
  borders_aib: {}
  borders_table: null
  command: {master: matrixnet -M, master_opts: -w 0.05 -g 3 -R old -S 0.
      -C 0.5 -W -v 25 -i 2000, nigger: matrixnet -N, nigger_opts: ''}
  formula_params: {FormulaID: 12, Pageno2: 1, Unstable: 1}
  learn_retries: 10
  min_niggers: 5
  percentiles: 10
ml:
  CTR0: 0.00142
  MIN_FORECASTED_CTR: 1.0e-10
```

这样的情况并非偶然，而是多次出现——有些词语还被用于变量名称中。

因此很有可能，开发团队是故意这么命名的。

比如，Yandex的开发者添加的注释是kill n*gger。

执行时这个函数就会在屏幕上显示：“ 请稍后我正在杀死n*gger ”。

```
66
67  sky run --hosts=$HOSTS_FILE "mkdir -p $WORK_DIR"
68  sky upload --hosts=$HOSTS_FILE -d $PROGRAM_PATH $PROGRAM_FILE $WORK_DIR/
69
70  stop_niggers()
71  {
72      echo Stop
73      echo Please wait until all niggers are terminated.
74      sky run --hosts=$HOSTS_FILE "pkill -f ^$WORK_DIR/$PROGRAM_FILE"
75      if [ -z $1 ]; then
76          exit 1
77      fi
78  }
79
80  trap stop_niggers SIGINT
81
82  ATTEMPT=0;
83
84  while [ ! -f ${OPERATION}.inc  -a  $ATTEMPT -lt $NUM_ATTEMPTS ]; do
85      ATTEMPT=$(($ATTEMPT+1))
```

眼看舆论愈演愈烈，Yandex官方也按捺不住，出面发表声明称：

他们的系统并未遭受黑客入侵，泄露源代码仓库的是一名前员工。

泄露源代码的人，并没有采取什么高明的技术手段，只是利用职务之便打包下载转存的。

至于代码中包含的种族歧视词语，Yandex称，使用这些词语不影响公司的服务，而且仅在内部使用。

不过，Yandex也依然对开发团队违反公司政策使用这类词语表示歉意。

虽然并未遭到网络入侵，但从员工能够直接转存整个公司的产品源码这一点来看，Yandex在内部管理上显然是有大问题的。

目前，这位员工窃取并公开源码的原因还暂不明确。

不过，Yandex的技术专家Bakunov解释称，此人泄露数据的动机或与政治有关。

因为这位员工并未试图将代码出售给商业竞争对手，而是直接放在了网上。

Bakunov补充道，泄露内容没有包含任何客户数据，不会对用户的隐私或安全构成直接风险，也不会直接威胁到Yandex的专有技术。

不过一些文件仍可能会暴露正在运行的服务，比如说“ blacklist.txt ”。

Bakunov还称，尽管泄密的部分不涉及敏感数据，但黑客针对性利用代码中的安全漏洞，只是时间问题。

换言之，Yandex增加了黑客暴露的风险，接下来可能会遭到大量攻击。

近几年来，由于缺乏管理约束，大厂内部员工“ 删库跑路 ”导致的安全事故早已屡见不鲜。

Yandex此次遭遇的重大泄露事件，也再次给各个科技公司敲响了警钟。

平时提到网络安全，大家一再强调如何抵御外部的攻击，殊不知来自内部的安全隐患才是最让人措不及防的。

本文链接：https://dqcm.net/zixun/16756795447227.html