

Queste de savoir

Statistiques sur les inscriptions,
bannissements et signalements

samedi 04 janvier 2025

Table des matières

Introduction	1
1. Analyse des inscriptions	1
2. Analyse des inscriptions via les réseaux sociaux	2
3. Analyse des bannissements permanents	2
4. Analyse des signalements	3
5. Analyse des signalements émis par l'antispam des biographies	4
Conclusion	4
Contenu masqué	4

Introduction

On est plusieurs à avoir constaté une augmentation du spam sur Zeste de Savoir ces dernières années. J'ai eu envie d'estimer l'ampleur de ce fléau donc j'ai fait quelques statistiques !

1. Analyse des inscriptions

Comptes créés...	depuis la création du site	en 2014	en 2015	en 2016	en 2017
Comptes utilisateurs	30328 (100 %)	1526 (100 %)	2465 (100 %)	2629 (100 %)	2736 (100 %)
Comptes utilisateurs activés	27017 (89 %)	1430 (94 %)	2243 (91 %)	2373 (90 %)	2480 (91 %)
Comptes utilisateurs activés et bannis	3726 (12 %)	8 (1 %)	75 (3 %)	52 (2 %)	49 (2 %)

TABLE 1.2. – Analyse des inscriptions

Les inscriptions sont relativement stables depuis 2015 avec une moyenne 2900 comptes créés chaque année. Environ 10% de ces comptes sont inactifs, c'est-à-dire que l'adresse de courriel n'a pas été validée par l'utilisateur en cliquant sur le lien reçu par courriel. Je n'imaginais pas que ce nombre soit aussi élevé ! Enfin, on observe depuis 2018 une croissance des comptes bannis de façon permanente. Ils représentent un quart des inscriptions ces deux dernières années !

2. Analyse des inscriptions via les réseaux sociaux

2. Analyse des inscriptions via les réseaux sociaux

Comptes créés...	depuis la création du site	en 2014	en 2015	en 2016	en 2017
<i>Utilisateurs inscrits</i>	27017	1430	2243	2373	2480
Utilisateurs inscrits via les réseaux sociaux	11162 (100 %)	31 (100 %)	518 (100 %)	898 (100 %)	1221 (100 %)
Utilisateurs inscrits via les réseaux sociaux et bannis	1224 (11 %)	0 (0 %)	9 (2 %)	14 (2 %)	9 (1 %)
Utilisateurs inscrits via Google	8706 (78 %)	26 (84 %)	502 (97 %)	616 (69 %)	684 (56 %)
Utilisateurs inscrits via Facebook	2446 (22 %)	5 (16 %)	6 (1 %)	282 (31 %)	537 (44 %)
Utilisateurs inscrits via Twitter ¹	10 (0 %)	0 (0 %)	10 (2 %)	0 (0 %)	0 (0 %)

TABLE 2.4. – Analyse des inscriptions via les réseaux sociaux

Environ 40 % des comptes créés sur le site l'ont été via les réseaux sociaux et la moitié des inscriptions se font via les réseaux sociaux ces dernières années. 80 % de ces comptes proviennent de Google et 20 % de Facebook.

On retrouve dans ce tableau les mêmes proportions de comptes bannis de façon permanente que dans le tableau précédent : 10 % des comptes depuis la création du site et environ 25 % ces deux dernières années.

3. Analyse des bannissements permanents

Bannissements de comptes créés...	depuis la création du site	en 2014	en 2015	en 2016	en 2017
Bannissements permanents	3778 ² (100 %)	7 (100 %)	86 (100 %)	52 (100 %)	49
Bannissements permanents pour spam ¹	3341 (88 %)	5 (71 %)	29 (34 %)	29 (56 %)	34

1. La possibilité de s'inscrire via Twitter a été enlevée en février 2015.

4. Analyse des signalements

Bannissements permanents pour spam ¹ (moins de 7 jours après l'inscription)	2855 (76 %)	3 (43 %)	28 (33 %)	21 (40 %)	23
Bannissements permanents pour spam ¹ (moins de 1 jour après l'inscription)	2624 (69 %)	2 (29 %)	23 (27 %)	21 (40 %)	21
Bannissements permanents pour spam ¹ (moins de 1 heure après l'inscription)	1371 (36 %)	2 (29 %)	18 (21 %)	17 (33 %)	11

TABLE 3.6. – Analyse des bannissements permanents

Ces deux dernières années, plus de 95 % des bannissements le sont pour cause de spam (publicité, arnaque...) avec une très bonne réactivité de la part de l'équipe du site car 80 % de ces bannissements ont lieu moins d'un jour après l'inscription et 45 % moins d'une heure après celle-ci !

4. Analyse des signalements

Signalements émis...	depuis la création du site	en 2016	en 2017	en 2018	en 2019
Signalements	6017 (100 %)	58 (100 %)	553 (100 %)	529 (100 %)	557 (100 %)
Signalements de profils	1242 (21 %)	0 (0 %)	0 (0 %)	0 (0 %)	50 (9 %)
Signalements de messages sur le forum	3901 (65 %)	57 (98 %)	473 (86 %)	441 (83 %)	437 (78 %)
Signalements de contenus	122 (2 %)	0 (0 %)	34 (6 %)	25 (5 %)	20 (4 %)
Signalements de commentaires	752 (12 %)	1 (2 %)	46 (8 %)	63 (12 %)	50 (9 %)

TABLE 4.8. – Analyse des signalements

1. Je n'ai gardé ici que les bannissements dont l'explication donnée par le modérateur contient le mot 'spam'.

2. Il y a bien 3778 bannissements permanents mais seulement 3726 comptes utilisateurs activés et bannis (cf. le premier tableau du billet). Un bannissement correspond ici à l'action de bannir et non pas à l'état actuel d'un compte utilisateur. Cette différence numérique s'explique donc par la possibilité d'enlever un bannissement (puis éventuellement de le remettre... puis de l'enlever à nouveau...).

5. Analyse des signalements émis par l'antispam des biographies

Plus de la moitié des signalements émis concernent des messages du forum et leur évolution est stable avec environ 450 signalements chaque année depuis 2015. Les signalements de profils et de commentaires ont eux nettement augmentés ces trois dernières années.

5. Analyse des signalements émis par l'antispam des biographies

Signalements émis...	depuis la création du site	en 2019	en 2020	en 2021
Signalements de profils	1242 (100 %)	50 (100 %)	91 (100 %)	70 (100 %)
Signalements de profils par l'antispam des biographies	783 (63 %)	0 (0 %)	0 (0 %)	0 (0 %)
Signalements de profils par l'antispam des biographies suivi d'un bannissement	775 (62 %)	0 (0 %)	0 (0 %)	0 (0 %)

TABLE 5.10. – Analyse des signalements émis par l'antispam des biographiques

Depuis juin 2020, l'[antispam des biographies](#) essaie de détecter le spam dans les biographies des comptes créés récemment et notifie les modérateurs lorsqu'une biographie lui semble suspecte. Au début, la notification se faisait par message privé à un groupe de modérateurs, mais depuis janvier 2022 le script utilise la fonction "Signaler" des profils. S'il est difficile d'estimer le nombre de faux négatifs, il n'y a en tous cas que très peu de faux positifs !

Conclusion

🔗 Code source du script Python qui a extrait ces statistiques de la base de données

Contenu masqué

Contenu masqué n°1 :

Code source du script Python qui a extrait ces statistiques de la base de données

```
1 from datetime import timedelta
2
3 from django.conf import settings
4 from django.contrib.auth.models import User, Group
5 from django.db.models import ExpressionWrapper, F, DurationField
6 from social_django.models import UserSocialAuth
7
```

```

8 from zds.member.models import Profile, Ban
9 from zds.utils.models import Alert
10
11 groupe_des_robots =
12     Group.objects.get(name=settings.ZDS_APP["member"]["bot_group"])
13
14 antisпам_user =
15     User.objects.get(username="Antispam des biographies")
16
17 def print_to_file(text):
18     with open("/opt/zds/table_output.md", "a") as f:
19         f.write(text)
20
21 def print_table(base, abscisses, ordonnées, coin=" ", titre=None):
22     table = "\n" + " | ".join([coin] + list(abscisses.keys())) +
23         "\n" + "|".join(["---"]*(len(abscisses)+1)) + "\n"
24     for index_y, (titre_y, filtre_y) in
25         enumerate(ordonnées.items()):
26         table += titre_y
27         if index_y == 0: # première ligne
28             valeurs_max = []
29             for index_x, filtre_x in enumerate(abscisses.values()):
30                 valeur = filtre_y(filtre_x(base)).count()
31                 if index_y == 0: # première ligne
32                     valeurs_max += [valeur]
33             try:
34                 pourcentage = round(valeur / valeurs_max[index_x]
35                                     * 100)
36             except ZeroDivisionError:
37                 pourcentage = 0
38             table += f" | {valeur} ({pourcentage} %)"
39         table += "\n"
40     if titre:
41         table += f"Table: {titre}\n"
42     table += "\n\n"
43     print_to_file(table)
44
45 ### Analyse des comptes utilisateurs
46
47 base =
48     Profile.objects.exclude(user__groups__in=[groupe_des_robots])
49 abscisses = {
50     "depuis la création du site": lambda x: x
51 } | {
52     f"en {année}": lambda x, année=année:
53         x.filter(user__date_joined__year=année) for année in
54         range(2014, 2025)
55 }
56
57 ordonnées = {
58     "Comptes utilisateurs": lambda x: x,

```

```

50     "Comptes utilisateurs activés[^utilisateurs-activés]": lambda
51         x: x.filter(user__is_active=True),
52     }
53     coin = "Comptes créés..."
54     titre = "Analyse des comptes utilisateurs"
55
56     print_table(base, abscisses, ordonnées, coin, titre)
57
58     print_to_file("\n
59         [^utilisateurs-activés]: Les comptes utilisateurs ne sont activés qu'une fo
60         \n")
61
62     ## Utilisateurs inscrits via les réseaux sociaux
63
64     base = UserSocialAuth.objects.filter(user__is_active=True)
65     abscisses = {
66         "depuis la création du site": lambda x: x
67     } | {
68         f"en {année}": lambda x, année=année:
69             x.filter(user__date_joined__year=année) for année in
70             range(2014, 2025)
71     }
72     ordonnées = {
73         "Utilisateurs inscrits via les réseaux sociaux": lambda x: x,
74         "Utilisateurs inscrits via les réseaux sociaux et bannis":
75             lambda x: x.filter(user__profile__can_read=False).filter(u
76                 ser__profile__end_ban_read=None),
77         "Utilisateurs inscrits via Google": lambda x:
78             x.filter(provider="google-oauth2"),
79         "Utilisateurs inscrits via Facebook": lambda x:
80             x.filter(provider="facebook"),
81         "Utilisateurs inscrits via Twitter[^twitter]": lambda x:
82             x.filter(provider="twitter"),
83     }
84     coin = "Comptes créés..."
85     titre = "Analyse des utilisateurs inscrits via les réseaux sociaux"
86
87     print_table(base, abscisses, ordonnées, coin, titre)
88
89     print_to_file("\n
90         [^twitter]: La possibilité de s'inscrire via Twitter a été enlevée en févr
91         \n")
92
93     ### Analyse des bannissements permanents
94

```



```
84 base = Ban.objects.annotate(depuis_inscription=ExpressionWrapper(F_
    ("pubdate") - F("user__date_joined"),
    output_field=DurationField())).filter(type__in=(
    "Bannissement illimité",
    "Ban définitif"))
85 abscisses = {
86     "depuis la création du site": lambda x: x
87 } | {
88     f"en {année}": lambda x, année=année:
        x.filter(user__date_joined__year=année) for année in
        range(2014, 2025)
89 }
90 ordonnées = {
91     "Bannissements permanents": lambda x: x,
92     "Bannissements permanents pour spam[^spam]": lambda x:
        x.filter(note__icontains="spam"),
93     "Bannissements permanents pour spam[^spam] (moins de 7 jours après l'in
        : lambda x:
        x.filter(note__icontains="spam").filter(depuis_inscription_
        __lte=timedelta(days=7)),
94     "Bannissements permanents pour spam[^spam] (moins de 1 jour après l'in
        : lambda x:
        x.filter(note__icontains="spam").filter(depuis_inscription_
        __lte=timedelta(days=1)),
95     "Bannissements permanents pour spam[^spam] (moins de 1 heure après l'in
        : lambda x:
        x.filter(note__icontains="spam").filter(depuis_inscription_
        __lte=timedelta(hours=1)),
96 }
97 coin = "Bannissements de comptes créés..."
98 titre = "Analyse des bannissements permanents"
99
100 print_table(base, abscisses, ordonnées, coin, titre)
101
102 print_to_file("\n_
    [^spam]: Je n'ai gardé ici que les bannissements dont l'explication donnée
    \n")
103
104 ### Analyse des signalements
105
106 base = Alert.objects
107 abscisses = {
108     "depuis la création du site": lambda x: x
109 } | {
110     f"en {année}": lambda x, année=année:
        x.filter(pubdate__year=année) for année in range(2016,
        2025)
```

```
111 }
112 ordonnées = {
113     "Signalements": lambda x: x,
114     "Signalements de profils": lambda x: x.filter(scope="PROFILE"),
115     "Signalements de messages sur le forum": lambda x:
116         x.filter(scope="FORUM"),
117     "Signalements de contenus": lambda x:
118         x.filter(scope="CONTENT"),
119     "Signalements de commentaires": lambda x:
120         x.filter(scope__in=("TUTORIAL", "ARTICLE", "OPINION")),
121 }
122 coin = "Signalements émis..."
123 titre = "Analyse des signalements"
124 print_table(base, abscisses, ordonnées, coin, titre)
125
126 ## Signalements par l'antispam des biographies depuis sa mise en
127 place en juin 2020
128
129 base = Alert.objects.filter(scope="PROFILE")
130 abscisses = {
131     "depuis la création du site": lambda x: x
132 } | {
133     f"en {année}": lambda x, année=année:
134         x.filter(pubdate__year=année) for année in range(2019,
135             2025)
136 }
137 ordonnées = {
138     "Signalements de profils": lambda x: x,
139     "Signalements de profils par l'antispam des biographies":
140         lambda x: x.filter(author=antispam_user),
141     "Signalements de profils par l'antispam des biographies suivi d'un ban
142         : lambda x:
143         x.filter(author=antispam_user).filter(profile__can_read=False)
144         .filter(profile__end_ban_read=None),
145 }
146 coin = "Signalements émis..."
147 titre = "
148     "Analyse des signalements émis par l'antispam des biographiques"
149
150 print_table(base, abscisses, ordonnées, coin, titre)
151
152 print_to_file("\n[[information]]\n
153     | L'antispam des biographies a été mis en place en juin 2020 mais le signa
154     )
155
```

[Retourner au texte.](#)