

Combo - Bug #9784

interdire googlebot sur les tentatives de callback

27 janvier 2016 10:08 - Thomas Noël

Statut:	Rejeté	Début:	27 janvier 2016
Priorité:	Normal	Echéance:	
Assigné à:		% réalisé:	0%
Catégorie:		Temps estimé:	0:00 heure
Version cible:		Planning:	
Patch proposé:	Non		
Description			
<p>Il arrive qu'on reçoive des requêtes Googlebot sur les URL de callback de lingo, très certainement parce qu'en cas d'erreur ces URL sont envoyées par les opérateurs à des adresses des régisseurs.</p> <p>Les URLs étant mauvaises, ça plante.</p> <pre>Internal Server Error: /lingo/callback/4/ Traceback (most recent call last): File "/usr/lib/python2.7/dist-packages/django/core/handlers/base.py", line 111, in get_response response = wrapped_callback(request, *callback_args, **callback_kwargs) File "/usr/lib/python2.7/dist-packages/django/views/generic/base.py", line 69, in view return self.dispatch(request, *args, **kwargs) File "/usr/lib/python2.7/dist-packages/django/views/decorators/csrf.py", line 57, in wrapped_view return view_func(*args, **kwargs) File "/usr/lib/python2.7/dist-packages/combo/apps/lingo/views.py", line 222, in dispatch return super(CallbackView, self).dispatch(*args, **kwargs) File "/usr/lib/python2.7/dist-packages/django/views/generic/base.py", line 87, in dispatch return handler(request, *args, **kwargs) File "/usr/lib/python2.7/dist-packages/combo/apps/lingo/views.py", line 215, in get return self.handle_callback(request, request.environ['QUERY_STRING'], **kwargs) File "/usr/lib/python2.7/dist-packages/combo/apps/lingo/views.py", line 182, in handle_callback payment_response = payment.response(backend_response) File "/usr/lib/python2.7/dist-packages/eopayment/__init__.py", line 138, in response return self.backend.response(query_string, **kwargs) File "/usr/lib/python2.7/dist-packages/eopayment/paybox.py", line 245, in response if d['erreur'][0] == '00000': KeyError: 'erreur' Request repr(): <WSGIRequest path:/lingo/callback/4/, GET:<QueryDict: {}>, POST:<QueryDict: {}>, COOKIES:{}, META:{u'CSRF_COOKIE': u'oL4k9YtGuLhD51BYNcgLPkvIFZmWXngp', 'HTTP_ACCEPT': 'text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8', 'HTTP_ACCEPT_ENCODING': 'gzip,deflate', 'HTTP_CONNECTION': 'close', 'HTTP_FROM': 'googlebot (at)googlebot.com', 'HTTP_HOST': 'mesdemarches.fontenay-sous-bois.fr', 'HTTP_USER_AGENT': 'Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/bot.html)'}></pre>			
Solution :			
<ul style="list-style-type: none">1 : solidifier un peu les checks dans le code, genre ne pas lire d['erreur'] sans vérifier que 'erreur' est dans le dico ; faire des 400 Bad Request dans ce cas.2 : peut-être aussi violemment interdire les accès quand "Googlebot" est présent dans le USER AGENT (pour le jour où ils taperait sur une URL valide...)			
Demandes liées:			
Lié à EOPayment - Development #9729: Vérifier la conformité d'une réponse ava...		Fermé	21 janvier 2016

Historique

#1 - 27 janvier 2016 11:11 - Benjamin Dauvergne

- Lié à *Development #9729*: Vérifier la conformité d'une réponse avant son traitement et émettre une *ResponseError* en cas de non conformité ajouté

#2 - 27 janvier 2016 11:14 - Benjamin Dauvergne

Je ne suivrai pas la voie 2 et je travaillerai principalement sur la validation correcte des requêtes entrantes dans eopayment, transformé en exception propre et rapporté par lingo par du 400 et des logs. J'ose penser que GoogleBot ne s'acharne pas quand il reçoit du 40x; maintenant on peut aussi réfléchir à servir un robots.txt dans nos applications Django auto-généré depuis les motifs d'URL avec un marqueur sur les vues (@no_crawler).

#3 - 27 janvier 2016 11:31 - Thomas Noël

Benjamin Dauvergne a écrit :

Je ne suivrai pas la voie 2 et je travaillerai principalement sur la validation correcte des requêtes entrantes dans eopayment, transformé en exception propre et rapporté par lingo par du 400 et des logs. J'ose penser que GoogleBot ne s'acharne pas quand il reçoit du 40x; maintenant on peut aussi réfléchir à servir un robots.txt dans nos applications Django auto-généré depuis les motifs d'URL avec un marqueur sur les vues (@no_crawler).

On a vu des requêtes de bot sur des sites de test avec un robots.txt qui exclu tout...

Mais oui oui oui pour se concentrer sur la voie 1, c'est le vrai soucis à régler.

#4 - 31 janvier 2017 21:30 - Frédéric Péters

- *Statut changé de Nouveau à Rejeté*

Gestion de la validation des paramètres reçus par eopayment, et code adapté dans combo, dans [#10260](#).