Best Practices for Spam Management in Jive

Best Practices for Spam Management in Jive

The purpose of this guide is to provide a detailed list of recommended ways to configure Jive to best handle any incoming spam in your community. The content in this guide is based off of Jive's own product documentation and tailored to Jive customers who are experiencing a sudden increase in spam activity.

The Jive platform comes with a series of spam prevention methods built into the product. Depending on the type and volume of spam you're seeing, as well as the type of community you manage, there will be different recommendations for how to best configure your Jive's spam prevention tools.

You can also review a complete list of spam prevention tools in our product documentation here: https://docs.jivesoftware.com/jive/7.0/community_admin/index.jsp?topic=/com.jivesoftware.help.sbs.online/admin/PreventingSpam.html.

The guide is intended for all Jive customers. The recommendations here will apply to Jive 6, Jive 7, Jive 8, and Jive Cloud.

Recommended Spam Prevention Tools Available in Jive

The following methods are Jive's recommended tools for preventing spam. These items are ranked according to effectiveness:

1. **Link moderation**: Moderate content that contains external URLs that are not whitelisted
2. **Link moderation (points theshold):** Configure link moderation so that all content with web links by users who do not have a certain amount of points is automatically moderated

3. **Configuring Content moderation (points threshold):** Configure content moderation so that all content by users who do not have a certain amount of points is automatically moderated
4. **Keyword interceptor**: Block or moderate content that contains keywords or phrases common to the spammers
5. **New user moderation:** Moderate new user accounts and have a human review new accounts
6. **Message governor interceptor:** Limit the frequency of posts
7. **Abuse reporting:** Allow community members to flag spam posts for review
8. **Banning user accounts or IPs:** Prevent certain user accounts from logging in or users from IP addresses from accessing Jive

You can see a complete list of spam prevention tools available in the Jive product in our product documentation.

How to Use Jive's Spam Prevention Features

Each spam prevention tool is managed through Jive's admin console. You will find details here on where each tool is configured and how to set it up.

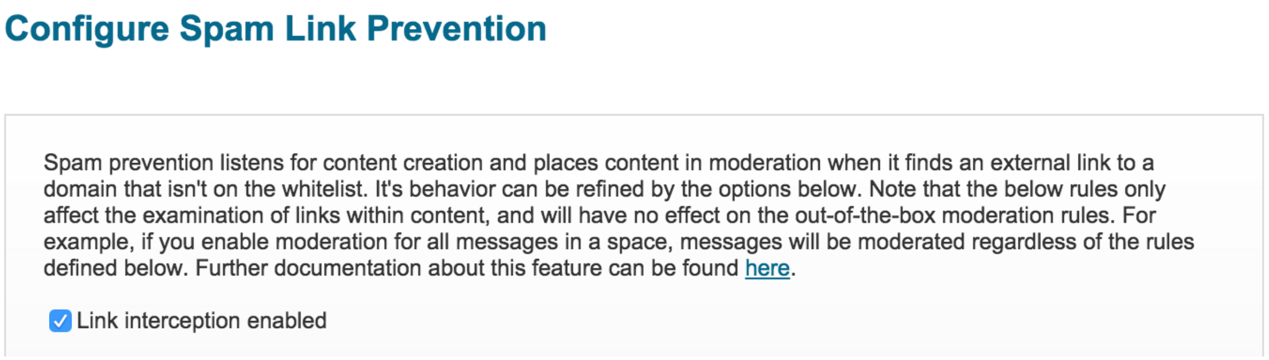
Configuring Content Moderation

Configuring Link moderation

Fastpath: Admin Console: System > Moderation > Configure Spam Prevention

Enabling Link moderation makes it so new content that contains a web link will automatically enter the moderation queue before it is publicly visible.

You can also also configure a list of domains that will bypass this link moderation functionality. It is recommended if your community members frequently link to sites that you trust.



Configure Spam Link Prevention

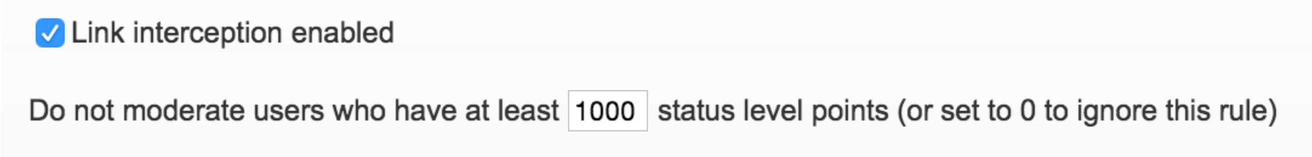
Spam prevention listens for content creation and places content in moderation when it finds an external link to a domain that isn't on the whitelist. It's behavior can be refined by the options below. Note that the below rules only affect the examination of links within content, and will have no effect on the out-of-the-box moderation rules. For example, if you enable moderation for all messages in a space, messages will be moderated regardless of the rules defined below. Further documentation about this feature can be found here.

☑ Link interception enabled

Configuring Link moderation (points threshold)

Fastpath: Admin Console: System > Moderation > Configure Spam Prevention

This tool allows you to configure Jive so that all **content created that includes a link** by users who do not have a certain amount of status points is automatically entered into moderation.



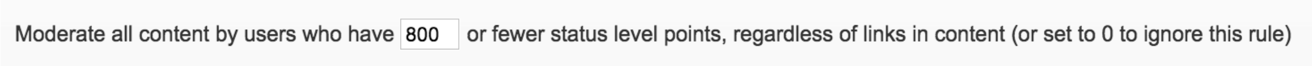
☑ Link interception enabled

Do not moderate users who have at least 1000 status level points (or set to 0 to ignore this rule)

Configuring Content moderation (points threshold)

Fastpath: Admin Console: System > Moderation > Configure Spam Prevention

This tool allows you to configure Jive so that all content created by users who do not have a certain amount of status points is automatically entered into moderation.



Moderate all content by users who have 800 or fewer status level points, regardless of links in content (or set to 0 to ignore this rule)

Advanced Gamification and Points
Jive Cloud: If you are using the Advanced Gamification module in **Jive Cloud** then the link moderation functionality will be based off of **Gamification points**.

Hosted and On-Premise 7.0.0 to 7.0.03: If you are using Advanced Gamification module in Jive 7.0.0 through 7.0.3 in Hosted or On-Premise, then the link moderation functionality will be based off of your **Status Points.** The link moderation functionality uses Jive's "status level points" and does not use Jive's Gamification point system.

Status level points are still accrued and applicable even if you are using Gamification, even if the status level points are not visible to your users.

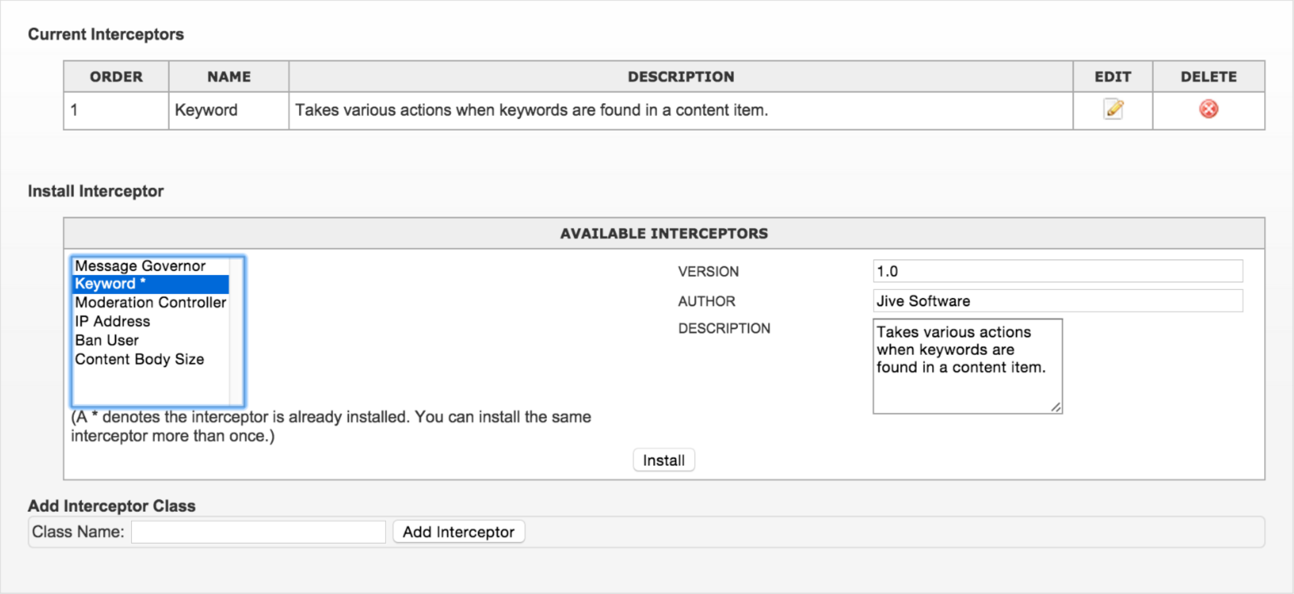
Starting in 7.0.4 and 8.0.0 Link Moderation will be using Gamification points, and not status points.

Configuring Keyword interceptor

Fastpath: Admin Console: System > Moderation > Moderation Configuration > Interceptors

The keyword interceptor allows you to prevent users from creating content that contains various keywords or phrases. You can either have these pieces of content go into moderation for an admin to approve, or you can block the creation of the content.

Tip: When configuring the keyword interceptor, **it is critical that you always copy and paste the keyword or phrase from the original spam post instead of retyping the text by hand**. The reason for this is that spammers will frequently use unusual or non-standard characters when posting content. If you re-type in the offending keyword or phrase you may not be correctly matching the targeted term.



Current Interceptors

ORDER	NAME	DESCRIPTION	EDIT	DELETE
1	Keyword	Takes various actions when keywords are found in a content item.	✎	✖

Install Interceptor

AVAILABLE INTERCEPTORS

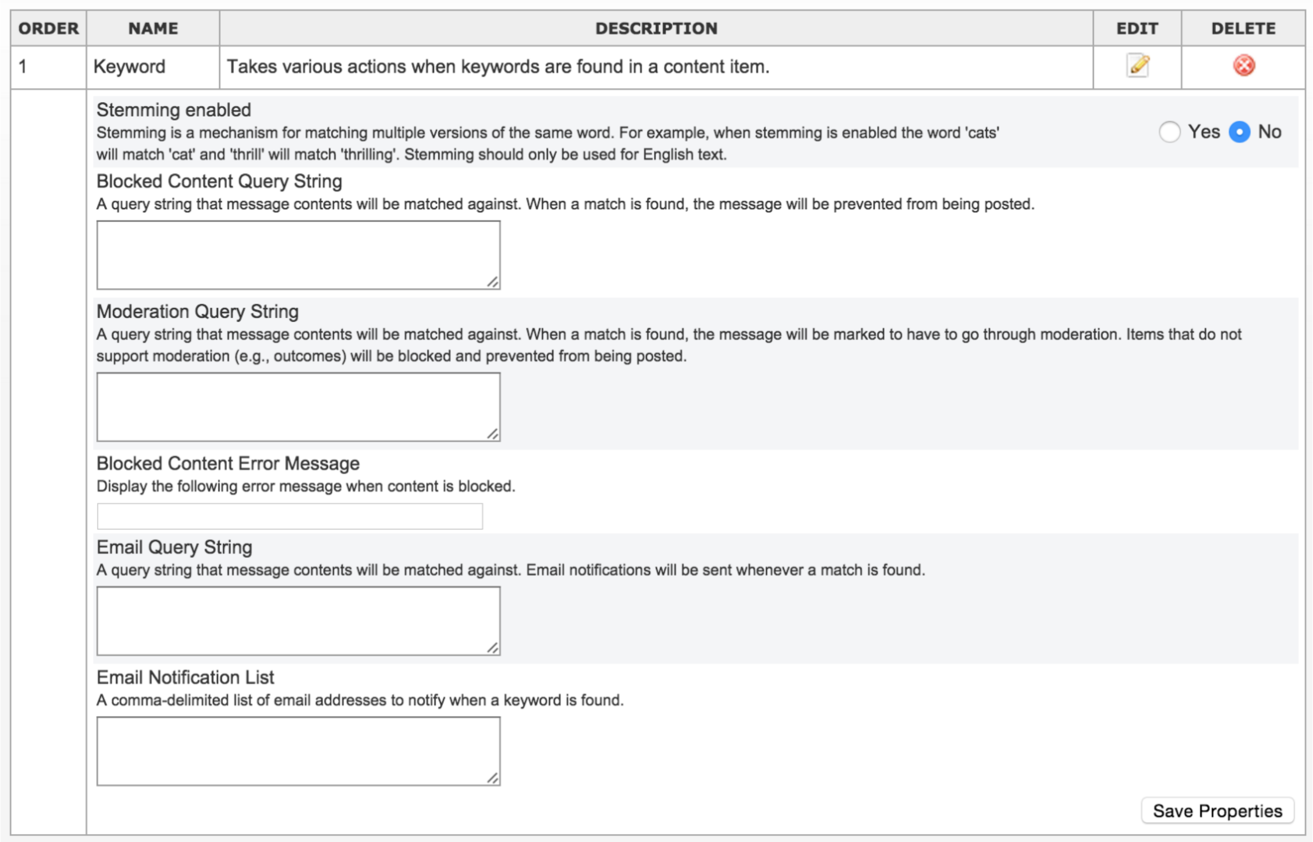
Message Governor
Keyword *
Moderation Controller
IP Address
Ban User
Content Body Size

VERSION	1.0
AUTHOR	Jive Software
DESCRIPTION	Takes various actions when keywords are found in a content item.

(A * denotes the interceptor is already installed. You can install the same interceptor more than once.)

Add Interceptor Class

Class Name:



ORDER	NAME	DESCRIPTION	EDIT	DELETE
1	Keyword	Takes various actions when keywords are found in a content item.	✎	✖

Stemming enabled
Stemming is a mechanism for matching multiple versions of the same word. For example, when stemming is enabled the word 'cats' will match 'cat' and 'thrill' will match 'thrilling'. Stemming should only be used for English text.
○ Yes ● No

Blocked Content Query String
A query string that message contents will be matched against. When a match is found, the message will be prevented from being posted.

Moderation Query String
A query string that message contents will be matched against. When a match is found, the message will be marked to have to go through moderation. Items that do not support moderation (e.g., outcomes) will be blocked and prevented from being posted.

Blocked Content Error Message
Display the following error message when content is blocked.

Email Query String
A query string that message contents will be matched against. Email notifications will be sent whenever a match is found.

Email Notification List
A comma-delimited list of email addresses to notify when a keyword is found.

***Global Interceptors note**: Configuring interceptors like the Keyword interceptor can be configured either for a specific space or for the entire 'global' community. To make sure that your interceptors are applying to content in all Spaces, Social Groups and personal content, you will need to always configure the interceptors as **Global Interceptors.** This can be done at Admin Console: System > Moderation > Moderation Configuration > Interceptors*

Configuring New user moderation

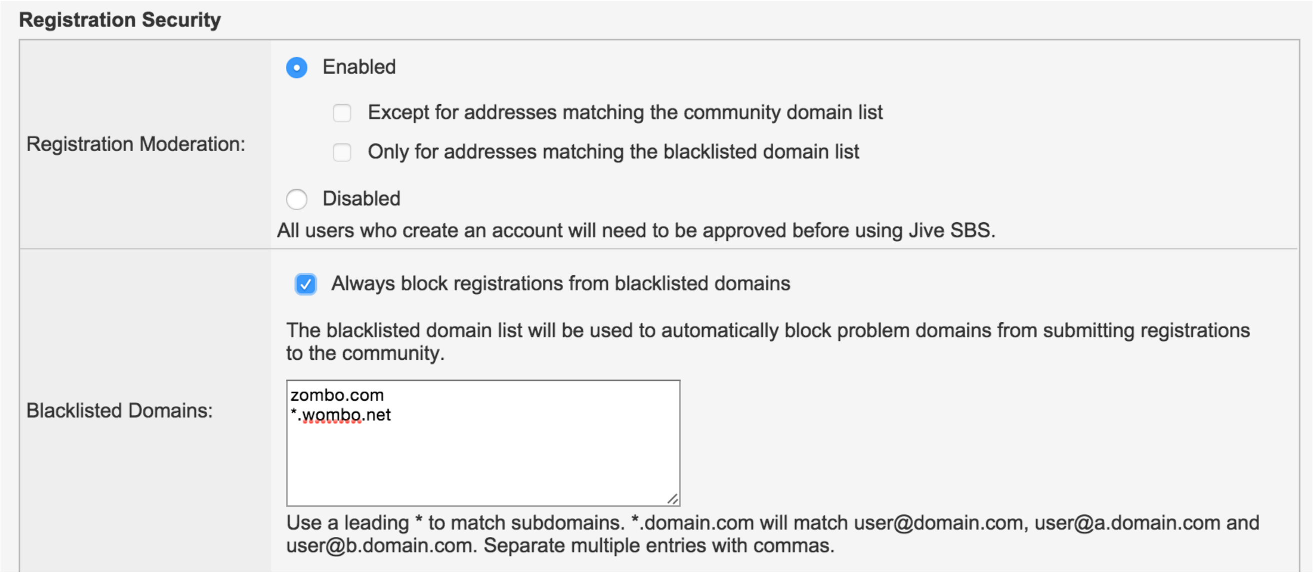
Fastpath: Admin Console: People > Settings > Registration Settings

If you allow users to create their own account, it can be helpful to turn on various registration moderation options in order to prevent spammers from creating new accounts. Please note, this method of moderation may be difficult to manage depending on the volume of new accounts created per day and your available community management resources. If your community uses SAML SSO for new user registration then this registration security configuration will not apply to your community.

To set up new user registration moderation:

1. Enable user self-registration by going to People > Settings > Registration Settings and select Enabled under "Allow user-created accounts." This allows users to create their own account from the login page. To learn more about the options on this page, see Configuring User Registration.

2. On the same page, under "Registration Moderation," select Enabled. This turns on the moderation feature for all new user registrations.
3. You can select Only for addresses matching the blacklisted domain list to limit registration moderation to those email address originating only from the domains that you blacklist. Alternatively, you can block registrations entirely from those you list in the blacklist box by selecting Always block registrations from blacklisted domains. To block or moderate all addresses from a domain, use an asterisk before the domain, e.g., *@domain.com.
    1. Please note, you cannot black list email subdomains in Jive 7 or older, e.g. *test.mymail.com. This issue has been fixed in Jive 8 and Jive Cloud
4. Make sure you have designated a Global Moderator(s). The application will send new user registration moderation requests to that user(s) first. If you don't have one, the application will send the request to the Full Access user(s).



Registration Security

Registration Moderation:

● Enabled

□ Except for addresses matching the community domain list

□ Only for addresses matching the blacklisted domain list

○ Disabled

All users who create an account will need to be approved before using Jive SBS.

Blacklisted Domains:

☑ Always block registrations from blacklisted domains

The blacklisted domain list will be used to automatically block problem domains from submitting registrations to the community.

```
zombo.com
*.wombo.net
```

Use a leading * to match subdomains. *.domain.com will match user@domain.com, user@a.domain.com and user@b.domain.com. Separate multiple entries with commas.

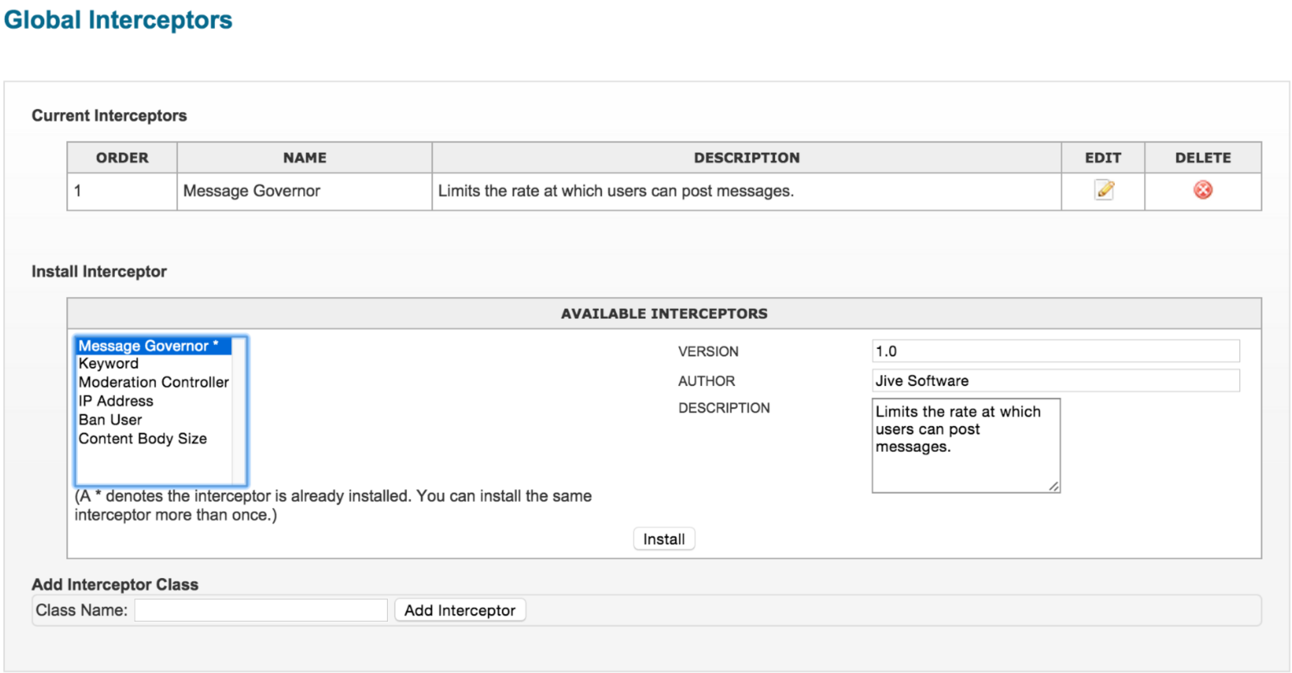
You can read more about setting up user registration moderation in our documentation.

Configuring Message governor interceptor

Fastpath: Admin Console: System > Moderation > Moderation Configuration > Interceptors

The message governor interceptor will allow you to prevent users from posting multiple pieces of content in quick succession. This can be helpful if you see that spammers are flooding your community with posts.

The default configuration for the message governor is 30 seconds for each post.



Global Interceptors

Current Interceptors

ORDER	NAME	DESCRIPTION	EDIT	DELETE
1	Message Governor	Limits the rate at which users can post messages.	✎	⊗

Install Interceptor

AVAILABLE INTERCEPTORS

- Message Governor *
- Keyword
- Moderation Controller
- IP Address
- Ban User
- Content Body Size

VERSION: 1.0

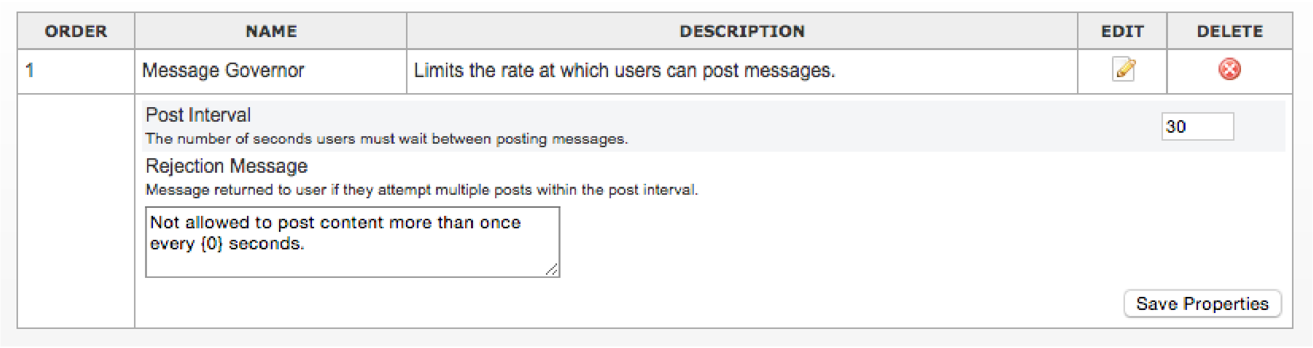
AUTHOR: Jive Software

DESCRIPTION: Limits the rate at which users can post messages.

(A * denotes the interceptor is already installed. You can install the same interceptor more than once.)

Add Interceptor Class

Class Name:



ORDER	NAME	DESCRIPTION	EDIT	DELETE
1	Message Governor	Limits the rate at which users can post messages.	✎	⊗

Post Interval

The number of seconds users must wait between posting messages.

30

Rejection Message

Message returned to user if they attempt multiple posts within the post interval.

Not allowed to post content more than once every {0} seconds.

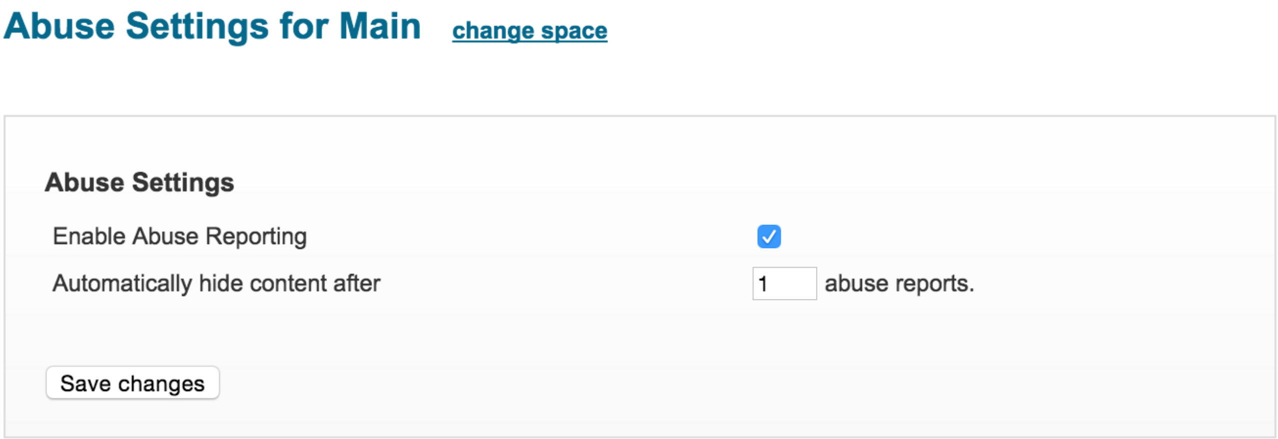
Configuring Abuse reporting

Fastpath: Admin Console: Spaces > Settings > Abuse Settings

Enabling the Abuse moderation in Jive will add a new button to content where any user can select to flag content as abuse. You can configure how many abuse reports are required before an item is immediately placed into moderation and removed from being publicly visible.

This tool can be helpful if you have spammers successfully posting content and you need to provide a quick way to hide this content as its found.

Abuse configuration



Abuse Settings for Main change space

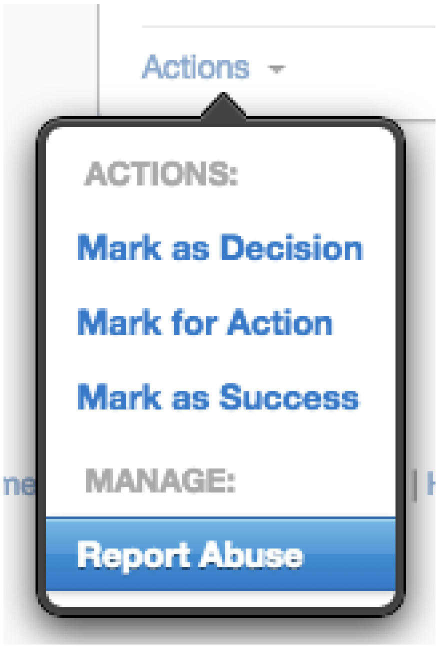
Abuse Settings

Enable Abuse Reporting ☑

Automatically hide content after 1 abuse reports.

Save changes

Marking an item as abuse



Configuring User and IP address bans

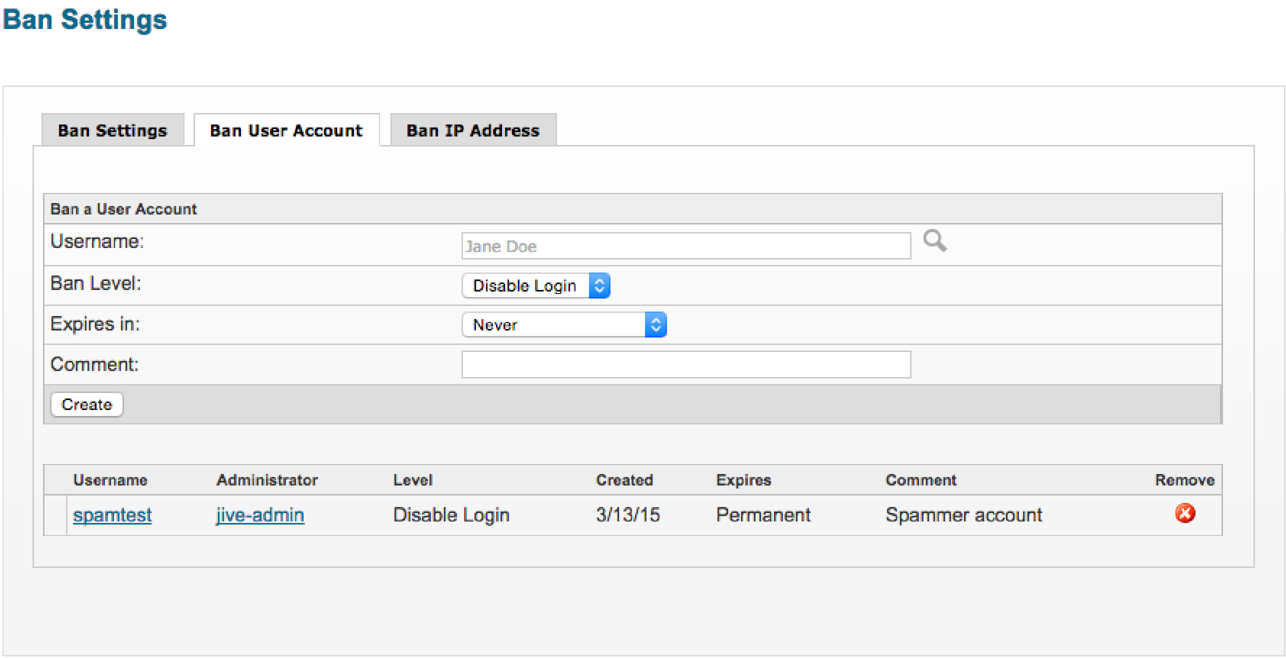
Fastpath: Admin Console: People > Settings > Ban Settings

You can ban users by either their Jive account or their IP address. This can be used to prevent users from logging into Jive.

If you wish to ban a user by IP address you will need open a new support case with Jive. Jive Support will need to examine the web server access logs to determine the IP address of the user who created the spam content.

Please note, a user's IP address may change, or they may post from a range of IP addresses, so this method of spam prevention is time consuming and is often considered a last resort. Banning by IP address will not work if you allow logged-out Guests to post content into your community. it is advised you turn off Guest posting if you are seeing this happen.

Banning a user by their User Account

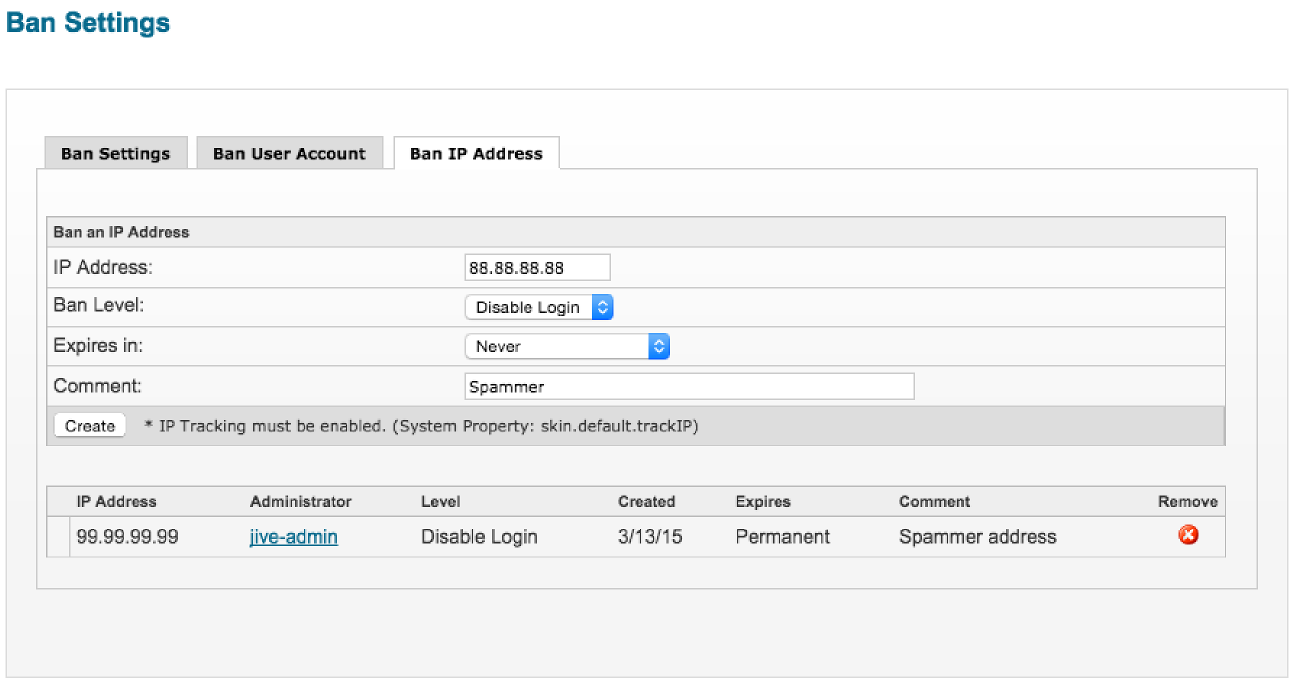


Ban a User Account

Username:	Jane Doe
Ban Level:	Disable Login
Expires in:	Never
Comment:	

Username	Administrator	Level	Created	Expires	Comment	Remove
spamtest	jive-admin	Disable Login	3/13/15	Permanent	Spammer account	✖

Banning a user by their IP address



Ban an IP Address

IP Address:	88.88.88.88
Ban Level:	Disable Login
Expires in:	Never
Comment:	Spammer

Create * IP Tracking must be enabled. (System Property: skin.default.trackIP)

IP Address	Administrator	Level	Created	Expires	Comment	Remove
99.99.99.99	jive-admin	Disable Login	3/13/15	Permanent	Spammer address	✖

Configuring Jive to Prevent Spam in Your Community

Configuring Jive to prevent spam is unique to each community and the type of spam being seen. There are several best practices that have been observed over time from communities who have had success in managing their spam:

1. Quarantining Spam
2. Common Spam Prevention Tools in Jive
3. Setting up Keyword Interceptors correctly

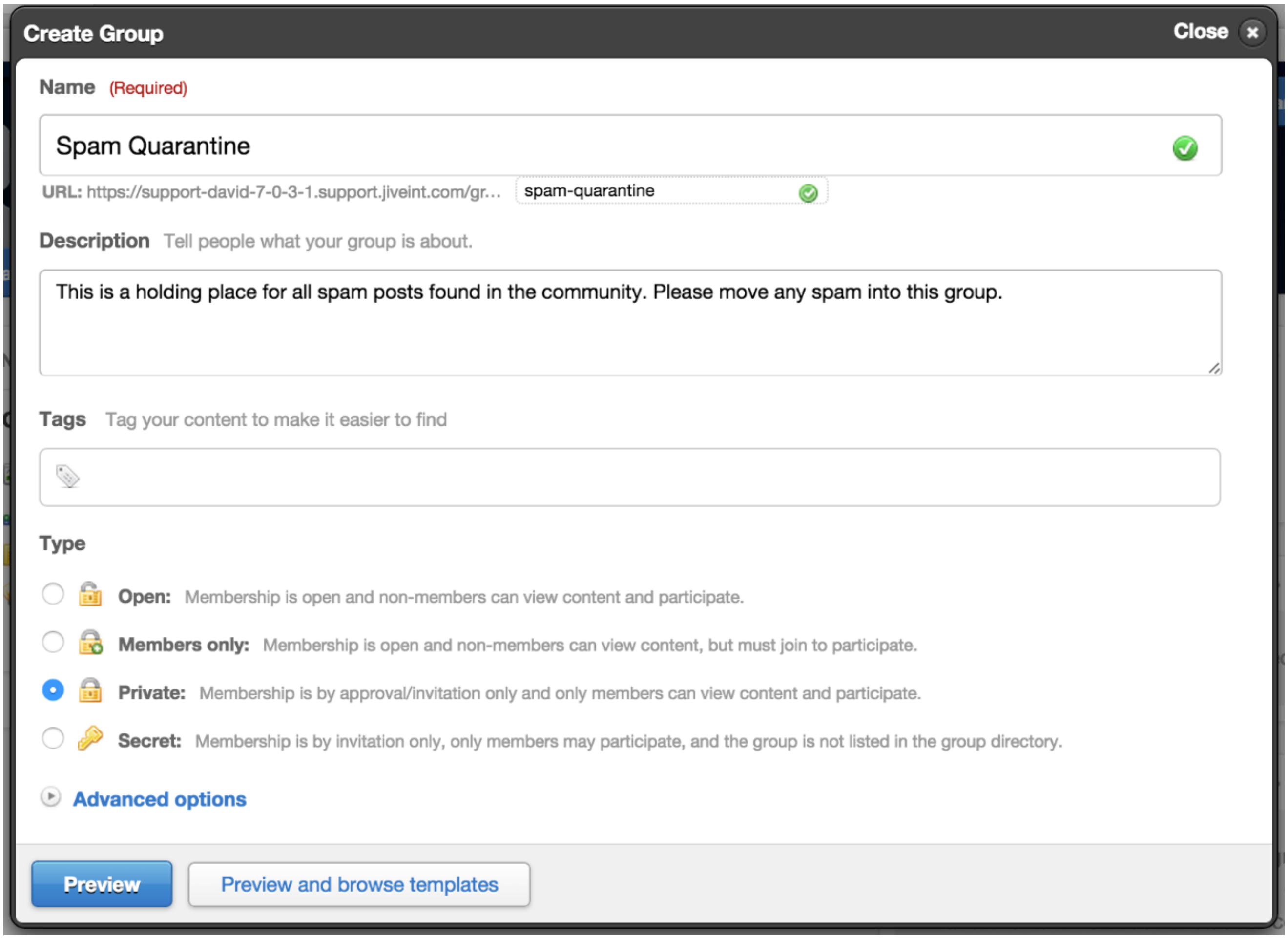
How to Quarantine New Spam

Before configuring any spam filtering rules in Jive it is recommended that you have an action plan in place for when you encounter spam in the community:

1. Immediately **disable the user** account that posted the spam
2. Move and **quarantine the spam content** into a dedicated private spam holding area in your community
    1. It is recommended that you set up a private social group in your community where you can move spam content

It is advised you keep the spam content on file while the spam event is happening so you can better understand the contents of the spam and build your spam filter accordingly. Deleting the user or the user's content will make this more difficult. It is advised that once the event is over you delete the spam content on your site.

To create a new spam quarantine area go to Create > Group, and create a new social group with the name of "Spam Quarantine". Be sure to set the group type to **Private** so that the spam content is not visible to your entire community.



Create Group

Name (Required)

Spam Quarantine

URL: https://support-david-7-0-3-1.support.jiveint.com/gr… spam-quarantine

Description Tell people what your group is about.

This is a holding place for all spam posts found in the community. Please move any spam into this group.

Tags Tag your content to make it easier to find

Type

○ **Open:** Membership is open and non-members can view content and participate.

○ **Members only:** Membership is open and non-members can view content, but must join to participate.

● **Private:** Membership is by approval/invitation only and only members can view content and participate.

○ **Secret:** Membership is by invitation only, only members may participate, and the group is not listed in the group directory.

Analyze Your Spam to Understand What Tools to Use

How you configure your Jive spam prevention tools is unique to each community depending on the types of users on the community, how they find and register on the site, how often they register new accounts, post new content, and how large your own community management team is.

There are several common tools we have seen being used across Jive's customers who have successfully managed spam:

1. Immediately **Disable user spam user accounts**
2. Turn on and configure **Link Moderation**
3. Turn on and configure **Keyword interceptors**
4. Turn on and configure **New user moderation**
5. Turn on and configure **Link moderation (points threshold)**

These are the tools that most communities have been successfully be using when encountering spam in their community. Please note that new user moderation may not be viable if your community has a large rate of new users being created and you do not have the available resources to manually review these new user accounts.

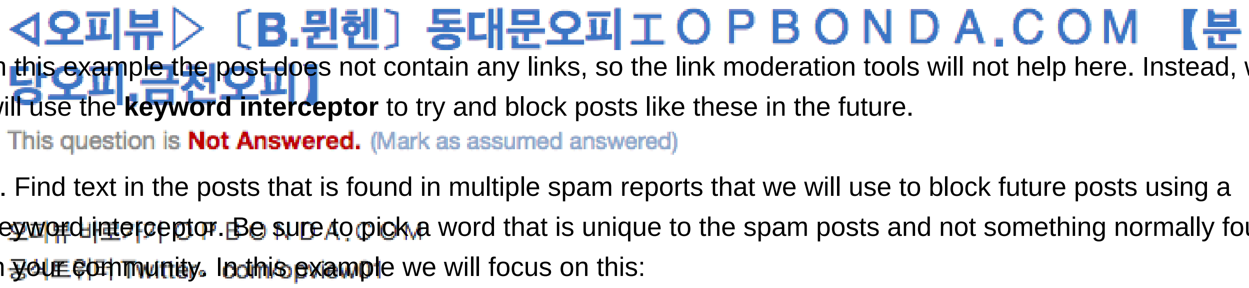
Example of Configuring a Keyword Interceptor Rules

The keyword interceptor is an powerful tool for allowing you to block content that only has certain keywords or phrases and allowing all other pieces of content.

There has been a recent rise in spam containing Korean content in Jive Communities. In this section we will provide an example of a spam post we've observed and the keyword interceptor rules used to block this type of post.

Please note, the actual contents of the spam posts will change over time. This means that you will need to adjust and adapt your keyword interceptors over time.

Example Spam Post

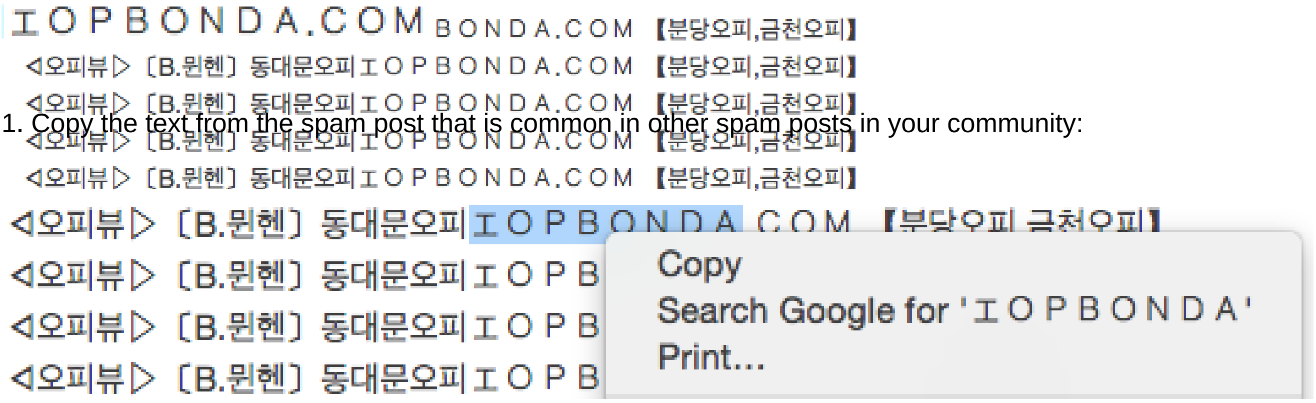


◁오피뷰▷〔B.뮌헨〕동대문오피ＩＯＰＢＯＮＤＡ.ＣＯＭ 【분당오피,금천오피】

This question is **Not Answered.** (Mark as assumed answered)

In this example the post does not contain any links, so the link moderation tools will not help here. Instead, we will use the **keyword interceptor** to try and block posts like these in the future.

1. Find text in the posts that is found in multiple spam reports that we will use to block future posts using a keyword interceptor. Be sure to pick a word that is unique to the spam posts and not something normally found in your community. In this example we will focus on this:

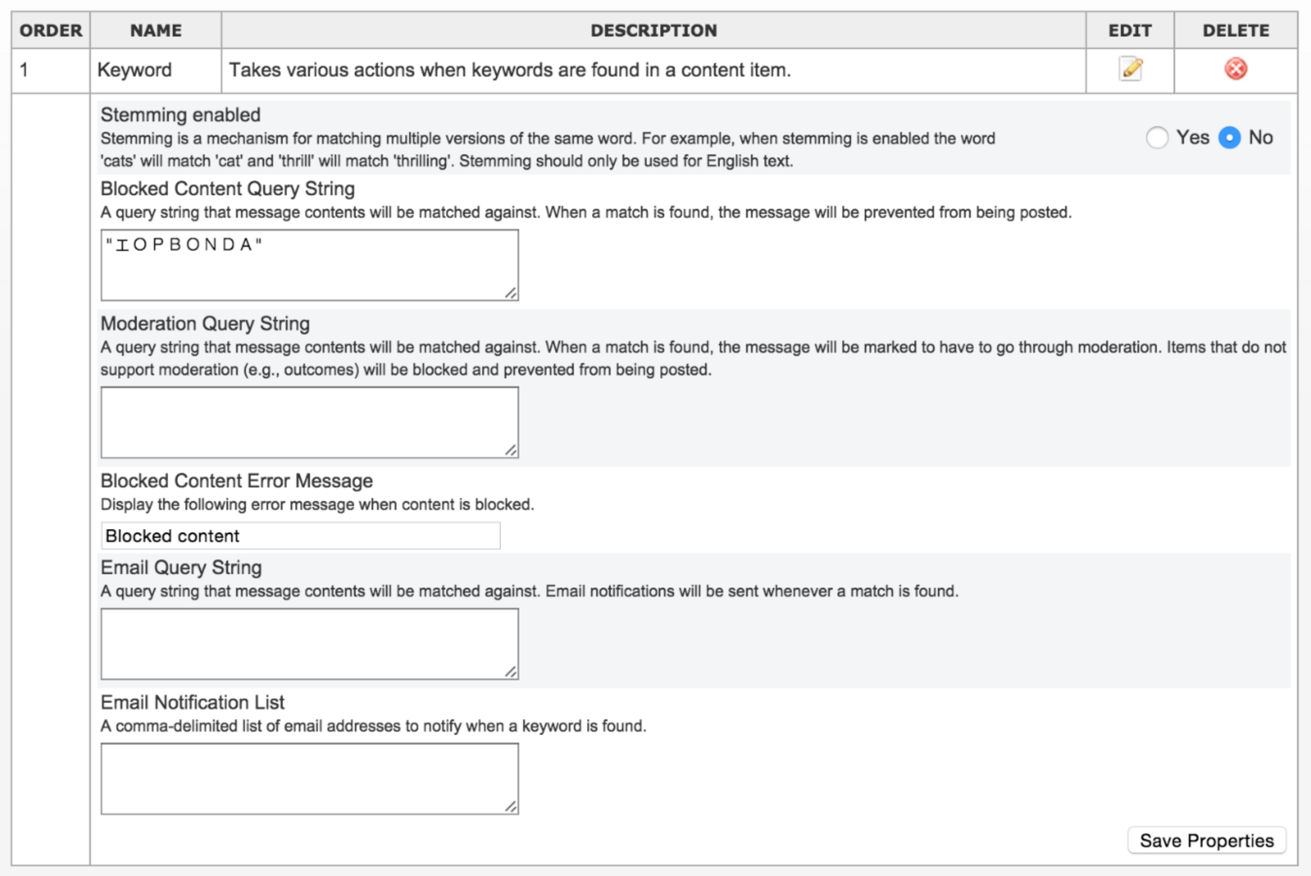


ＩＯＰＢＯＮＤＡ.ＣＯＭ ＢＯＮＤＡ.ＣＯＭ 【분당오피,금천오피】
◁오피뷰▷〔B.뮌헨〕동대문오피ＩＯＰＢＯＮＤＡ.ＣＯＭ 【분당오피,금천오피】
◁오피뷰▷〔B.뮌헨〕동대문오피ＩＯＰＢＯＮＤＡ.ＣＯＭ 【분당오피,금천오피】
◁오피뷰▷〔B.뮌헨〕동대문오피ＩＯＰＢＯＮＤＡ.ＣＯＭ 【분당오피,금천오피】
◁오피뷰▷〔B.뮌헨〕동대문오피ＩＯＰＢＯＮＤＡ.ＣＯＭ 【분당오피,금천오피】
◁오피뷰▷〔B.뮌헨〕동대문오피ＩＯＰＢＯＮＤＡ.ＣＯＭ 【분당오피,금천오피】
◁오피뷰▷〔B.뮌헨〕동대문오피ＩＯＰＢ
◁오피뷰▷〔B.뮌헨〕동대문오피ＩＯＰＢ
◁오피뷰▷〔B.뮌헨〕동대문오피ＩＯＰＢ

1. Copy the text from the spam post that is common in other spam posts in your community:

2. Create a new Keyword Interceptor in the admin console

3. Paste the copied text into the "Blocked Content Query String" text field. **Be sure to wrap the text in double quote characters.**



ORDER	NAME	DESCRIPTION	EDIT	DELETE
1	Keyword	Takes various actions when keywords are found in a content item.	✎	⊗

Stemming enabled
Stemming is a mechanism for matching multiple versions of the same word. For example, when stemming is enabled the word 'cats' will match 'cat' and 'thrill' will match 'thrilling'. Stemming should only be used for English text.
○ Yes ● No

Blocked Content Query String
A query string that message contents will be matched against. When a match is found, the message will be prevented from being posted.
"エ O P B O N D A"

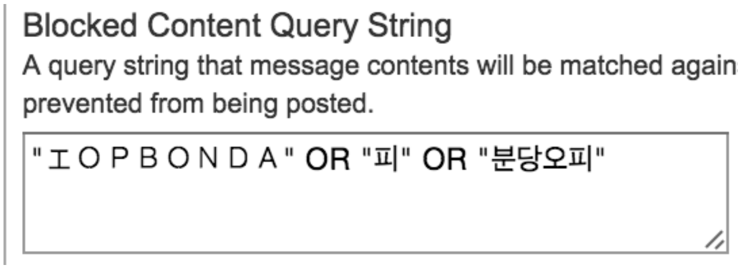
Moderation Query String
A query string that message contents will be matched against. When a match is found, the message will be marked to have to go through moderation. Items that do not support moderation (e.g., outcomes) will be blocked and prevented from being posted.

Blocked Content Error Message
Display the following error message when content is blocked.
Blocked content

Email Query String
A query string that message contents will be matched against. Email notifications will be sent whenever a match is found.

Email Notification List
A comma-delimited list of email addresses to notify when a keyword is found.

As you need to add more keywords or phrases to your keyword interceptor be sure to separate them with " OR ":



Blocked Content Query String
A query string that message contents will be matched against. When a match is found, the message will be prevented from being posted.
"エ O P B O N D A" OR "피" OR "분당오피"

If you're still having issues with managing spam or have questions please contact Jive Support by opening a new support case.